Title
Observation and measurement of photon-induced WW production in pp collisions at sqrt(s) = 13 TeV using the ATLAS detector

Author
McCormack, William Patrick

Publication Date
2021

Peer reviewed|Thesis/dissertation

Observation and measurement of $\gamma\gamma \to W^+W^-$ in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector

by

William Patrick McCormack III

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Dr. Maurice Garcia-Sciveres, Co-chair
Professor Marjorie Shapiro, Co-chair
Professor Yury Kolomensky
Professor Karl Van Bibber

Spring 2021

Observation and measurement of $\gamma\gamma \to W^+W^-$ in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector

William Patrick McCormack III

Abstract

Observation and measurement of $\gamma\gamma \to W^+W^-$ in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector

by

William Patrick McCormack III

Doctor of Philosophy in Physics

University of California, Berkeley

Dr. Maurice Garcia-Sciveres, Co-chair

Professor Marjorie Shapiro, Co-chair

While CERN's Large Hadron Collider (LHC) is primarily a high-energy proton or heavy ion collider, it can also be used as a high-energy photon collider. This is possible because the highly boosted electromagnetic fields surrounding the circulating ions behaves like a flux of photons. A search for the process $\gamma\gamma \to W^+W^-$ was conducted in the 139 fb^{-1} of proton-proton collision data collected in LHC Run 2 between 2015 and 2018. Events were tagged using the $e^{\pm}\nu\mu^{\mp}\nu$ decay channel of the W^+W^- system, where the event was determined to be photon-induced due to a lack of charged hadronic activity near the leptons. The process was observed with a significance of 8.4σ, with a measured fiducial cross section of 3.13 ± 0.31 (stat.) $\pm$ 0.28 (syst.) fb. This fiducial cross-section agrees with Standard Model expectations. A strategy to improve the $\gamma\gamma \to W^+W^-$ measurement with the use of a novel low-p_{T} tracking algorithm in ATLAS is also presented. The use of low-p_{T} tracks improves the rejection of QCD-induced backgrounds to $\gamma\gamma \to W^+W^-$, leading to an expected increase in statistical significance of 3–4σ.

And God said, Let there be light: and there was light.

To my parents

Contents

List of Figures

List of Tables

Acknowledgments

The acknowledgements section is the most important part of any thesis; as they say, raising a scientist takes a village. I count myself very lucky that I have never had a lack of supportive mentors and teachers.

First, I have to thank my many colleagues at LBNL. To Maurice: thank you for taking me on as a SULI student six years ago and then being a great advisor. Thank you for always reminding me of the bigger picture and never running out of fresh ideas. To Simone: I couldn't have asked for a better mentor; thank you for showing me how to actually carry out an analysis, for your expertise, and most of all for always being available to help. To Aleksandra: thank you for showing me how to be a serious researcher and for helping me out with all my problems, big and small. To Juerg: thank you for many interesting discussions and for helping me to remember to keep my code and ntuples comprehensible. To Ben: thank you for so many interesting projects, helping me when I was just getting started, and for countless pieces of advice. To Timon: thank you for staying patient with me in the lab and for basically teaching me how hardware works. To Marjorie: thank you for always being sympathetic and for being next door to answer whatever questions I had.

I also have to thank Kevin Lannon, my research advisor at Notre Dame, for always giving me interesting projects to work on and for many helpful discussions about how to be a physicist. And I also want to thank Mr. Adams for making my first real encounter with Physics so interesting. There probably isn't a better high school Physics teacher out there.

Thank you to all the friends I've made at Berkeley for making grad school an enjoyable experience. Thank you to Rebecca, Jennet, Cesar, Greg, Ryan, and Tom-Erik for the most entertaining lunches a DOE site has ever hosted. Thank you to the 7A crew, Reed, the Sams, Liz, Best, and Isaac for making our first year entertaining. Thanks to all the Soft Scatterers and Tsar Bomba for being the most fun teams in their league. And a special thanks to Neha for being a friend through classes, GSI-ing, and research.

And above all, thank you to my family and to Anna. To Anna: thank you for being you and for keeping me sane these past years. To Megan: thank you for endless laughs and for being the best Disney partner around. And Mom and Dad, no son has ever been luckier. Thank you for everything.

Chapter 1

Introduction

The Standard Model is the most rigorously tested theory in the history of science. It mathematically encapsulates three of the four fundamental forces of nature, Electromagnetism, the Weak Force, and the Strong Force, and all of the known fundamental particles. And just as Maxwell unified Electricity and Magnetism into the theory of Electromagnetism, the Standard Model unifies Electromagnetism and the Weak Force into a combined Electroweak theory.

Electroweak theory describes a rich phenomenology of Boson interactions with only a few fundamental parameters, which have been well measured within the past century. However, not all predicted interaction have been observed yet. One such interaction is the process $\gamma\gamma \to W^+W^-$, or the conversion of two photons into two W Bosons. The leading order Feynman diagrams for this process are shown in Figure 1.1. Because the Electroweak theory is overconstrained, it is important to test all of its predictions. A deviation from expectation could be a hint at the nature of Beyond the Standard Model (BSM) physics. In particular, the quartic vertex in the right-most diagram of Figure 1.1 can be strongly influenced by BSM physics, and it is difficult to probe through any other mechanism.

Observing this process is an experimental challenge: initial state photons with a center of mass energy of 160 GeV are needed. There is currently no photon collider that can access such energies, so we must turn to photons radiated from the 6.5 TeV protons at the

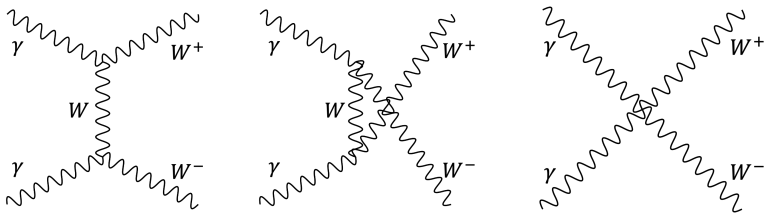


Figure 1.1: Leading order Feynman diagrams that contribute to $\gamma\gamma \to WW$ scattering.

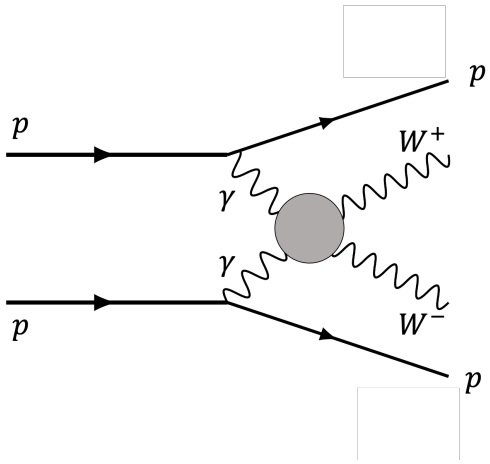


Figure 1.2: Illustration of $\gamma\gamma \to W^+W^-$ at a proton collider. In this diagram, both protons remain intact after radiating photons, but one or both could dissociate in reality.

Large Hadron Collider (LHC). Figure 1.2 illustrates such an interaction, showing a case where both protons remain intact after radiating photons. While the protons at the LHC provide access to the $\gamma\gamma \to W^+W^-$ process, they also introduce significant backgrounds, as QCD-induced WW production has a much higher cross-section that that predicted for $\gamma\gamma \to W^+W^-$. Discriminating the photon-induced signal from background in the LHC environment requires accurate charged particle reconstruction. While the final state W bosons in $\gamma\gamma \to W^+W^-$ will be isolated, QCD-induced WW will be surrounded by hadronic material due to the direct interaction of the initial state protons.

This document details the first ever observation of $\gamma\gamma \to W^+W^-$. The analysis was performed using the 139 fb^{-1} of proton-proton collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector throughout 2015 – 2018. The fiducial cross-section was measured to be 3.13 ± 0.31 (stat.) ± 0.28 (syst.) fb. This agrees with the Standard Model within two standard deviations, constituting a signal strength of $1.33^{+0.14}_{-0.14}$(stat.)$^{+0.22}_{-0.17}$(syst.) relative to the Standard Model expectation.

It also details the development of low-p_{T} tracking for use in nominal pileup environments. The use of low-p_{T} tracks will improve signal vs. background discrimination, enhancing the $\gamma\gamma \to W^+W^-$ analysis and any future analysis of photon-induced phenomena with ATLAS.

This thesis is structured as follows:

- Part I explains the background knowledge needed to understand this thesis. Chapter 2 explains the phenomenology of the $\gamma\gamma \to W^+W^-$ process, relevant historical and contemporary measurements, and how the process can be used to constrain BSM physics. Chapter 3 explains practical experimental aspects of the analysis, including the LHC and the construction and design of the ATLAS detector. Chapter 4 continues the discussion of Chapter 3, detailing how particles are reconstructed by ATLAS. Lastly,

Chapter 5 explains the simulations used to understand what signal and background will look like in ATLAS.

- Part II details the complete $\gamma\gamma \to W^+W^-$ analysis without low-p_{T} tracks. Chapter 6 explains the strategy behind this analysis, laying out the steps needed to make an observation. Chapters 7 – 10 explain corrections that are made to the simulation in order to make an accurate prediction of background rates. Chapter 11 introduces the signal region and main control regions used to isolate signal and study major backgrounds in the analysis. Chapter 12 details sources of uncertainty in the analysis and Chapter 13 presents the final result.

- Part III introduces the concept of low-p_{T} tracking and motivates its use for a re-analysis of $\gamma\gamma \to W^+W^-$. Chapter 15 illustrates how low-p_{T} tracks will improve the analysis, and Chapter 16 details the implementation of the algorithm used to reconstruct them. Chapter 17 explains how high-quality low-p_{T} tracks are selected for the re-analysis and new cuts that will be applied to events for signal vs. background discrimination. Lastly, Chapter 18 provides the first comparison of ATLAS data with low-p_{T} tracks in nominal pileup to simulation, proving with data that low-p_{T} tracking improves background rejection for the $\gamma\gamma \to W^+W^-$ analysis and presenting the expected results of the re-analysis.

Additionally, the appendices detail various projects that the author worked on throughout graduate school. Appendix A is most relevant to the $\gamma\gamma \to W^+W^-$ analysis presented in the main body of the thesis, detailing the optimization of the criteria used to select events for Part II. Appendices B – F present projects to improve tracking in ATLAS, implement a new luminosity measurement algorithm, discriminate quark and gluon jets, and calibrate thin silicon sensors. Appendix G contains a lecture series which provides an introduction to the mathematics underlying the Standard Model intended for an undergraduate-level audience.

Part I

Foundations of Particle Experiments

Chapter 2

Electroweak and Photon-Induced Physics

This chapter lays out the relevant background needed to understand the rest of this thesis, including a brief summary of the Electroweak theory and a theoretical and experimental introduction to photon-induced physics. A gentler introduction[1] to the history and mathematics behind this chapter can be found in Appendix G.

Section 2.1 describes the Standard Model with particular focus on the Electroweak Sector. Section 2.2 covers more practical aspects of performing cross-section calculations of hadron collisions, particularly for photon-induced processes at hadron colliders. Section 2.3 provides a survey of recent and historical experimental results relevant to this thesis. Lastly, Section 2.4 looks to the unknown, describing possible modifications to the Standard Model and the role that the $\gamma\gamma \to W^+W^-$ process plays in constraining such modifications.

2.1 The Standard Model

The Standard Model (SM) of particle physics is the great triumph of physics in the second half of the 20^{th} century. It describes almost every particle interaction that has ever been observed, making predictions that agree with experiment over many orders of magnitude of energy and rareness. Similarly, most field interactions predicted by the SM have been observed. However, there are still SM processes that have not yet been observed due to energy constraints or experimental difficulty. This thesis presents the observation and measurement of a process predicted by the SM that had not been previously established: the conversion of two photons into two W Bosons ($\gamma\gamma \to W^+W^-$).

[1] Aimed at the level of a curious undergraduate student

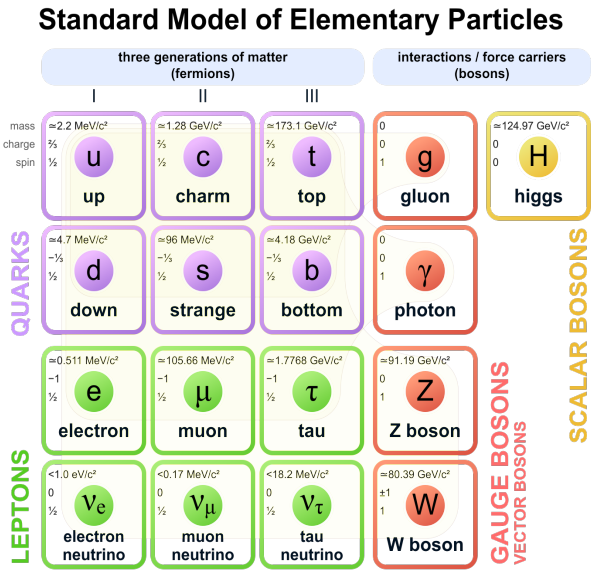


Figure 2.1: The particles of the Standard Model. There are 12 Fermionic particles, four vector Bosons, and one scalar particle, called the Higgs Boson.

2.1.1 Particle Content of the Standard Model

The SM is constructed using the tools discussed in Appendix G. There are Fermion fields, three gauges and their corresponding vector Bosons, and one scalar field that breaks gauge symmetries when it acquires a vacuum expectation value (*vev*). Some of the Fermion and Boson fields acquire mass through this symmetry breaking.

The particle content of the SM is shown in Figure 2.1. All of the particles in this figure are "fundamental" in the sense that they have no known constituents or dimension. They are all of the fundamental particles that have ever been observed experimentally. Many other well-known particles, such as protons and neutrons, are actually composite particles, being composed of fundamental particles just as atoms are actually composed of electrons, protons, and neutrons. The mass, charge, and spin of the particles are expressed in the figure.

There are 12 Fermions in the SM, which are arranged in three generations. The third generation particles typically have the highest mass and the first generation the lowest. Otherwise each generation has the same interactions as the other generations row-by-row. For example, the electron, muon, and tau all interact in the same way with respect to all of the forces in the SM, though the muon and tau are both able to decay because they have higher mass than the electron. The Fermions are further divided into "leptons" and "quarks". The chief difference between leptons and quarks is that quarks can interact with gluons while leptons cannot.

There are four vector Bosons, which carry the three forces that are expressed by the SM: the strong nuclear force (also known as Quantum Chromodynamics or QCD), the weak force, and Electromagnetism. The gluon and the photon are both massless and electrically neutral, the Z Boson is massive and electrically neutral, and the W Boson is massive and has an electric charge equal to that of the electron (there is a $+e$ and $-e$ version of the W Boson). The gluon is the only Boson that is charged under the strong force. There is also the scalar Higgs Boson, which is massive and electrically neutral.

2.1.2 The Standard Model Lagrangian

The starting point of constructing the SM is its three gauge symmetries. There is an $SU(3)_C$ gauge, where the C stands for "color", an $SU(2)_L$ gauge, where the L stands for "left", and an $U(1)_Y$ gauge, where the Y stands for "hypercharge". A complex doublet scalar Boson (the Higgs Boson), which is a doublet under $SU(2)_L$ and charged under $U(1)_Y$, is used to break the $SU(2)_L$ and $U(1)_Y$ symmetries[2]. The $U(1)$ symmetry of Electromagnetism emerges from this symmetry breaking. Before the introduction of the Higgs's *vev*, there are 8 massless vector Bosons from $SU(3)_C$, 3 massless Bosons from $SU(2)_L$, and 1 massless Boson from $U(1)_Y$. The 8 $SU(3)_C$ Bosons are treated as the same particle: the gluon. The 3 $SU(2)_L$ and 1 $U(1)_Y$ Boson combine to become the W Bosons, the Z Boson, and the photon.

The Fermions of the SM are introduced as separate, massless left-handed spinors and massless right-handed spinors. The Higgs mechanism is used to combine the massless spinors into the massive particles that we observed experimentally. The introduction of massless single-handed spinors is not just for aesthetics. The left-handed spinors are combined into $SU(2)_L$ doublets, while the right-handed spinors are *not* charged under ${SU(2)_L}$[3]. Therefore, in full, the Fermion fields in the SM are:

$$\begin{aligned} &L: \begin{pmatrix}\nu_L^e\\ e_L\end{pmatrix},\ \begin{pmatrix}\nu_L^\mu\\ \mu_L\end{pmatrix},\ \begin{pmatrix}\nu_L^\tau\\ \tau_L\end{pmatrix},\ \begin{pmatrix}u_L\\ d_L\end{pmatrix},\ \begin{pmatrix}c_L\\ s_L\end{pmatrix},\ \begin{pmatrix}t_L\\ b_L\end{pmatrix} \\ &R: e_R,\ \mu_R,\ \tau_R,\ u_R,\ d_R,\ c_R,\ s_R,\ t_R,\ b_R. \end{aligned} \tag{2.1}$$

The names of the quarks are up, down, charm, strange, top, and bottom. The leptons are the electron, muon, and tau, along with their respective neutrinos. In the SM, there are no right-handed neutrinos, so the neutrinos are massless particles. It is actually experimentally known that the neutrinos *do* have a very small mass, but it is unclear whether they acquire their mass through the Higgs mechanism. If the neutrinos do get mass through the Higgs mechanism, then there may indeed be right-handed neutrinos.

Given the particle content discussed in the last two paragraphs, the Lagrangian for the SM can then be written as:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Fermion}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \tag{2.2}$$

[2]Mathematical details are explained in Section G.2.6 of Appendix G

[3]The underlying mathematics of this mechanism are also laid out in Section G.2.6

The $\mathcal{L}_{\text{gauge}}$ component is associated with the gauge Bosons:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}(G^a_{\mu\nu})^2 - \frac{1}{4}(W^a_{\mu\nu})^2 - \frac{1}{4}B^2_{\mu\nu}. \tag{2.3}$$

Here, terms like $W_{\mu\nu}$ represent $W_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g f^{abc} W^b_\mu W^c_\nu$, where f^{abc} is the structure function of the group associated with the gauge Boson. The G^a_μ fields are the gluons, eight for the eight generators of $SU(3_C)$. The W^a_μ fields are the three fields of the $SU(2)_L$ group. The B_μ field is associated with the $U(1)_Y$ group; this is the only Abelian gauge group in the SM, so it has $f^{abc} = 0$.

The $\mathcal{L}_{\text{Fermion}}$ component of the Lagrangian can be written as:

$$\mathcal{L}_{\text{Fermion}} = \sum_{\text{Fermions}} i\bar{\psi}\not{D}\psi, \tag{2.4}$$

where the sum runs over each of the Fermion fields in Eq. 2.1. For the left-handed fields, ψ is in the form of an $SU(2)$ doublet, and for the right-handed field, ψ is a singlet. D_μ is the covariant derivative, which takes the form

$$D_\mu = \partial_\mu - i g_c q_c G^a_\mu \frac{\lambda^a}{2} - i g_L q_L W^a_\mu \frac{\sigma^a}{2} - i g_Y q_Y B_\mu. \tag{2.5}$$

The three g parameters are the overall coupling strengths of the three gauges, and the three q parameters represent the "charges" of the Fermion fields under the gauge relative to the coupling strength. Each field has a different set of q values as expressed in Table 2.1. The λ^a are the 3x3 matrices that generate $SU(3)$, also known as the Gell-Mann matrices [1], and the σ^a matrices are the Pauli matrices[4]. In both cases, the factors of $\frac{1}{2}$ are just a conventional normalization. After the Higgs field breaks the $SU(2)_L$ and $U(1)_Y$ gauges down to $U(1)_{EM}$, the electric charges of the particles are given by q_y for the right-handed Fermions, $(\frac{1}{2} + q_y)$ for the left-handed Fermions on the top of their doublet, and $(-\frac{1}{2} + q_y)$ for the left-handed Fermions on the bottom of their doublet.

The $\mathcal{L}_{\text{Higgs}}$ component of the Lagrangian can be written as:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu H)^\dagger (D_\mu H) + \mu^2 H^\dagger H - \lambda (H^\dagger H)^2, \tag{2.6}$$

where the covariant derivative is the same as in Eq. 2.5, with the gauge charges of the Higgs field from Table 2.1.

Lastly, the $\mathcal{L}_{\text{Yukawa}}$ component of the Lagrangian can be written as:

$$\mathcal{L}_{\text{Yukawa}} = -y\bar{L}^i_L H l^i_R + h.c. - Y^d_{ij}\bar{Q}^i_L H d^j_R + h.c. - Y^u_{ij}\bar{Q}^i_L \tilde{H} u^j_R + h.c. \tag{2.7}$$

The sum over the repeated i index in the first term represents a sum over all of the lepton generations (or "flavors"). The Higgs *vev* will generate mass terms for the bottom fields

[4] see Eq. G.3

Field	q_C	q_L	q_Y
L_L	0	1	$-1/2$
Q_L	1	1	1/6
e_R	0	0	-1
ν_R	0	0	0
u_R	1	0	2/3
d_R	1	0	$-1/3$
H	0	1	1/2

Table 2.1: The charges of the Standard Model Fermion fields and Higgs Boson under the three gauges. Each generation of Fermion has the same charges, so only the first generation is written explicitly. The L_L field stands for the left-handed ν^e-electron doublet, and the Q_L field stands for the left-handed u-d doublet. A right-handed neutrino is included here for reference; it would not couple to any gauge.

in the lepton $SU(2)$ doublets: the electron, muon, and tau. In the two explicitly written quark terms, the sums over i and j run over the three generations. The first quark term will generate mass for the down-type quarks. The $\tilde{H}$ field that was introduced is just $\tilde{H} = i\sigma_2 H$, which allows the Higgs vev to appear in the top of the doublet, generating mass terms for the up-type quarks. For the quark terms, there are Yukawa matrices Y^d_{ij} and Y^u_{ij}. These are 3x3 matrices, where the indices run over the generations. There are significant physical consequences due to these matrices. The matrices allow the mass terms to be mixed between generations. We are free to choose a basis where there are no mass-mixing terms between the generations, in which case what we find is that the W Bosons allows direct interactions between the different generations of quarks in a single Fermion current[5]. This is in contrast to the lepton sector of the SM. There, the W Boson does not mix the generations: if a W Boson couples to a lepton current, the current must contain only one generation, such as a $\nu^e - e$ pair. If a W Boson couples to a u-quark, it could be part of a $u - d$ current, a $u - s$ current, or a $u - b$ current for example.

In total, there are 18 real parameters in the SM.

- Three gauge couplings: g_C, g_L, g_Y
- Six quark masses, three lepton masses
- Four mixing parameters associated with the Yukawa matrices for the quarks
- The μ and λ from $\mathcal{L}_{\text{Higgs}}$ (which can alternatively be expressed as m_h and the vev)

[5]If we choose to write the SM Lagrangian in the mass basis after symmetry breaking, the mixing between generations shows up in $\mathcal{L}_{\text{Fermion}}$

This list can be expanded if we allow for neutrino masses through the introduction of right-handed neutrinos. In this case, there are three additional neutrino masses, and intergeneration mixing parameters for leptons, just as in the quark case.

The SM is a highly overconstrained system in the sense that every particle physics process should be a function of those 18 parameters alone. From another perspective, this implies that every particle physics experiment constitutes a measurement of at least one of those 18 parameters. If two experiments find contradictory values for the same parameter, that would imply that the SM needs to be modified somehow. In the next section, we will explore some of the rich phenomenology of Electroweak physics, which relies only on a small subset of the SM parameters. The parameters that govern the $\gamma\gamma \to W^+W^-$ process[6] have been measured in other experiments using other processes. Because of this, accurate predictions of the production rate of $\gamma\gamma \to W^+W^-$ can be made. If our measurement of the process disagrees with the prediction, that would suggest that some Beyond the Standard Model (BSM) physics occurred.

2.1.3 Interactions Among the Vector Bosons

The parts of the SM that relate to the $SU(2)_L$ and $U(1)_Y$ gauge symmetries and their breaking due to the Higgs mechanism are known as "Electroweak physics"[7].

We can start from a Lagrangian of the form

$$\mathcal{L} = -\frac{1}{4}(W^a_{\mu\nu})^2 - \frac{1}{4}B^2_{\mu\nu} + (D_\mu H)^\dagger (D_\mu H) + \mu^2 H^\dagger H - \lambda (H^\dagger H)^2, \tag{2.8}$$

using the covariant derivative from Eq. 2.5. When the Higgs field acquires its *vev*, $v = \frac{m}{\sqrt{\lambda}}$, it can be written as

$$H = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} + \frac{h}{\sqrt{2}} \end{pmatrix}. \tag{2.9}$$

Substituting this form for H into Eq. 2.8 generates mass terms for the Bosons:

$$(D_\mu H)^\dagger (D_\mu H) = g^2 \frac{v^2}{8}\Big[(W^1_\mu)^2 + (W^2_\mu)^2 + \Big(\frac{g'}{g}B_\mu - W^3_\mu\Big)^2\Big]. \tag{2.10}$$

The mass terms for the B and W^3 are mixed, but can be separated by defining the fields

$$\begin{aligned} Z_\mu &= \cos\theta_w W^3_\mu - \sin\theta_w B_\mu \\ A_\mu &= \sin\theta_w W^3_\mu + \cos\theta_w B_\mu, \end{aligned} \tag{2.11}$$

where the mixing angle θ_w is defined by $\tan\theta_w = \frac{g_Y}{g_L}$. The Z field here is just the massive SM Z, and A is the massless photon. The mass of the Z Boson is

$$m_Z = \frac{1}{2\cos\theta_w} g_L v.$$

[6]The signal process in this thesis

[7]More detail on the mathematics can be found in Section G.2.6 of Appendix G

This is purely a function of the g_Y, g_L, μ, and λ parameters in the SM.

To make the SM W Bosons, which have charge of $\pm e$, we need to form linear combinations of the W^1 and W^2 Bosons:

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

We will find that the identification $e = g_L \sin\theta_w = g_Y \cos\theta_w$ simplifies upcoming equations. This is the coupling strength of the W Boson to the photon, which is the same as electric charge of the electon. Because of this this, we can see that electric charge is a function of the g_Y and g_L parameters of the SM. We can also take a moment to note from Eq. 2.10 that the mass of the W Boson is given by

$$m_W = \frac{v}{2} g_L.$$

The masses of the Z and W Bosons are related by $m_W = m_Z \cos\theta_w$. All of these values are functions of g_Y, g_L, μ, and λ, and the SM tells us unambiguously that the W Boson has a smaller mass than the Z Boson.

If we write out all of the interaction terms in Eq. 2.8 after symmetry breaking explicitly, we find terms that correspond to ZWW, AWW, $WWWW$, $ZZWW$, $AAWW$, and $AZWW$ interactions. The coupling strengths of all of these interactions depend on different combinations of e and trigonometric functions of θ_w. This thesis is concerned primarily with the interactions of photons and W Bosons; the terms corresponding to these interactions are:

$$\begin{aligned}\mathcal{L}_{\gamma+\mathrm{W}} = & -ie\Big[\partial_\mu A_\nu (W_\mu^+ W_\nu^- - W_\nu^+ W_\mu^-) \\ & + A_\nu(-W_\mu^+ \partial_\nu W_\mu^- + W_\mu^- \partial_\nu W_\mu^+ + W_\mu^+ \partial_\mu W_\nu^- - W_\mu^- \partial_\mu W_\nu^+)\Big] \\ & + e^2(A_\mu W_\mu^+ A_\nu W_\nu^- - A_\mu A_\mu W_\nu^+ W_\nu^-).\end{aligned} \tag{2.12}$$

It should not be surprising here that the coupling strengths in both of these terms rely only on e. The Feynman diagrams associated with these terms are shown in Figure 2.2. The forms of the terms dictating the other 3- and 4-pronged interactions among the gauge Bosons are exactly as in Eq. 2.12 up to the coupling strengths, which are presented in Table 2.2.

It is also possible to expand out the Eq. 2.8 to find terms describing interactions between the Higgs Boson and massive gauge Bosons and terms expressing Higgs self-interactions:

$$\begin{aligned}\mathcal{L}_{\mathcal{H}} = & -g\frac{m_h^2}{4m_W} h^3 - \frac{g^2}{32}\frac{m_h^2}{m_W^2} h^4 \\ & + 2\frac{h}{v}(m_W^2 W_\mu^+ W_\mu^- + \frac{1}{2} m_Z^2 Z_\mu^2) + \frac{h^2}{v^2}(m_W^2 W_\mu^+ W_\mu^- + \frac{1}{2} m_Z^2 Z_\mu^2).\end{aligned} \tag{2.13}$$

We can note in particular that the coupling strength of the Higgs Boson to the W Boson is proportional to m_W^2, and that the coupling strength of the Higgs Boson to the Z Boson is proportional to m_Z^2. The Higgs Boson *does not* couple to the massless photon or gluon.

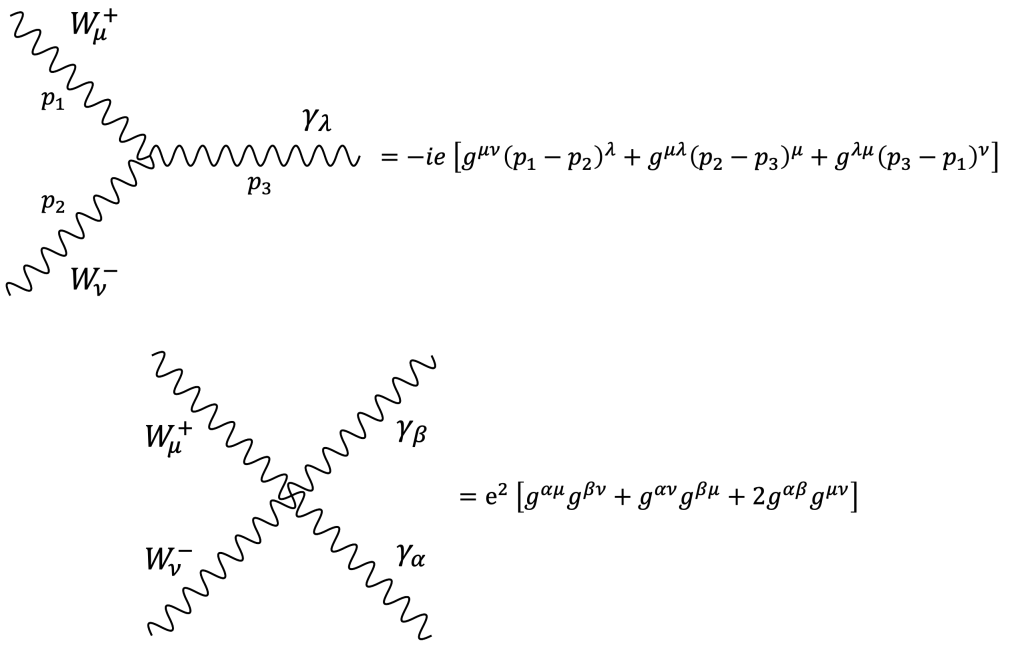


Figure 2.2: Feynman diagrams for 3-pronged γWW vertex (top) and the 4-pronged $\gamma\gamma WW$ vertex in the Electroweak sector of the Standard Model. Associated Feynman rule used in calculation of scattering amplitudes also provided.

Vertex	Strength
ZWW	$e \cot\theta_w$
γWW	e
$WWWW$	$e^2 \csc^2\theta_w$
$ZZWW$	$e^2 \cot^2\theta_w$
$\gamma\gamma WW$	e^2
γZWW	$e^2 \cot\theta_w$

Table 2.2: Coupling strengths of the gauge Boson interaction vertices in the Standard Model.

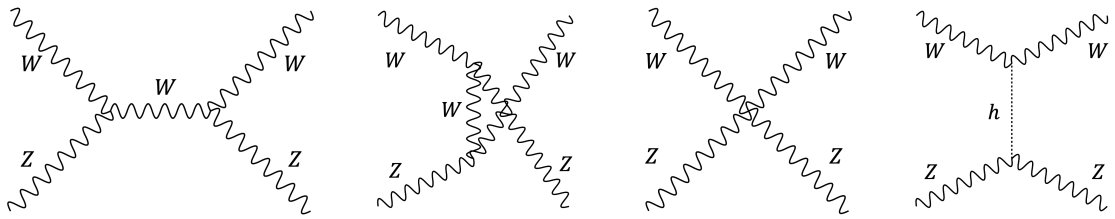


Figure 2.3: Feynman diagrams of the leading order processes contributing to $WZ \to WZ$ scattering. Without the fourth diagram, the cross section would diverge at high energy.

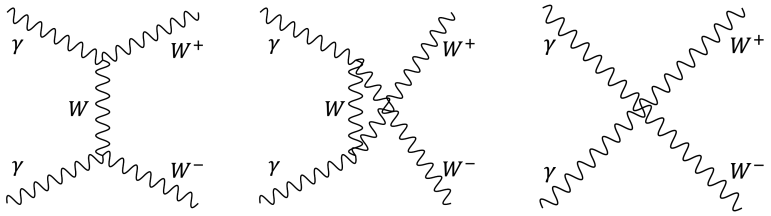


Figure 2.4: Feynman diagrams of the leading order processes contributing to $\gamma\gamma \to WW$ scattering.

The coupling of the Higgs Boson to the massive vector Bosons has important consequences in the SM. In a theory with massive vector Bosons but no Higgs Boson, cross sections for vector Boson scattering are proportional to $\frac{E^2}{m^2}$, where m is the Boson mass and E is the scattering energy. Since the mass is fixed, that would imply the unphysical consequence that the cross section should increase as a function of energy with no upper bound. The introduction of the Higgs Boson and its interaction with the massive Bosons precisely cancels this high-energy behavior, yielding physically meaningful results. For example, there are four leading order Feynman diagrams associated with the scattering process $WZ \to WZ$, as shown in Figure 2.3. The high-energy behavior of the fourth diagram exactly subtracts the high-energy behavior associated with the first three.

The divergence associated with massive Boson only scattering is not a concern for the $\gamma\gamma \to W^{\pm}W^{\mp}$ process that this thesis seeks to observe, as the photon is massless. The leading order Feynman diagrams associated with this process are shown in Figure 2.4. This process is unique in that it *only* involves Electroweak Boson vertices at leading order, making it a powerful probe of the gauge structure of the SM. In order to produce two on-shell W Bosons, both of which will have a mass of 80.4 GeV [2], the total energy of the two initial state photons must be at least 160.8 GeV.

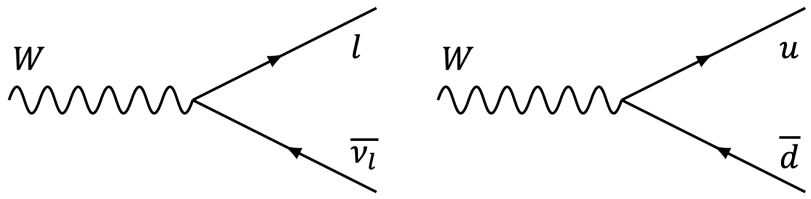


Figure 2.5: Feynman diagrams that dictate the decay of W Bosons. Decay can be into either a lepton and neutrino or into two quarks.

2.1.4 W Boson Decay

When an on-shell W Boson is produced, it decays quickly, with a half-life of 3×10^{-25} s [2]. The decay of a W Boson proceeds via one of the Feynman diagrams in Figure 2.5. There are two decay options: either a lepton and neutrino pair are created or a quark pair is created. If the decay is to a quark pair, one of the quarks must be an up-type quark, and the other must be a down-type quark. No decay may involve the top quark because its mass is higher than the W Boson's. The charge of the W Boson dictates the charge of the decay particles, as charge must be preserved. W^- Bosons decay to an $e + \bar{\nu}_e$ pair 10.9% of the time, a $\mu + \bar{\nu}_\mu$ pair 10.7% of the time, and a $\tau + \bar{\nu}_\tau$ pair 10.6% of the time. Decays to quark pairs occur the other 67.8% of the time.

2.1.5 Quantum Effects on Couplings

So far this discussion of the SM has involved only leading-order effects and Feynman diagrams with no loops. Tree diagrams essentially encapsulate classical physics; for example, performing calculations using the leading order diagrams for the interaction of the electron and the photon will give the same results as the Maxwell equations. Loops, however, are a purely quantum mechanical effect that have profound consequences.

The most important consequence in the context of this thesis is the running of coupling constants. When calculating the scattering of electrons and positrons in Quantum Electrodynamics one of the second order diagrams looks like that in Figure 2.6. The Bosons in the diagram are photons, but any Fermion that couples to the photon can run around the loop, so this diagram effectively corresponds to multiple diagrams, one for each charged Fermion type.

When these diagrams are added to the leading order diagrams, the results can be interpreted as causing the value of the electric charge[8] to change as a function of energy. If the theory involves only a photon and an electron, the functional relationship between electric

[8]Which is to say the strength of the coupling of charged particles and photons

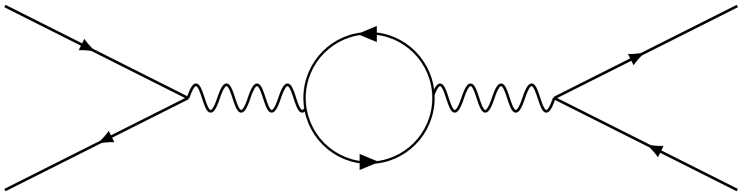


Figure 2.6: One of the second order diagrams that contribute to $2 \to 2$ Fermion scattering in Quantum Electrodynamics. The Bosons in the diagram are photons, and any Fermion that couples to the photon can appear in the loop.

charge and energy at high energy is

$$e^2_{\text{eff}}(Q) = \frac{e^2_R}{1 - \frac{e^2_R}{12\pi^2} \ln \frac{Q^2}{m^2}}. \tag{2.14}$$

Here Q is the energy of the interaction being studied, m is the electron mass, and e_R is just the electric charge, which has been *measured* at some energy value. A typical choice is $e_R = e_{\text{eff}}(m)$. Because the charge changes as a function of energy it is no longer just a value that is written into the equation; a measurement *must* be made at at least one energy, so that the value can be predicted for other energies. We should note here that as Q increases, the denominator of Eq. 2.14 gets smaller, meaning that e increases. The electric charge increases as a function of energy. Admittedly, this running is pretty subtle over the ranges of energies that are consequential for everyday life, but it is an experimentally confirmed effect. Based on this formula, when Q becomes very large, the electric charge diverges towards infinity. This is not actually a physical consequence, rather a breakdown in the perturbative framework used to derive Eq. 2.14[9]. Luckily though, e remains small enough for perturbation theory to work at energies that are currently experimentally accessible.

A dramatically different behavior occurs for the coupling of the strong force, which is associated with the $SU(3)_C$ gauge and g_c coupling constant. Gluons can self interact, so in addition to diagrams like Figure 2.6, there are second-order diagrams that involve gluon loops (which replace the Fermions in the middle with gluons) and gluon bubbles, which attach a gluon loop to the middle propagator using the 4-point interaction vertex. Physically, the consequence is that the strong coupling constant runs as

$$\alpha_s(\mu) = \frac{2\pi}{\beta_0} \frac{1}{\ln \frac{\mu}{\Lambda_{\text{QCD}}}}. \tag{2.15}$$

[9]In perturbation theory, processes are calculated by summing up diagrams that contain progressively more factors of e. When e is small, that means that only a few diagrams are important. As e becomes large, more and more diagrams become relevant, up to the point that every conceivable diagram must be considered to make an accurate calculation. Unfortunately, that's impossible, as there are infinitely many possible diagrams. See Section G.2.7 in Appendix G for more details

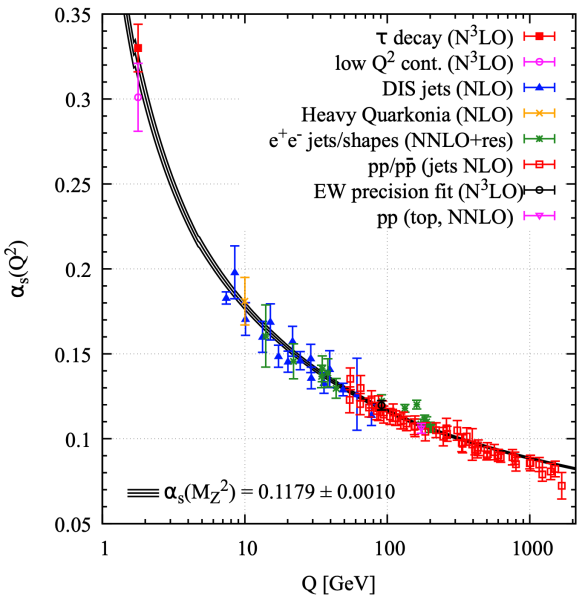


Figure 2.7: Running of the strong coupling constant α_s as a function of energy, Q [2].

Here, $\alpha_s = \frac{g_c^2}{4\pi}$ is used to make the equations simpler, μ is the observation energy, and $\beta_0 = 11 - \frac{2n_f}{3}$ (n_f is the number of quarks), which comes from group theory considerations. Λ_{QCD} is the energy at which $\alpha_s = 1$, or in other words, where perturbation theory breaks down, as every diagram would be of equal importance. As long as there are fewer than 17 quarks (which is the case in the SM), then the numerator in Eq. 2.15 is positive, meaning that the coupling actually decreases at higher energy and is stronger at lower energy. In fact, Λ_{QCD} is about 1 GeV, which is about the energy of the proton mass. The fact that perturbation theory breaks down at *low* energy is problematic in the sense that effects related to the strong force are largely incalculable at low energy. For example, while the strong force binds the quarks in a proton together and the protons and neutrons of a nucleus together, it is very difficult to calculate these binding energies. However, there are phenomenological models based on data observations that allow for approximations to be made[10]. The change in α_s is a real phenomenon, and the value of α_s has been measured in many experimental contexts, as demonstrated in Figure 2.7.

The running of the strong force coupling also means that it is impossible to observe "bare" quarks. When an object that has color charge, such as a quark, moves through space and time, the strong force pulls particle-antiparticle pairs into existence to create color-neutral objects. This process is known as "hadronization". Hadrons are color-neutral composite particles composed of quarks, bound together by the strong force. Baryons, such

[10]Such models are important to the simulations described in Chapter 5, and lead to the need for the corrections described in Chapter 8

as the proton and neutron, are a class of hadrons formed from three quarks, where each quark carries one of the three colors (combining all three colors makes a color-neutral sum). Mesons, such as pions and kaons, are formed by quark-antiquark pairs, where the color and anti-color cancel each other.

2.2 Proton-Proton and Photon-Photon Collisions

2.2.1 Cross Sections and Proton Distribution Functions

Some of the most important values calculated using the SM are scattering cross sections, which are normally denoted with σ. A cross section represents the probability that a particular process will occur. For example, we might want to calculate the probability that an electron and positron with 4-momenta p_e and p_p collide and create a $\mu^-\mu^+$ pair. That probability would be total cross section of this particular process, denoted as $\sigma_{e^-e^+\to\mu^-\mu^+}$. Such a calculation is useful if we only care about the creation of the muons, regardless of their kinematics. If we instead calculated the probability that the muon goes off in a particular direction with the anti-muon travelling in some other direction, that would be known as a "differential cross section". If we wanted to calculate the probabilty that the electron and positron interact at all, regardless of the final state, that would be called a "total cross section". Process-specific, differential, and total cross sections are all calculable using perturbation theory as long as the coupling value is sufficiently small at the scattering energy scale.

Cross sections in particle physics are conceptually related to the cross-sectional area of an object, with both carrying units of area (or distance^2). A larger value means a higher probability of interaction. The unit of cross section that will be used throughout this thesis is the "barn"[11] (denoted by b). The conversion between barns and m^2 is 1b = 10^{-28}m^2 = 100 fm^2, where a femtometer[12], fm, is 10^{-15} m. Total cross sections for a variety of scattering interactions are presented as a function of center-of-mass energy in Figure 2.8. Center-of-mass energy is the energy in a reference frame where the two incoming or scattering particle have equal and opposite-direction momentum, meaning that the sum of their 4-momenta is just a vector of the form $(m, 0, 0, 0)$, where m is the sum of the energies of the two particles. The energy sum, m, is also commonly denoted as $\sqrt{s}$, which comes from nomenclature used in Feynman diagram calculations. Figure 2.8 shows that the total cross section for proton-proton interactions at a center-of-mass energy of about 10,000 GeV – equivalently, 10 TeV – is about 100 mb. Similarly the total cross section for two photons to interact and produce hadrons is about 1 μb when $\sqrt{s} = 10$ TeV.

Scattering cross sections for particles such as photons, electrons, and positrons are all calculable from the SM, as these are fundamental particles. Historically, many experiments have accelerated electrons or positrons to high energies, and observed the resulting interac-

[11] The larger the side of the barn, the more likely you are to hit it after all.

[12] For context, the radius of the proton is just shy of 1 fm.

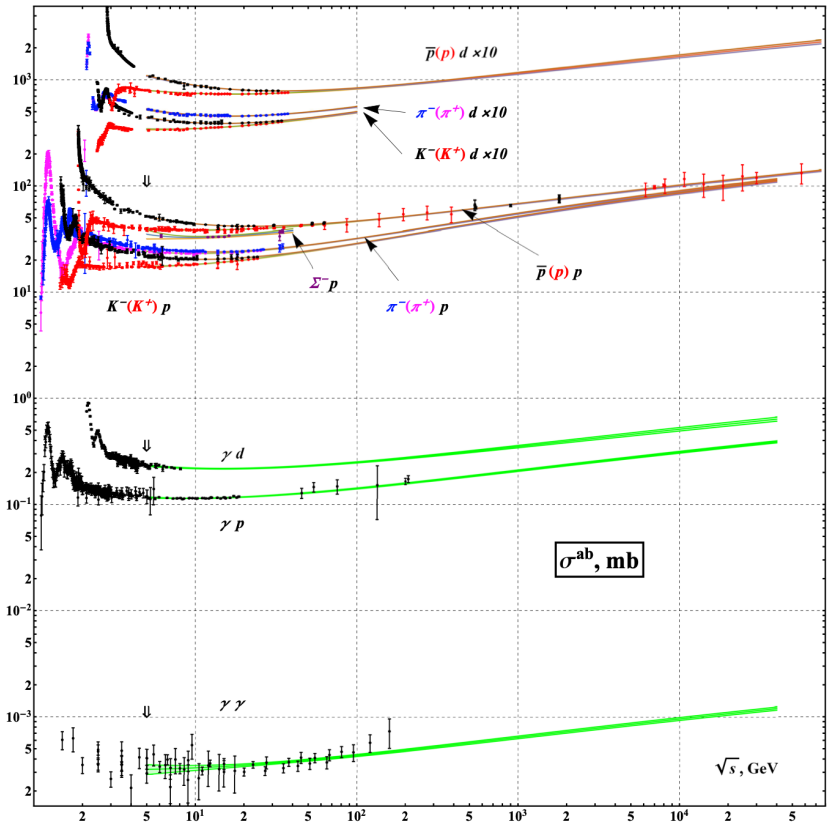


Figure 2.8: Total cross sections for various proton+hadron interactions, including proton-proton interactions, as a function of the center-of-mass energy of the scattering, $\sqrt{s}$. Cross sections proton+photon and photon+photon interactions that yield hadrons in the final state are also shown [2].

tions to see if they match SM expectations. This class of experiment is known as lepton collider experiments. One can also collide high energy protons or other heavier ions. These setups are called hadron colliders, and they can access higher energies than lepton colliders. However, protons are *not* fundamental and do not appear in the SM. It is actually very difficult to calculate total proton-proton scattering cross sections. The scattering of QCD composite particles is physics that emerges from the SM but is not nicely encapsulated in a clean, perturbative framework.

Fortunately, we are not totally out of luck when using proton colliders. When protons of sufficiently high energy collide, the collision can actually be thought of as a collision of the protons' constituent particles. This is because the coupling of the strong force becomes weaker at higher energies. Rather than behaving as a single strongly bound object, a proton behaves as a collection of loosely bound quarks and gluons when probed at sufficiently high

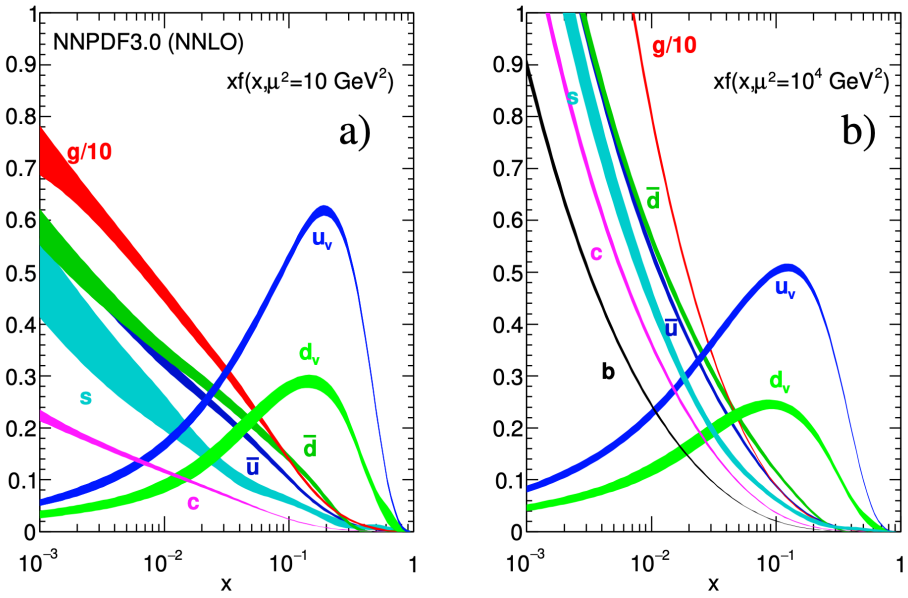


Figure 2.9: Parton Distribution Functions for two different energy scales. The x-axis in both plots is the Bjorken x, which represents a fraction of the proton's total momentum. The y-axis for both is $xf(x,\mu)$, where f is the PDF [2].

energy. While a proton is nominally a bound state of 3 quarks, that is only half of the story. Those 3 quarks are known as "valence quarks", but they are bound together by a sea of gluons and $q\bar{q}$ pairs. Collectively the valence quarks, gluons, and sea quarks are known as "partons". The partons which comprise a proton are fundamental particles that appear in the SM, so they can be used to calculate cross sections in a straightforward way. The trick is that we must know what fraction of the partons are, say, up-quarks or gluons at a given proton energy scale. Many experimentalists and theorists have done a great deal of work to measure and derive these values, called "Parton Distribution Functions" (PDFs). Example PDFs for two different proton energy scales are shown in Figure 2.9. The y-axis for both is $xf(x,\mu)$, where f is the PDF, x is the "Bjorken x", which represents a fraction of the proton's total momentum carried by the parton, and μ is the proton energy scale. Here, f is the PDF, which is a probability distribution function. The sum of all of the partons' momenta exactly equals the total momentum of the proton.

To calculate a cross section in a proton-proton collision, we must use the PDFs as

$$\sigma_{ab\to X} = \int\int dx_a\ dx_b\ f_a(x_a,\mu)\ f_b(x_b,\mu)\ \hat{\sigma}_{ab\to X}(p_a,p_b). \tag{2.16}$$

For example, we might be interested in the cross section for the process $u\bar{u} \to W^+W^-$. Here, a and b are the two types of partons that are colliding, u and $\bar{u}$. We must integrate over all

of the possible momenta of these partons within the proton, where $p_i = x_i P$ (P being the total momentum of the proton). These integrals are weighted by the PDFs. The $\hat{\sigma}$ is the parton-level cross section for given momentum of the partons, which is calculable from the SM.

We are typically interested in collisions involving high-energy partons. In the case of $u\bar{u} \to W^+W^-$, the two partons must have sufficient energy to actually produce a W^+W^- pair. When the partons carry such high energy, the strong coupling constant is small, and the partons behave essentially as free particles. When such a process occurs in a proton-proton collision, it is said to factorize into the hard-scatter and an underlying event (UE). The underlying event involves the interactions of the rest of the two protons. These interactions involve lower-energy partons and are not necessarily perturbative, but they can result in many particles (mostly pions and kaons) being emitted in random directions[13].

2.2.2 Equivalent Photon Approximation

The goal of this thesis is to observe the process $\gamma\gamma \to W^+W^-$. In order to produce two W Bosons, the colliding photons must have very high energy- far above what is achievable using a laser or synchrotron radiation. Unfortunately, neutral particles like photons cannot be accelerated using conventional means, so a high energy photon collider is difficult to build. However, in proton-proton collisions, we are able to directly study one type of high-energy neutral particle interaction, gluon-gluon collisions, as gluons carry non-negligible fractions of the proton momentum.

Photons are not partons, but photon-photon interactions *do* occur at hadron colliders. These collisions are understood through the Equivalent Photon Approximation[14] (EPA), which interprets the highly boosted Electric field around accelerated hadrons as a flux of photons [3, 4, 5].

Photons are particle interpretations of the propagation of the Electromagnetic field through free space. They travel at the speed of light in and have direction. It is important to note, however, that the Electric and Magnetic fields are transverse to the direction of photon motion as shown in Figure 2.10.

A proton at rest produces an electric field that points directly away from the proton, and there is no Magnetic field. This field can be interpreted as a somewhat complicated superposition of low-energy photons, but when the field is boosted, things simplify. If we embed the Electric and Magnetic fields in the Electromagnetic field strength tensor[15], then

[13] Underlying Events play a major role in this thesis, particularly as a discriminant between our $\gamma\gamma \to W^+W^-$ signal and various backgrounds. They will be discussed in greater detail in Chapter 8 and Part III

[14] Sometimes referred to as the Weizsäcker-Williams Approximation.

[15] See Eq. G.24 in Appendix G. This appendix also has a more detailed review of Electromagnetism and Relativity

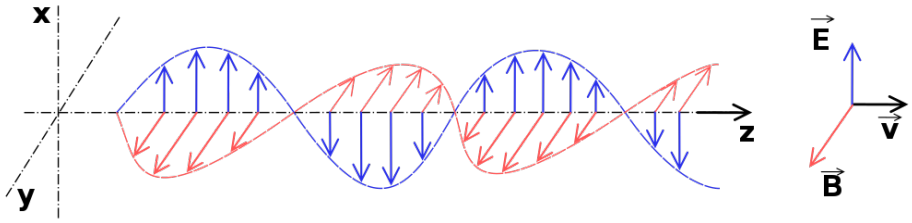


Figure 2.10: The Electric and Magnetic fields are transverse to the direction of photon propagation [6].

a Lorentz boost in the x-direction takes the form of

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & 0 & 0 \\ -E_y & 0 & 0 & 0 \\ -E_z & 0 & 0 & 0 \end{pmatrix} \rightarrow$$

$$F'_{\mu\nu} = \begin{pmatrix} \cosh\beta_x & \sinh\beta_x & 0 & 0 \\ \sinh\beta_x & \cosh\beta_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & 0 & 0 \\ -E_y & 0 & 0 & 0 \\ -E_z & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cosh\beta_x & \sinh\beta_x & 0 & 0 \\ \sinh\beta_x & \cosh\beta_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.17}$$

After performing the matrix multiplications, we find

$$\begin{aligned} |E'_x| &= |E_x| \\ |B'_x| &= 0 \\ |E'_\perp| &= |\cosh\beta_x \; E_\perp| \\ |B'_\perp| &= |\sinh\beta_x \; E_\perp|. \end{aligned} \tag{2.18}$$

Here the prime indicates the field strength in the boosted frame, and the $\perp$ symbol denotes the transverse direction, which in this case is the y-z plane. As the speed of the boosted frame relative to the stationary proton approaches the speed of light ($v \rightarrow 1$), $\cosh\beta_x$ becomes very large, and $\sinh\beta_x \rightarrow \cosh\beta_x$. In fact, at the Large Hadron Collider[16], the protons that are being collided travel at a speed of $v = 0.99999999c$, and they have $\cosh\beta_x = 7460$. This means that $|E'_x| \ll |E'_\perp|$, so $|E'_x|$ can effectively be ignored. We also have that $|E'_\perp| \approx |B'_\perp|$, which is what we would expect for the Electric and Magnetic fields of a photon. The boosting of the Electric field is visualized in Figure 2.11.

This approximation is used in a similar way to PDFs for parton collisions. The flux of photons as a function of photon frequency, ω, is given heuristically as [7]

$$n(\omega) = \frac{2}{\pi} Z_p^2 \; \alpha \; \ln \frac{\cosh\beta}{\omega R}. \tag{2.19}$$

[16]see Chapter 3

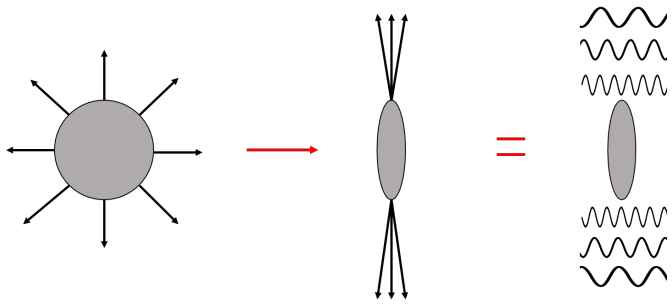


Figure 2.11: Visualization of the Equivalent Photon Approximation. The Electric field around a stationary proton (left), can be approximated as a flux of photons after a large boost (right). The compression of the photon is due to relativistic length contraction. Adapted from [3]

Here, $\alpha = \frac{1}{137}$, which is related to the strength of the electric charge. This equation is approximate, but valid for any accelerated nucleus[17], where Z_p is the number of protons in the nucleus, and R is the radius of the nucleus. In the case of a single proton, $Z_p = 1$ and $R \approx 0.8$ fm[18]. In truth, the photon flux depends strongly on the impact parameter, or distance from the proton.

These fluxes can be used to calculate simplified photon-photon interaction cross sections as [7]

$$\sigma_{\gamma\gamma\to X} = \int\int d\omega_1\ d\omega_2 \frac{n(\omega_1)}{\omega_1}\frac{n(\omega_2)}{\omega_2}\ \hat{\sigma}_{\gamma\gamma\to X}(\omega_1,\omega_2). \tag{2.20}$$

This equation is reminiscent of Eq. 2.16. To find the total cross section associated with a $\gamma\gamma \to X$ process, we have to integrate over all of the photon frequencies present. For each combination of two frequencies, there is some cross section $\hat{\sigma}_{\gamma\gamma\to X}(\omega_1,\omega_2)$ that can be calculated from the SM. For $\gamma\gamma \to W^+W^-$, that calculation would be done using the Feynman diagrams in Figure 2.4. A full calculation in EPA extends Eq. 2.19 with an integral over impact parameter and retains the impact parameter-dependence of the flux functions; thorough mathematical details of EPA can be found in Ref. [8] and Ref. [9].

2.2.3 Proton Dissociation

In EPA, the photons are radiated off of the proton as a coherent unit. However, because quarks can radiate gluons, there is a non-zero photon PDF [10, 11, 12]; a photon emitted by

[17] A proton is a Hydrogen nucleus.

[18] Photon-photon collisions can also be studied in heavy-ion colliders. Heavy ions bring an increase in the approximate photon flux thanks to the factor of Z_p^2 in Eq. 2.19; however, while more photons may collide, they typically have lower energy than would be seen in the photon-photon interactions at a proton collider. The smaller the ion R becomes, the higher the flux of high-frequency photons.

one proton might interact with a photon "inside" the other proton. In general, this photon PDF is known well thanks to electron-proton deep inelastic scattering results [13].

In the case that a high-energy photon is emitted or absorbed by parton, the proton can dissociate. This is to say that the proton becomes a spray of hadronic material as the proton's partons separate and hadronize. The proton's dissociation involves non-perturbative QCD, so it is difficult to calculate the expected spectra for the directions and momenta of the outgoing particles. However, the particles are expected to be forward[19], potentially out of the range of detection[20]. In the UE of a typical proton-proton interaction, there is a higher probability that hadronic material is emitted transverse to the motion of the original proton. It is not too hard to imagine why this is the case by considering momentum conservation. After a proton radiates a photon, it still has significant momentum in its original direction of motion. Conversely, a more typical proton-proton collision studied at the LHC is relatively head-on: lower-energy partons travelling in opposite directions can interact with each other, so the products of these interactions have less net momentum in the direction of either proton.

In general, proton dissociation can occur for one or both of the protons. When there is no dissociation, the interaction is labelled as *elastic*, as no energy is transferred into breaking up the binding of the proton. If only one of the protons dissociates, that collision is labelled as *single dissociative*, and if both protons dissociate it is *double dissociative*. These three types of collisions are visualized in Figure 2.12. Regardless of the protons' status after the emission of the photons, all of the interactions can be called *photon-induced*, and they all provide access to the $\gamma\gamma \to W^+W^-$ process. Because of the presence of outgoing protons or their decay products, at a proton collider this process can be expressed as $pp(\gamma\gamma) \to p^{(*)}W^+W^-p^{(*)}$, where $p^{(*)}$ means that the proton can remain intact or dissociate.

EPA only accounts for photons coherently radiated from protons, so to account for dissociative effects it is typical to define a cumulative photon PDF which is a sum of the photon "PDF" from EPA and the PDF resulting from quark-radiated photons. Ref. [8] provides mathematical details on how to perform cross-section calculations with this cumulative PDF, and furthermore how to compute cross-sections with rapidity[21] restrictions on quarks after photon radiation. This can be done by modifying the DGLAP equations [14, 15, 16] which govern the energy-scale evolution of PDFs.

2.2.4 Survival Factors

While EPA is a good way to approximate photon collisions at hadron colliders, it decouples the proton and the photon in a way that is not entirely realistic[22]. In EPA, the interacting photons are on-shell, but in reality they are virtual particles emitted by the protons or their

[19] A "forward" particle is one travelling with direction similar to the incoming particles

[20] See Chapter 3

[21] A measure of how transverse a particle is to the initial proton direction (see Chapter 3).

[22] Beyond the dissociative effects discussed in Section 2.2.3

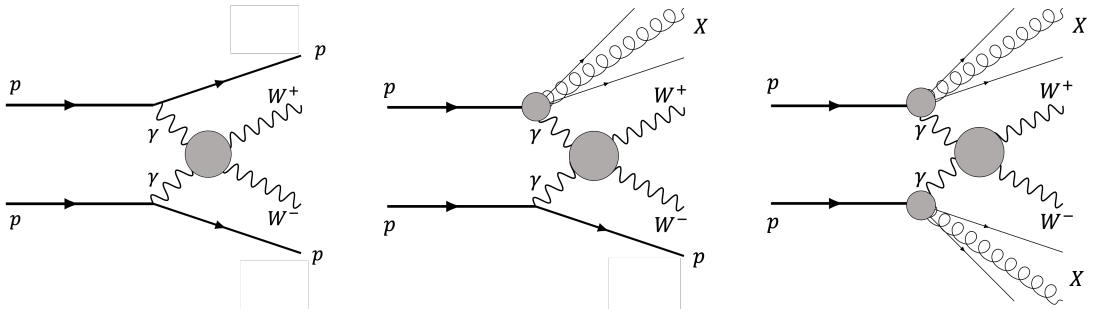


Figure 2.12: Illustrations of the three photon-induced processes that contribute to $\gamma\gamma \to W^+W^-$ production at a proton collider. From left to right, the illustrations represent elastic, single dissociative, and double dissociative production.

constituents. The use of EPA is justified when the energy of the colliding photons far exceeds their virtuality, which is the case for photon collisions at the LHC[23].

However, there are still corrections that can be made to improve the accuracy of predictions using this framework. In particular, we can note that in EPA, the protons that emit the photons are effectively ignored. While the protons may not exert a great deal of influence on the photon-photon collision[24], the protons may still interact *with each other*[25] [9, 8]. This important fact can influence the overall cross-sections computed with EPA and the phenomenology of photon-induced events.

The proton-proton multiple interactions (MI) in otherwise photon-induced events, depicted in Figure 2.13 with the skinny grey oval, primarily occur at low energy scales. These interactions are due to the strong force in an energy regime where QCD is strong and non-perturbative. While a precise calculation of this MI effect with perturbation theory is not possible, approaches have been developed to parameterize its effects [17, 18, 19, 8, 9]. The main effect of MI is encapsulated in a survival factor, S^2. The term "survival factor" refers to the survival of the "rapidity gap" between the two hadronic final states[26]. Explicitly, we are interested in how often there are no additional particles produced transverse to the initial direction of the protons due to proton MI, and S^2 is the fraction of photon-induced events that do not have central hadrons. S^2 depends on the specific color-neutral exchange between the two protons[27], but here of course we only focus on the case of photon exchange. It also depends on the impact parameter of the two protons, which is their physical separation. A large distance between the protons leads to less MI.

[23] quark-radiated photons tend to be more virtual, with higher relative transverse momentum.

[24] After the photons have been emitted, that is.

[25] Either before or after photon emission.

[26] Either the two final state protons, or their remains in the case of dissociation.

[27] In general, photon-induced processes at hadron colliders can be called "exclusive", where an exclusive event is typically defined as one where the primary interaction between the initial state hadrons does not involve any color flow.

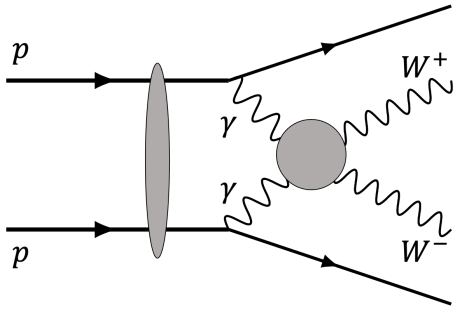


Figure 2.13: Multiple interactions of protons in a $\gamma\gamma \to W^+W^-$ process. The skinny grey region indicates low-momentum proton-proton interactions, while the circular region is a stand-in for the different vertices in the $\gamma\gamma \to W^+W^-$ process.

For photon-induced processes, S^2 must be calculated separately for the cases when both photons are emitted coherently from the protons, when one is a quark-radiated photon and one is coherently emitted, and when both are quark-radiated photons. Cases involving quark-radiated photons involve much smaller impact parameters, and therefore much smaller survival factors.

Two modern approaches to survival factor calculations can be found in Ref. [17] and Ref. [8, 9, 20]. Both methods encapsulate the survival factor in a suppression form factor that is integrated over proton-proton impact parameters, though the S^2 predictions of the first tend to be little lower. The first method predicts an S^2 of 0.7 for the case of photon-induced high-$m_{\ell\ell}$ di-muon production[28] and an S^2 of 0.65 for $\gamma\gamma \to W^+W^-$. The second method predicts an S^2 of 0.86 for the high-$m_{\ell\ell}$ di-muon case and 0.82 for $\gamma\gamma \to W^+W^-$. These predictions agree within uncertainty with a measurement of $\gamma\gamma \to l^+l^-$ performed by ATLAS [21].

2.3 Experimental Status of Photon-Induced Physics

The success of the SM can be summarized succinctly with Figures 2.14 and 2.15. These plots show that the SM correctly predicts cross sections that differ by over 10 orders of magnitude and span a wide range of phenomena, from vector Boson interactions to quark scattering. No other theory in any discipline has matched the precision and breadth of the SM.

2.3.1 Recent and Historical Measurements

The earliest published experimental results regarding photon-induced processes at accelerators date back to the 1990s. These were primarily observations of e^+e^- production in

[28] With $m_{\ell\ell} > 160$ GeV

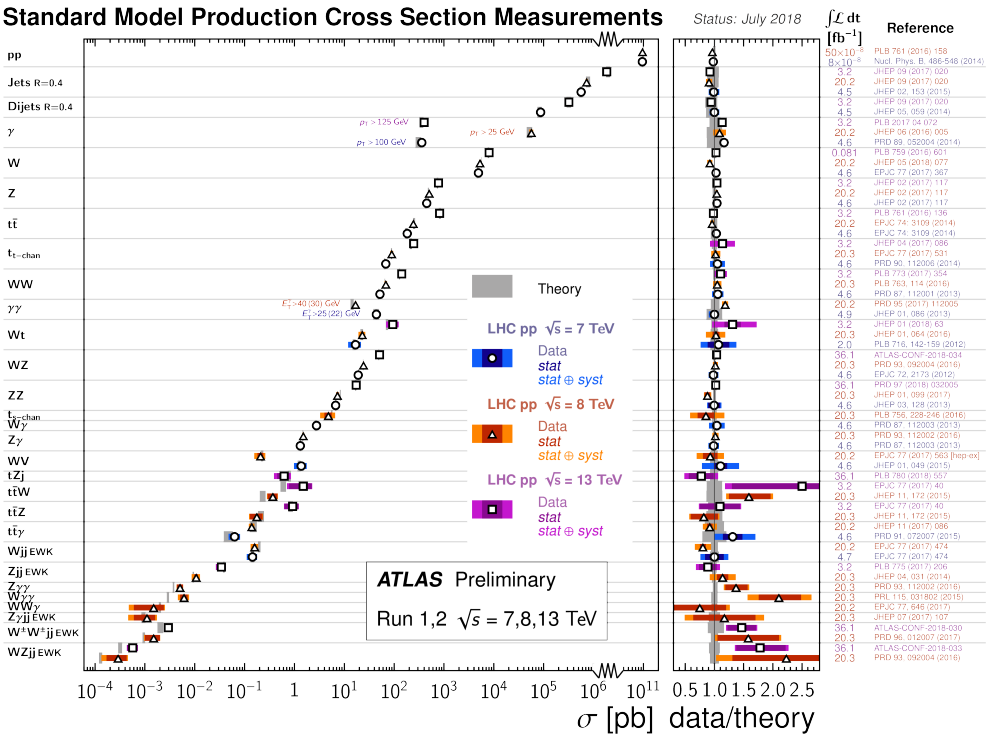


Figure 2.14: Summary of Standard Model cross section measurements performed by the ATLAS Collaboration using proton proton collisions at $\sqrt{s} = 7, 8, 13$ TeV [22].

ultra-peripheral heavy ion collisions, which is dominated by the $\gamma\gamma \to e^+e^-$ process. The leading order Feynman diagram for this process is shown in Figure 2.16. Heavy-ion collisions have enhanced photon fluxes relative to proton-proton collisions thanks to the factor of Z_p^2 in Eq. 2.19, and production of e^+e^- pairs requires the lowest center of mass energy of the two-photon initial state. Some early results are:

- The CERES/NA45 observed e^+e^- production with 10 MeV $< m_{\ell\ell} <$ 100 MeV when Sulfur at 200 GeV/u[29] collided with a stationary Platinum Target at the CERN Super Proton Synchrotron (SPS)[30] [23]. A separate group observed e^+e^- production when SPS Sulfur ions at the same energy collided with stationary Aluminum, Gold, and Palladium targets [24].

- At the Lawrence Berkeley National Lab Bevelac, Uranium ions at 0.96 GeV/u and Lanthanum ions at 0.405, 0.96, and 1.3 GeV/u produced e^+e^- pairs when colliding with Silver, Gold, and Copper targets [25].

[29] GeV per nucleon
[30] See Chapter 3

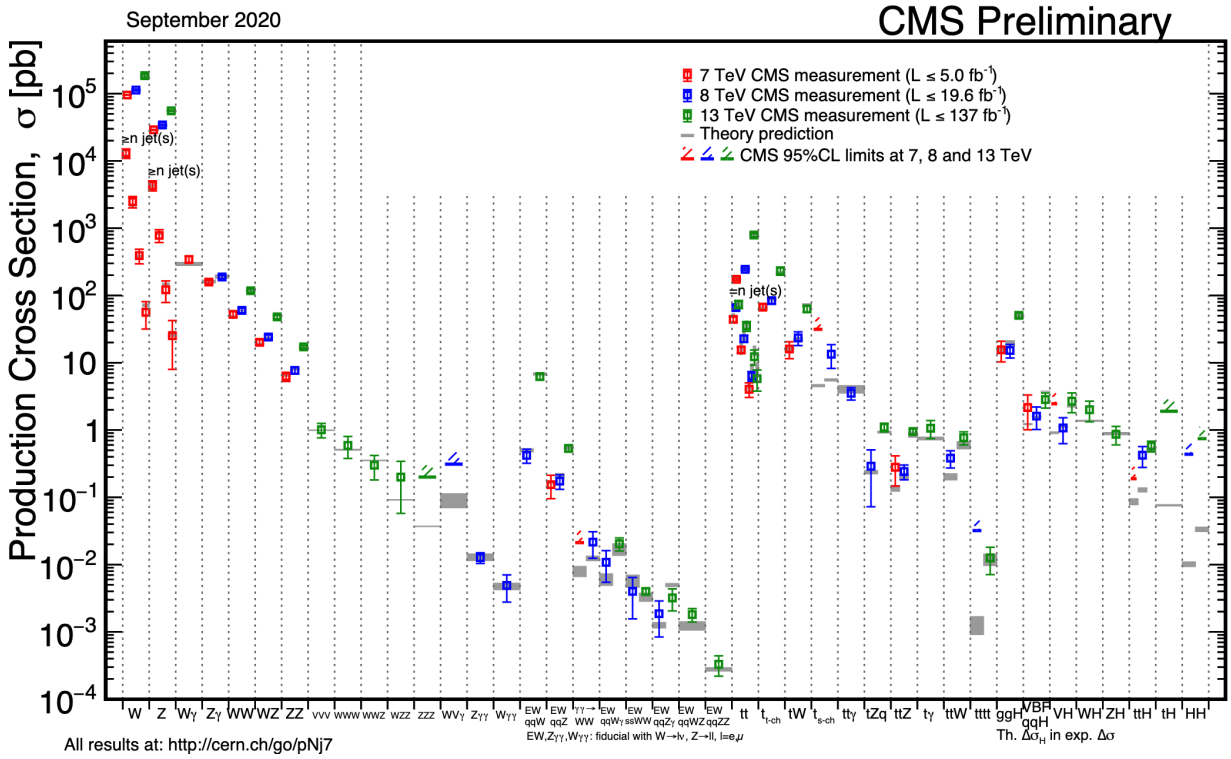


Figure 2.15: Summary of Standard Model cross section measurements performed by the CMSs Collaboration using proton proton collisions at $\sqrt{s} = 7, 8, 13$ TeV [22].

- The STAR Collaboration at the Relativisitic Heavy Ion Collider (RHIC) observed e^+e^- production in Gold-Gold collisions at a nucleon-nucleon pair center of mass energy of $\sqrt{s_{NN}} = 200$ GeV [26]. This experiment involves collisions of two accelerated beams of ions, rather than an accelerated beam and a stationary target.

There were also some results from electron-proton colliders in the early 2000s. For example, the H1 and ZEUS Collaborations at the DESY's HERA accelerator also observed photon-induced e^+e^- *and* $\mu^+\mu^-$ production [27, 28, 29]. These measurements were performed at center of mass energies of $\sqrt{s} = 301$ and 319 GeV.

The first photon-induced processes observed in hadron colliders were photon-induced e^+e^- and $\mu^+\mu^-$ production in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV at FermiLab's Tevatron [30, 31, 32]. The first observation, which was $\gamma\gamma \to e^+e^-$ in Ref. [30], was established with 16 events with $10 < m_{\ell\ell} < 40$ GeV, 2 of which were expected to be background.

Within the past 10 years, the ATLAS and CMS collaborations have produced many measurements of photon-induced phenomena, which have either matched SM predictions or

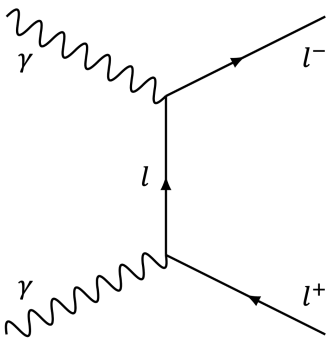


Figure 2.16: Feynman diagram depicting the process $\gamma\gamma \to l^+l^-$.

have been used to improve survival factor modeling[31].

Of course, a benchmark measurement is still photon-induced di-lepton production. This process has been observed by ATLAS in proton-proton collisions at $\sqrt{s} = 7$ TeV and 13 TeV [33, 34], and has even been observed with reconstructed final state protons[32] [21]. In Ref. [33], ATLAS found fiducial cross-sections of 0.43 ± 0.04 pb for the e^+e^- channel, and 0.63 ± 0.04 pb for the $\mu^+\mu^-$ channel[33]. In Ref. [34], the fiducial cross-section of the $\mu^+\mu^-$ was measured to be 3.12 ± 0.16 pb; part of the increase comes from difference in the leptons selections, and part is due to an increase in the cross-section with $\sqrt{s}$ – the cross-section scales as $\sigma \propto (\ln\sqrt{s})^3$. The photon-induced di-lepton process has also been observed in lead-lead collisions at a nucleon-nucleon center of mass energy of $\sqrt{s_{NN}} = 5.02$ TeV [35]. Similar measurements have been produced by CMS [36, 37].

Another impressive result is the observation of light-by-light scattering, or $\gamma\gamma \to \gamma\gamma$. This is a process that can only proceed via a quantum loop, with the dominant loop shown in Figure 2.17. This is a dramatic quantum effect in the SM, as light cannot interact with itself in classical Electromagnetism. This process was first observed by ATLAS is lead-lead collisions at a $\sqrt{s_{NN}} = 5.02$ TeV [38, 39]. In this context, an "observation" means a 5σ significance rejection of the hypothesis that light-by-light scattering does not occur. This means that we are at least 99.99997% sure that we have observed the process. A result with a significance of at least 3σ is called "evidence" for a process. ATLAS measured a fiducial cross-section for this process of 78 ± 15 nb. This indicates 99.7% confidence that a process has been observed. CMS has established evidence for light-by-light scattering, also in lead-lead collisions at a nucleon-nucleon center of mass energy of 5.02 TeV [40].

[31] The impact-dependent form factors mentioned in Section 2.2.4 rely on experimental input.

[32] See Section 3.2.6

[33] A fiducial cross-section is the cross-section after applying cuts on various observables, such as lepton kinematics. Different requirements were used to select electrons and muons, leading to some of the difference in the fiducial cross-sections

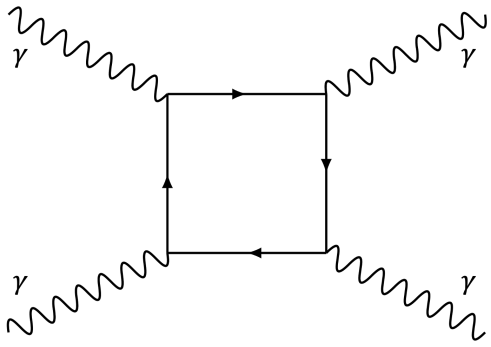


Figure 2.17: Feynman diagram depicting the process $\gamma\gamma \to \gamma\gamma$. The particles in the loop can be any charged Fermion.

Prior to the work presented in this thesis, evidence for $\gamma\gamma \to W^+W^-$ had been established by both ATLAS and CMS in proton-proton collisions at both 7 and 8 TeV [41, 42, 43]. In Ref. [41], ATLAS measured the fiducial cross-section to be 6.9 ± 2.6 fb, when requiring one W to decay to an electron and the other to decay to a muon. This thesis focuses on an analysis with ATLAS, using 13 TeV proton-proton collisions.

2.4 Beyond the Standard Model

While the SM accurately describes almost every particle interaction that has every been observed, there are phenomena that are not explained by the SM, suggesting that it is not complete or that it must be modified somehow. Some of these open problems are:

- Gravity is not explained by the SM. Currently, the best theory of gravity is Einstein's General Relativity, which is not quantum mechanical in any way. At extremely short distance scales and at high spacetime curvatures, this description of gravity is expected to be insufficient.

- Dark matter is not primarily composed of any of the particles in the SM. Despite this fact, Dark Matter accounts for about 85% of the mass in the universe, as deduced from various Astrophysical observations [44, 45]. New particles are needed to explain Dark Matter.

- The SM does not explain why the universe is primarily composed of matter rather than anti-matter. As the universe cooled after the Big Bang, some process or processes favored the existence of matter over anti-matter, but most of the interactions in the SM treat matter and anti-matter symmetrically [46]. There are some subtle SM

mechanisms that favor matter over anti-matter, but they are not sufficient to explain the predominance of matter in the current universe [47].

- It has been experimentally confirmed that neutrinos have mass [48], though it is not known how they acquire their mass. It could be through the introduction of right-handed neutrinos as discussed in Section 2.1.2, but it could be through other mechanisms [49].

There are also some theoretical open questions in the SM. While in some sense, these are primarily aesthetic issues rather than inconsistencies, two major examples are [50]:

- It is strange that the mass of the Higgs Boson is so "low" compared to the Planck scale, which is about 10^{19} GeV. The mass of a quantum field is related to its propagation through space and time, and this propagation is in turn sensitive to quantum effects, such as loops. If there is some quantum theory of gravity at high energy, then the quantum corrections to the mass of any fundamental scalar field, such as the Higgs, should be sensitive to this physics. The fact that the Higgs mass is *not* near the Planck mass suggests that some miraculous fine-tuning exists in the SM. This problem is referred to as the "hierarchy problem".

- Almost every possible term that respects the symmetries of the SM *is* in the SM. However, there is one term that could be added to the $SU(3)$ sector that would cause matter and anti-matter to be treated slightly differently. Any asymmetry that would result from this term has never been observed. It is unclear why this term is not present, or alternatively why it has a coefficient that is nearly 0. This is referred to as the "strong CP" problem.

Many "Beyond the Standard Model" (BSM) theories have been proposed to resolve one or more of these observational and theoretical issues. The following list barely scratches the surface:

- In "Supersymmetry" every Fermion in the SM gains a new Bosonic counterpart, and every Boson gains a Fermion partner [51]. This theory is designed to resolve the hierarchy problem and provides a dark matter candidate.

- If right-handed neutrinos exist, they can acquire a unique mass term in the SM Lagrangian, known as a "Majorana mass" [52]. This term can force the regular neutrino masses to be small, introduce a mechanism to generate the universe's matter-anti-matter asymmetry, and can provide a dark matter candidate.

- A new $U(1)$ symmetry can be introduced to the SM to resolve the strong CP problem [53]. This solution can also provide a dark matter candidate.

- String theory [54] and Quantum Loop Gravity [55] have both been proposed as potential quantum theories of gravity.

2.4.1 Effective Field Theory and the Importance of $\gamma\gamma \to W^+W^-$

Most BSM models have consequences that should be experimentally verifiable. If a new field is introduced, then the resulting particle should be able to be discovered directly. However, if this particle is too massive to be produced at current colliders or too difficult to detect directly, its presence can still have physical consequences due to quantum effects. These mostly manifest as modifications of SM predictions for the interactions of the known particles. A popular mathematical framework for "parameterizing our ignorance" and allowing for BSM effects in the SM is known as Effective Field Theory (EFT).

In the SM Lagrangian, terms such as $(F_{\alpha\beta})W^2$ or $F_{\alpha\mu}F_{\alpha\nu}W_\mu W_\nu$ weren't allowed due to gauge symmetry considerations. Such terms *are* Lorentz invariant, and can be added to the SM after the $SU(2)\times U(1)$ symmetry is broken by the Higgs mechanism. However, quantum corrections, similar those that make the Higgs mass sensitive to high-energy physics, actually *suppress* terms like these, which are said to have high mass-dimension. Every factor of a vector field, scalar field, or derivative adds 1 to the mass-dimension of a term, so the two examples given above have mass-dimension of 6. These terms would actually modify the $\gamma\gamma WW$ coupling predictions of the SM, but they are suppressed by a factor of $\frac{1}{\Lambda^2}$, where Λ is the energy scale at which the new physics exists. If new physics only exists at the Planck scale, then these terms would be highly suppressed, but new physics could exist at the order of 1-10 TeV.

In general, every term of high mass-dimension, d_i, has a unique coefficient, $\frac{f_i}{\Lambda^{d_i-4}}$, allowing the terms to contribute with different strengths. It is unknown *a priori* what f_i or Λ should be, but in the SM, $f_i = 0$. Lists of potential $d_i = 6$ and $d_i = 8$ terms that could affect the $\gamma\gamma \to W^+W^-$ are given in [56, 57, 58, 59]. Terms that affect the $\gamma\gamma W$ vertices in Figure 2.4 are said to induce anomalous trilinear gauge couplings, and terms that affect the $\gamma\gamma WW$ vertices induce anomalous quartic gauge couplings. The $\gamma\gamma \to W^+W^-$ scattering process is especially relevant in searching for (and placing constraints on) anomalous couplings [60], and studies have been performed in the past at both LEP [61] and the Tevatron [62]. Experiments at these colliders were not able to find the $\gamma\gamma \to W^+W^-$ process, suggesting that a large enhancement above the SM production rates does not occur. The LHC produces much more data though and at significantly higher energy, increasing sensitivity to anomolous couplings by more than an order of magnitude [63]. Both ATLAS and CMS have placed constraints using LHC proton-proton data at 7 and 8 TeV [41, 43, 42].

A widely used EFT Lagrangian is be expressed as [57]

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_{\dim-6} \frac{c_i}{\Lambda^2}\mathcal{O}_i + \sum_{j=1,2} \frac{f_{S,j}}{\Lambda^4}\mathcal{O}_{S,j} + \sum_{j=0,\ldots,9} \frac{f_{T,j}}{\Lambda^4}\mathcal{O}_{T,j} + \sum_{j=0,\ldots,7} \frac{f_{M,j}}{\Lambda^4}\mathcal{O}_{M,j}. \quad (2.21)$$

Here, the $\mathcal{O}_i$, $\mathcal{O}_{S,j}$, $\mathcal{O}_{T,j}$, and $\mathcal{O}_{M,j}$ operators refer to specific combinations of fields. For the dimension-8 operators, normally only combinations of the W, Z, B, and H fields are considered, but there are some dimension-6 terms that involve Fermions. Analyses that interpret their results in light of this framework often place constraints on the c/Λ^2 or f/Λ^4 coefficients.

The $\gamma\gamma \to W^+W^-$ process is not the only one that can be used to constrain these factors [64, 58]. Its relevance to the dimension-6 and $\mathcal{O}_{S,j}$ operators is generally considered marginal, with analyses such as generic Vector Boson Scattering with jets, and QCD-induced WW, WZ, or ZZ production placing stronger limits. However, $\gamma\gamma \to W^+W^-$ is considered more relevant for the dimension-8 $\mathcal{O}_{T,j}$ and $\mathcal{O}_{M,j}$ operators, which are

$$\begin{aligned}
\mathcal{O}_{T,0} &= \mathrm{Tr}[W_{\mu,\nu}W^{\mu,\nu}] \times \mathrm{Tr}[W_{\alpha,\beta}W^{\alpha,\beta}] \\
\mathcal{O}_{T,1} &= \mathrm{Tr}[W_{\alpha,\nu}W^{\mu,\beta}] \times \mathrm{Tr}[W_{\mu,\beta}W^{\alpha,\nu}] \\
\mathcal{O}_{T,2} &= \mathrm{Tr}[W_{\alpha,\mu}W^{\mu,\beta}] \times \mathrm{Tr}[W_{\beta,\nu}W^{\nu,\alpha}] \\
\mathcal{O}_{T,5} &= \mathrm{Tr}[W_{\mu,\nu}W^{\mu,\nu}] \times B_{\alpha,\beta}B^{\alpha,\beta} \\
\mathcal{O}_{T,6} &= \mathrm{Tr}[W_{\alpha,\nu}W^{\mu,\beta}] \times B_{\mu,\beta}B^{\alpha,\nu} \\
\mathcal{O}_{T,7} &= \mathrm{Tr}[W_{\alpha,\mu}W^{\mu,\beta}] \times B_{\beta,\nu}B^{\nu,\alpha} \\
\mathcal{O}_{T,8} &= B_{\mu,\nu}B^{\mu,\nu} \times B_{\alpha,\beta}B^{\alpha,\beta} \\
\mathcal{O}_{T,9} &= B_{\alpha,\mu}B^{\mu,\beta} \times B_{\beta,\nu}B^{\nu,\alpha}
\end{aligned} \tag{2.22}$$

and

$$\begin{aligned}
\mathcal{O}_{M,0} &= \mathrm{Tr}[W_{\mu,\nu}W^{\mu,\nu}] \times \left[(D_\beta H)^\dagger D^\beta H\right] \\
\mathcal{O}_{M,1} &= \mathrm{Tr}[W_{\mu,\nu}W^{\nu,\beta}] \times \left[(D_\beta H)^\dagger D^\mu H\right] \\
\mathcal{O}_{M,2} &= [B_{\mu,\nu}B^{\mu,\nu}] \times \left[(D_\beta H)^\dagger D^\beta H\right] \\
\mathcal{O}_{M,3} &= [B_{\mu,\nu}B^{\nu,\beta}] \times \left[(D_\beta H)^\dagger D^\mu H\right] \\
\mathcal{O}_{M,4} &= \left[(D_\mu H)^\dagger W_{\beta,\nu} D^\mu H\right] \times B^{\beta,\nu} \\
\mathcal{O}_{M,5} &= \left[(D_\mu H)^\dagger W_{\beta,\nu} D^\nu H\right] \times B^{\beta,\mu} \\
\mathcal{O}_{M,6} &= \left[(D_\mu H)^\dagger W_{\beta,\nu} W^{\beta,\nu} D^\mu H\right] \\
\mathcal{O}_{M,7} &= \left[(D_\mu H)^\dagger W_{\beta,\nu} W^{\beta,\mu} D^\nu H\right].
\end{aligned} \tag{2.23}$$

In particular, initial projections indicate that the $\gamma\gamma \to W^+W^-$ analysis should be able to place leading constraints on the $\mathcal{O}_{T,0}$, $\mathcal{O}_{T,5}$, $\mathcal{O}_{T,6}$, and $\mathcal{O}_{T,7}$ operators and the $\mathcal{O}_{M,0}$, $\mathcal{O}_{M,1}$, $\mathcal{O}_{M,2}$, $\mathcal{O}_{M,3}$, $\mathcal{O}_{M,4}$, and $\mathcal{O}_{M,5}$ operators.

Another experimental handle that allows for stronger limit placement is the fact that these operators do not only cause potential enhancements to (or deficits in) the overall cross-section. For example, they can cause enhancements in certain kinematic regimes, such as at higher $m_{\gamma\gamma}$, or distort spectra, such as the expected momentum of the W Bosons [41, 43, 42]. Typically, $\gamma\gamma \to W^+W^-$ analyses look for leptonic decays of the W Bosons, so differential cross-section distributions in quantities such as $m_{\ell\ell}$ or $p_{\mathrm{T},\ell\ell}$ can be made. A differential distribution shows the cross section as a function of some kinematic variable, enabling comparisons to the spectrum shape predicted by the SM or by EFT-enhanced models.

Chapter 3

The Large Hadron Collider and the ATLAS Detector

This chapter summarizes the experimental underpinnings of this thesis, from the acceleration of charged particles at the Large Hadron Collider (LHC) to the various components of the ATLAS detector.

3.1 Synchrotrons and the Large Hadron Collider

3.1.1 Particle Accelerators

As discussed in Section 2.2.2, high energy photon collisions are experimentally accessible at lepton or hadron colliders: energetic, slightly-virtual photons emitted by the high-energy charged particles can interact. While photons cannot be accelerated by conventional means, charged particles *can* be with simple Electromagnetism. Electric fields exert forces on charged particles, causing acceleration. In fact the electron-volt (eV) unit is defined as the amount of energy an electron would gain if accelerated over a potential difference of 1 V in vacuum (remember that a Electric field is created by potential difference). The greater the potential difference, the greater the energy a charged particle would gain.

While establishing a potential difference in the TeV range is practically impossible, accelerators can push protons to energies in the 1-10 TeV range through successive accelerating "kicks". The first type of accelerator to realize this concept was the "cyclotron", which was incidentally developed at Berkeley [65]. An additional insight in the development of this type of accelerator is that if the charged particles are travelling in a circle, then they can be kicked many times using a small number of accelerating units. The "cyclo" in the cyclotron's name comes from this circular motion principle. Cyclotrons use a radiofrequency (RF) alternating current source to accelerate particles over a gap between two semicircular bending regions, where the alternating orientation of the field kicks the charged particles in the correct direction as they pass between the semicircles. They use a fixed magnetic field to

bend the particles in a circle, and as the speed of the particle increases, its radius of motion increases. The maximum accessible energy with a cyclotron is limited by its radius and by relativistic effects. As a particle approaches the speed of light, the frequency of its passes between the hemispheres becomes out of phase with the RF source driving it.

The prevailing type of accelerator for modern high energy physics is the synchrotron [66, 67]. The synchrotron overcomes the cyclotron's limitations with two key insights. First, synchrotrons are fixed-radius accelerators. Where the cyclotron's magnets had fixed field strength, the bending magnetic field for a synchrotron is adjusted as the particles speed up, causing them to bend with a set radius. Additionally, the RF sources that drive the particles have adjustable frequency, allowing the kicks to be carefully synchronized with the particles' arrival times. This synchronization is where the synchrotron derives its name. In principle, the synchrotron can accelerate particles to arbitrarily high energies, but this is not achievable in reality. Charged particles will lose energy through synchrotron radiation when they are accelerated by the bending magnets, and there are limits to how fast RF fields can be made to oscillate. Another limiting factor is the achievable magnetic field strength for bending particles. Increasingly high fields are needed to hold increasingly energetic particles at the same radius of curvature. Most magnet setups can only exert a limited range of field strengths, so acceleration is often performed in multiple stages, where particles are injected into rings of increasing radius as they gain energy.

3.1.2 The Large Hadron Collider

The world's largest and highest energy collider, the Large Hadron Collider (LHC) [68], is located at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland. The LHC sits in a 27 km circumference tunnel that is about 100 m underground, passing beneath the Franco-Swiss border. This tunnel was previously used by the Large Electron-Positron (LEP) collider [69], which, as the name suggests, was an electron-positron collider. An illustration of the complete accelerator complex is given in Figure 3.1. The protons that will eventually be sent into the LHC are initially stored in a bottle of Hydrogen gas. The Hydrogen is ionized, yielding protons that are passed through a linear accelerator, the Linac2, and the circular Proton Synchrotron (PS) Booster, reaching energies of 50 MeV and 1.4 GeV after the two steps respectively [70]. From the booster, the protons are accelerated to 25 GeV in the PS and are injected into the Super Proton Synchrotron (SPS). The SPS was final accelerator in the CERN complex in the 1980s, accelerating protons and anti-protons to an energy around 250 GeV; it was used to discover the W and Z bosons [71, 72, 73, 74].

After being accelerated to 450 GeV by the SPS, the protons are injected into the LHC. In Run 1 of the LHC, which was in 2009-2012, protons were accelerated to either 3.5 TeV or 8 TeV, depending on the year, yielding a center of mass collision energy of $\sqrt{s} = 7$ or 8 TeV. In Run 2 of the LHC, which was from 2015-2018, the protons were brought to 6.5 TeV, for a collision energy of $\sqrt{s} = 13$ TeV. After reaching their maximum energy, the protons circulate in the LHC for several hours.

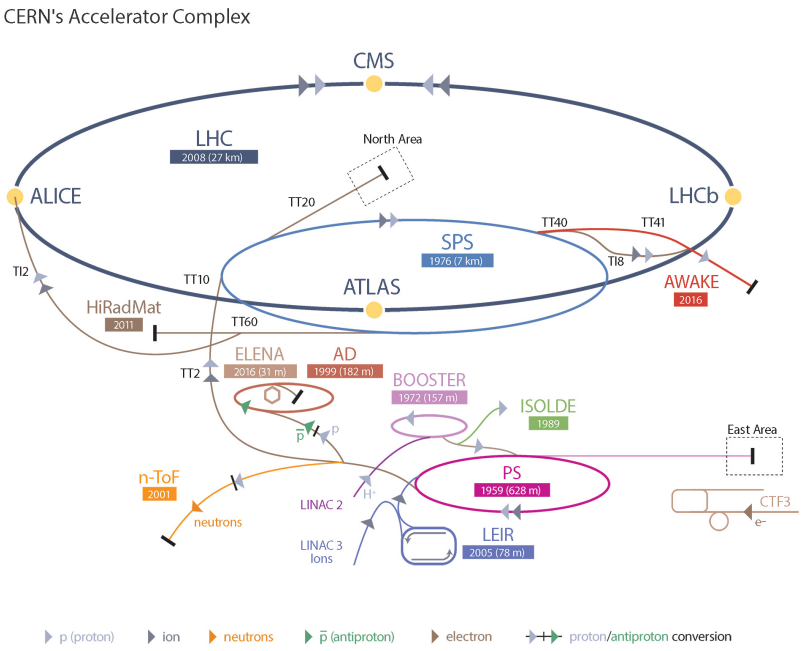


Figure 3.1: Map of the CERN accelerator complex. A proton starts its journey through the accelerator in the linear accelerator labelled LINAC2. It then is sent into the BOOSTER and then the Proton Synchrotron (PS). From the PS it is sent into the Super Proton Synchrotron (SPS), and from the SPS it is sent into the LHC. The proton's energy increases at each subsequent step. [75].

The LHC "ring" itself is in truth comprised of eight arcs and eight "insertions" [70]. The different insertions are where protons are injected into the LHC, extracted from the LHC, accelerated, or collided. There are two proton beams in the LHC, running in opposite directions; for most of the LHC's length, there are two beampipes that keep the beams separated by 192 mm. The beampipes are kept at ultrahigh vacuum to prevent beam protons from interacting with gas in the pipes. At "interaction points", the two beams are passed through each other in a single beampipe. Four experiments are built around the interaction points: ATLAS [76], CMS [77], LHCb [78], and ALICE [79].

Sixteen total RF cavities are used to accelerate the beams (eight cavities per beam), where each cavity delivers 2 MV at 400 MHz [70]. The RF nature of the cavities is key to the actual structure of the LHC beams, which is composed of circulating *bunches* of about 1.2×10^{11} protons each. The RF cavities create an oscillating electric field, and as protons circulate, those on the front end of a bunch receive a smaller kick than protons in the back of the bunch, thus preserving the bunch structure of the beam. At a maximum fill, there are 2808 bunches circulating in each beam of the LHC at any given moment. At interaction

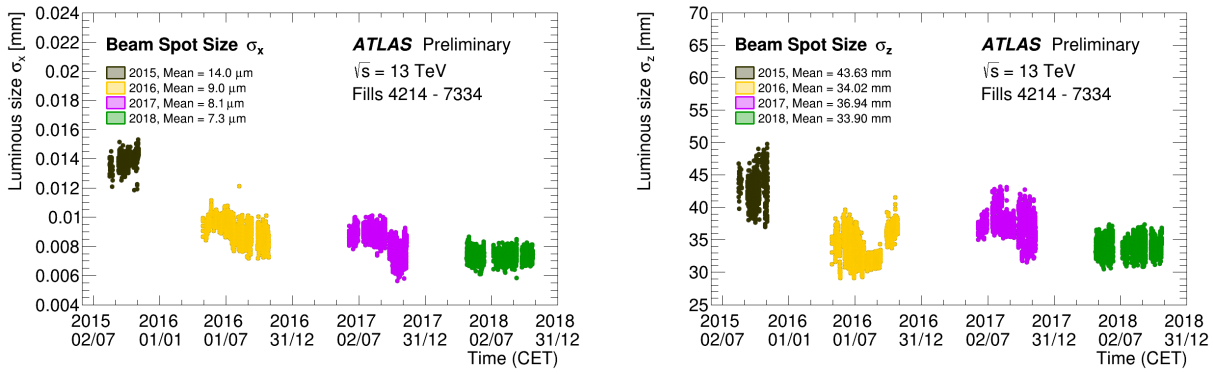


Figure 3.2: Beamspot size information for Run 2 of the LHC as recorded by ATLAS [80]. The left plot shows the typical transverse size of the beamspot, and the right plot shows the typical longitudinal size, where in both directions, the density of interactions is modelled as a Gaussian distribution.

points, bunches pass through each other once every 25 ns (assuming each bunch is filled for both beams). For some perspective, the protons in the LHC travel at about 99.999999% of the speed of light, and circulate the LHC 11,245 times per second.

A system of magnets keep the beams rotating in the beampipes and focused [70]. Superconducting dipole magnets are built around the pipes in the arc sections of the LHC, which bend the beams. Series of quadrupole magnets are used to focus the beams, which naturally disperse due to electromagnetic repulsion between the protons that make up the bunches.

When the beams pass through each other in the interaction points, the region where the bunches overlap is called the "Luminous Region" or "beamspot". This is the physical location where proton-proton collisions actually occur. ATLAS monitors the size and shape of the beamspot, and this information for LHC Run 2 is shown in Figure 3.2. The left-hand plot shows the transverse size of the beam as a function of time, and the right-hand plot shows the longitudinal size. In the transverse direction, the interactions were typically contained in a roughly circular region with a diameter of about 10 μm, but occurred in a region with a length of about 42 mm in 2015 and 35 mm in 2016-2018 [80]. This was determined by fitting a Gaussian function to the z distribution of reconstructed vertices [81] (see Section 4.1). For context, that is bit narrower than a typical human hair, though it may be about as long as a hair, depending on the barber.

3.1.3 Luminosity

To predict the number of times a particular process, such as $\gamma\gamma \to W^+W^-$, will occur, we must know both the cross section for the process (see Sec. 2.2) and the "luminosity" delivered by the LHC. The luminosity is primarily a function the number of protons delivered to each

experiment per second, the beam size, and the overlap of the beams. It can be functionally expressed as [2]

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y}. \tag{3.1}$$

Here, f is the frequency with which a bunch circulates the LHC: 11,245 Hz. There are two bunches passing through each other, and n_1 and n_2 are the numbers of protons per bunch. The σ_x and σ_y values are the transverse root mean squared sizes of the beam, assuming that both beams have identical shape. Equation 3.1 additionally assumes that the center of the beams overlap; if the beams are off-center the equation must be modified such that the denominator encapsulates the overlapping area of the beams. Heuristically, the units of luminosity carry a factor of $1/area$ and $1/time$. Instantaneous luminosity is often cited in units of $\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

The total number of times a process occurs is given by the formula

$$N_p = \sigma_p \int \mathcal{L}(t)dt. \tag{3.2}$$

Here, $\mathcal{L}$ is the instantaneous luminosity, which can change as a function of time, N_p is the total number of times a process is predicted to occur, and σ_p is the process's cross section. The time integral is performed over the period for which we are trying to predict N_p. The quantity $\int \mathcal{L}(t)dt$ is often called the "integrated luminosity", and it is often expressed in units of b^{-1} (or nb^{-1}, pb^{-1}, or fb^{-1} as the case may be). This makes calculating N_p easier, given that σ_p is often expressed in barns. To be explicit, 1 fb^{-1} corresponds to more pp collisions than 1 pb^{-1}, and 1 pb^{-1} is more than 1 nb^{-1}, etc. To observe processes with smaller and smaller cross sections, you need more and more integrated luminosity.

The integrated luminosity as a function of time delivered to and recorded by the ATLAS experiment over the course of LHC Run 2 is shown in Figure 3.3 [82]. The integrated luminosity of data considered good for physics analysis is also shown [83]. This is the data that will be used in this thesis, corresponding to 139 fb^{-1} of proton-proton data. This value was determined using the LUCID-2 detector [84] with the algorithm described in Ref. [85]. More information on luminosity measurements is given in Appendix D, which also describes a luminosity algorithm I developed based on Pixel Cluster Counting.

3.1.4 Pileup

Despite the fact that there are about 1.2×10^{11} protons in each bunch, the bunches are mostly empty space. Almost all of the protons in the opposite-moving bunches in the interaction point simply pass by each other without interacting. However, *some* of the protons must interact for there to be data. The probability distribution for the mean number of interactions per bunch crossing as recorded by ATLAS for LHC run 2 is shown in Figure 3.4 [82]. The distributions are normalized to integrated luminosity and are shown for Run 2 in its entirety and for each year separately. The fewest numbers of collisions per crossing were seen in 2015, with the highest rate of collisions seen in 2017 and 2018. On average, 33.7 interactions

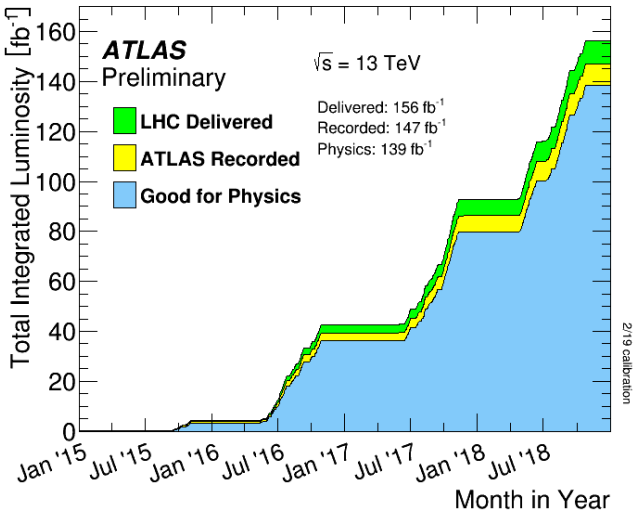


Figure 3.3: Integrated luminosity delivered to and recorded by the ATLAS experiment as a function of time in LHC Run 2 [82]. The integrated luminosity of data considered good for physics analysis is shown in blue, totalling 139 fb^{-1} over all of 2015-2018.

per crossing occurred in Run 2, with a range extending down to 10 and up to 70. The definitions of luminosity and integrated luminosity already account for the possibility of multiple interactions per crossing, so no modifications are needed.

We will find that the density of interactions is a very important factor in the search for $\gamma\gamma \to W^+W^-$. Figure 3.5, shows the typical interaction density in the middle of the interaction point as a function of the number of interactions per bunch crossing for a few months of 2017 data-taking [80]. For this time period, the most data was collected in beam conditions with 60 interactions per crossing, and a density of about 0.75 interactions per mm.

3.2 ATLAS Detector

Having learned how particles interact in Chapter 2 and how to collide high energy protons in Section 3.1, the last major piece of the puzzle is learning how to *observe* proton collisions. There is no microscope that is strong enough to observe fundamental particles, and even if you did stand near the interaction point with a lens of some sort, you wouldn't last long due to the radiation. Somewhat paradoxically, large, multi-component detectors must be used to measure fundamental particles precisely. One such detector, the ATLAS detector [76], was used to collect the data used in this thesis. This section presents the design details of the detector and how the components are used to detect particles.

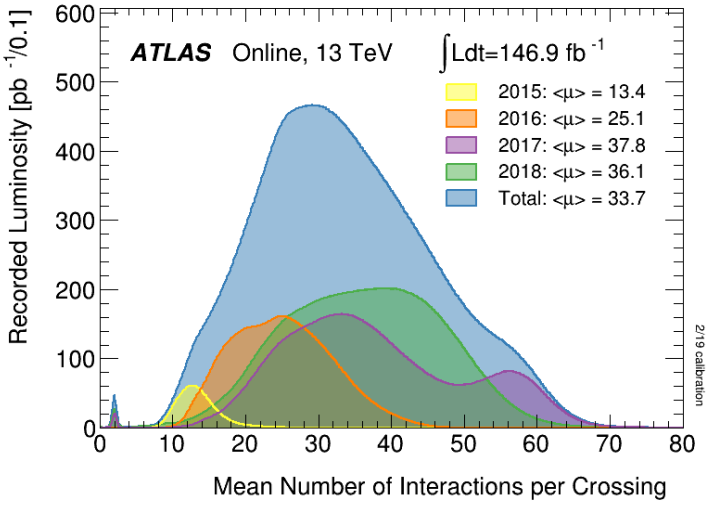


Figure 3.4: Probability distribution functions of the mean number of proton-proton interactions per bunch crossing as recorded by ATLAS in LHC Run 2 normalized to integrated luminosity, both cumulatively and for each year individually [82].

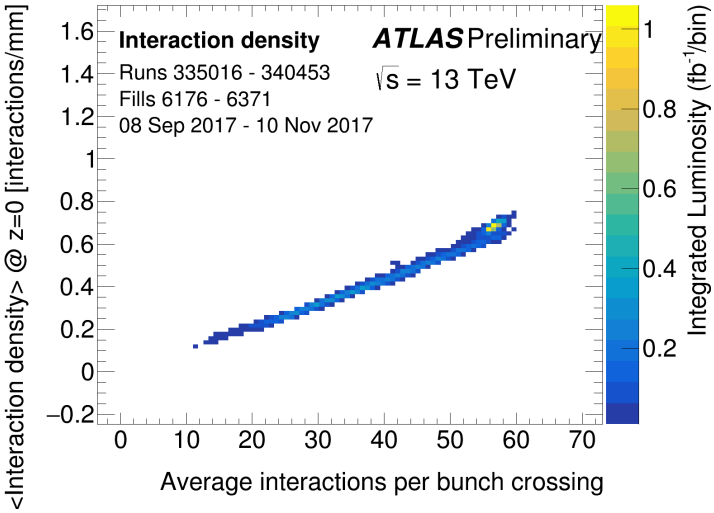


Figure 3.5: Typical interaction density in the middle of the interaction point as a function of the number of interactions per bunch crossing for a few months of 2017 data-taking [80].

3.2.1 Overview

The ATLAS detector, pictured in Figure 3.6 [86], is a cylindrical multi-purpose detector built around one of the LHC interaction points. It is about 44m long and 25m high, providing nearly 4π coverage around the proton-proton collisions. To mathematically refer to the positions and trajectories of particles, a right-handed Cartesian coordinate system is defined such that the positive x-axis points towards the inside of the LHC ring, the y-axis points up, and the z-axis is along the beamline, in the direction of oncoming protons. The (0, 0, 0) position is in the center of the detector, which is symmetric in the positive-z and negative-z directions. One can also use a polar coordinate system based on the Cartesian one, where the ϕ angle is the inclination from the x-axis, spinning counterclockwise around the z-axis from 0 to 2π, and r is the perpendicular distance from the z-axis. A θ angle is often used, which is the inclination from the positive-z axis, extending from 0 to π.

There are two magnet systems in ATLAS [76]: a solenoid magnet provides a 2 T field that points in the z-direction for the innermost section of the detector, and a toroidal magnet system creates a magnetic field for the outermost sections of the detector. The magnetic fields allow for the measurement of charged particles' momenta using the formula $p_{\mathrm{T}} = qBR$, where p_{T} is the momentum of the particle transverse to the magnetic field, q is the charge of the particles, B is the strength of the magnetic field, and R is the bending radius of the particle. In the context of ATLAS and similar experiments, this relationship can be rewritten as $p_{\mathrm{T}} = 0.3BR$, as long as the momentum is in GeV, the particle's charge is the same as that of an electron, the magnetic field is in T, and the radius is in m.

Because the magnetic field due to the solenoid is in the z-direction and therefore parallel to the beamline, the component of a particle's momentum transverse to the beamline is an important quantity:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}.$$

A last important quantity used in describing position and trajectory in ATLAS is "pseudorapidity". Pseudorapidity, or η, is a transformation of the angle θ:

$$\eta = -\ln\Big[\tan\Big(\frac{\theta}{2}\Big)\Big]. \tag{3.3}$$

For massless particles, pseudorapidity is equivalent to "rapidity", which can be defined as

$$y = \frac{1}{2}\ln\Big(\frac{E + p_z}{E - p_z}\Big).$$

For massive particles, pseudorapidity is a good approximation of rapidity, especially when the particle's energy is large compared to its mass. The introduction of this rapidity might seem arbitrary, but it is related to the rapidity that was discussed in the context of relativity (see e.g. Eq. G.10). Differences in rapidity are actually Lorentz invariant. This is important because while collisions in the LHC occur in the center-of-mass frame of the incoming protons, the fundamental particles that collide (the quarks, gluons, or in our case, photons) *don't*

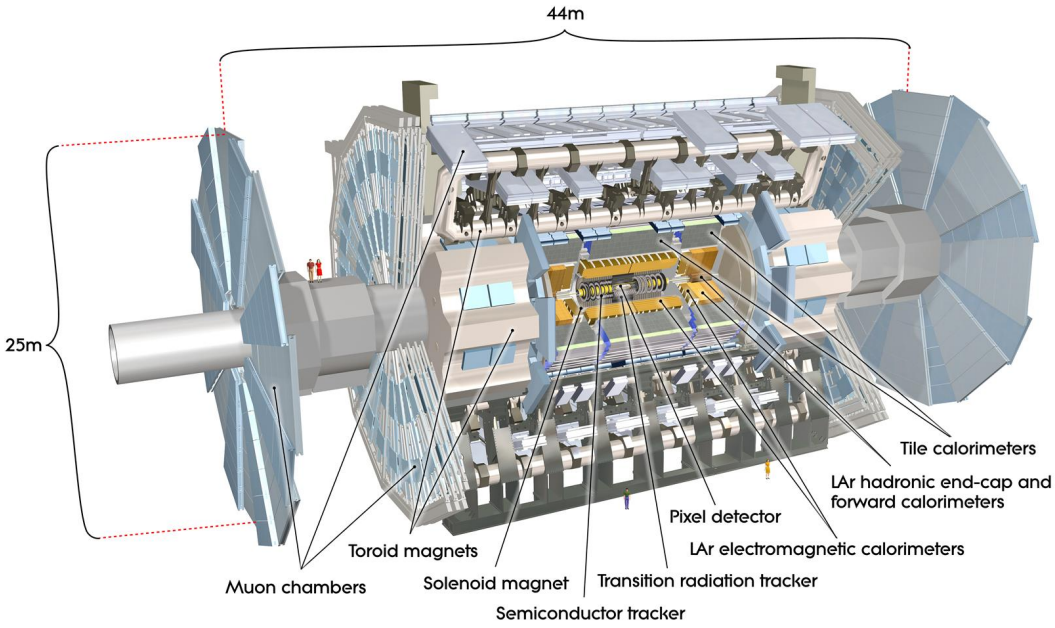


Figure 3.6: The ATLAS detector with various subsystems labelled. Average-sized humans are shown in red for scale [86].

necessarily have perfectly balanced momentum in the z-direction. Furthermore, particle production in proton-proton collisions is roughly isotropic in η rather than θ [87]. A few lines illustrating η in physical space are drawn in Figure 3.7.

The trajectories of particles are most often specified in ATLAS using η and ϕ, and the angular distance between two particles is given as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

3.2.2 Inner Detector

When a particle is emitted by a proton-proton collision at the ATLAS interaction point, the first object it encounters is the beampipe [88]. This is not an active element of the detector, just a 0.8 mm thick Beryllium pipe with an inner radius of 23.5 mm. Most particles pass through the beampipe without interacting; it constitutes 0.32% of a radiation length (X_0), which is the distance over which a high-energy electron loses all but $1/e$ of its energy.

Immediately outside of the beampipe is the inner detector [76, 89, 90], which has a primary purpose of reconstructing charged particles with $|\eta| < 2.5$ and measuring their momenta. The inner detector is schematically laid out in Figure 3.7. Broadly speaking, there are three subcomponents of the inner detector: the Pixel Detector and Insertable B-Layer (IBL), the Semiconductor Tracker (SCT), and the Transition Radiation Tracker

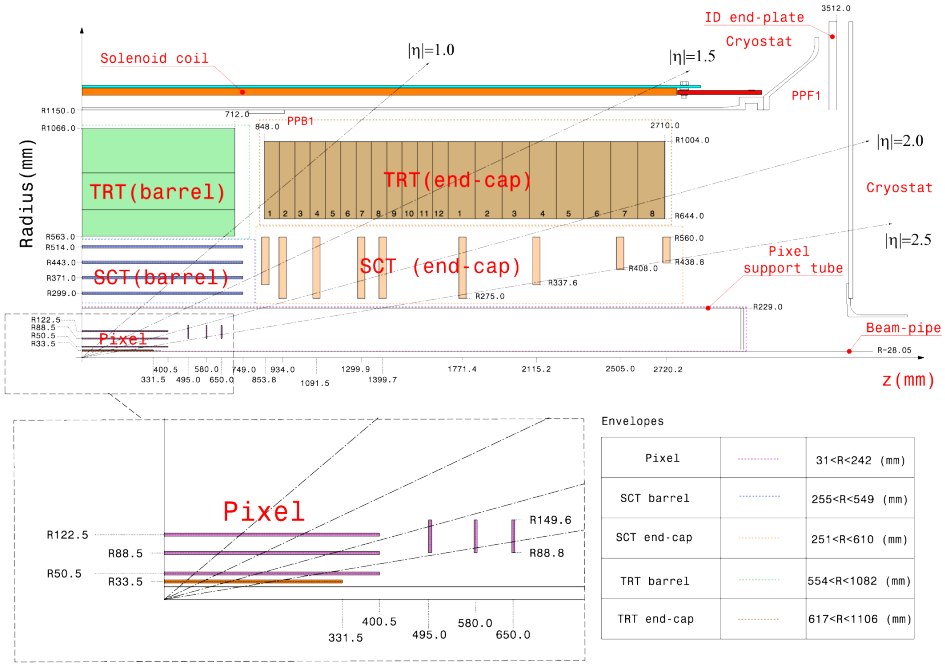


Figure 3.7: Cross-sectional layout of ATLAS inner detector [88]. The different components of the detector, including the IBL are specified, along with their radial information. Guiding lines are also included to illustrate the η values for particles passing through various components of the detector.

(TRT).

The IBL and Pixel Detector

The IBL and Pixel Detector comprise the components of the detector closest to the interaction point [76, 89, 91, 90]. They are four layers of silicon pixel modules, forming cylinders around the beampipe with radii of 33.5 mm (IBL), 50.5 mm, 88.5 mm, and 122.5 mm (Pixel Detector). A "pixel" is a small, rectangular region of silicon. When a charged particle passes through the silicon, charge is liberated within the silicon that can be read out by attached electronics. As a charged particle passes through multiple layers of the detector, it's position at the layers' radii is known to within the resolution of the pixels, allowing its trajectory to be calculated.

The IBL was inserted into the ATLAS detector, along with a new beam pipe, between Run 1 and Run 2 of the LHC. It's main purpose was to improve the resolution of measurements of where charged particles originate, both in z and in r. This origin resolution is important for identifying B-hadrons, which is where the IBL gets its name. The IBL consists of 14 staves, laid out around the beampipe as illustrated in Figure 3.8 (a) and (b). The pixels

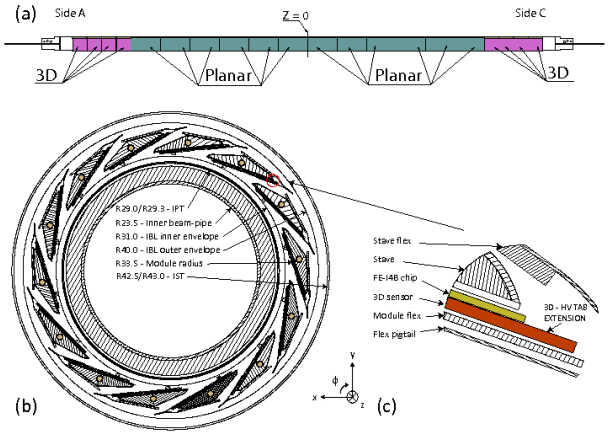


Figure 3.8: Illustration of the ATLAS Insertable B-Layer (IBL) [90]. A cross-sectional image in the bottom left illustrates how the staves are oriented around the beam pipe, and the top image illustrates how the modules are arranged on each stave.

on the staves have a width in the ϕ-direction of 50 μm, a length in the z-direction of 250 μm, and are either 200 μm or 230 μm thick, depending on whether they are "planar" or "3D" pixels (see Figure 3.8). In the barrel layers, the size of the pixels allow for position measurements with a resolution of about 10 μm in the ϕ-direction and about 70 μm in the z-direction. Each module consists of 336 columns and 80 rows of pixels. There are about 12 million total pixels in the IBL.

The three layers of the Pixel Detector are called the "B-layer", "Layer 1", and "Layer 2" in order of increasing radius and are arranged similarly to those of the IBL. There are additionally two disks with three layers each, as shown in Figure 3.7. The pixels in the Pixel Detector have width in the ϕ-direction of 50 μm, a length in the z-direction of 400 μm, and are 250 μm thick. The $r\phi$ and z resolutions are about 10 μm and 120 μm, respectively, in the barrel layers. Each module consists of 160 columns and 18 rows of pixels. There are about 80.4 million pixels in the Pixel Detector.

Appendix F details a calibration method I developed for the charge readout of silicon pixel detectors.

The Semiconductor Tracker (SCT)

The Semiconductor Tracker (SCT) [76, 92, 93] comprises the next four layers of the detector beyond the Pixel Detector. The radii of the layers in the barrel component of the SCT are 299 mm, 371 mm, 443 mm, and 514 mm. There are also two endcaps with nine disks each, as shown in Figure 3.7. The active elements of the SCT are made of silicon, like the IBL and Pixel Detector, but instead of pixels, these layers have silicon strips. A silicon strip detects charged particles in a similar way to a pixel detector, but instead of being composed

of relatively small rectangles, the detecting units are long, thin strips. The strips in the SCT have a width of 80 μm and a length of approximately 12 cm. If a single strip was isolated parallel to the beamline, it would allow for a position resoltion of about 23 μm in the ϕ-direction and about 3.5 cm in the z-direction.

However, each of the four layers of the SCT consists of sheets of strips that are glued back-to-back of each other. The strips of one side are oriented with their long side parallel to the beamline. The other side is rotated with respect to its counterpart by 40 mrad (such that the strips' long sides form an "x"). Combining the hit information in the two sides allows for a position resolution of about 17 μm in the ϕ-direction and about 580 μm in the z-direction.

In the full SCT, there are about 6.3 million strips.

If a charged particle passed through all of the barrel layers of the IBL, Pixel Detector, and SCT, it would leave 12 hits: 4 in the silicon pixels and 8 in the strips (2 per SCT layer). Most charged particles that pass through the endcaps of either system would leave 12 hits as well.

The Transition Radiation Tracker (TRT)

The Transition Radiation Tracker (TRT) [76, 89, 94] is a drift tube tracker that extends from a radius of 563 to 1066 mm in the barrel region. There is also an end-cap component to the TRT as indicated in Figure 3.7. A drift tube (or straw) tracker consists of a large number of long skinny tubes that are filled with a gaseous mixture that is ionized as charged particles pass through. A wire runs along the center of the tube, and an electrical potential difference is maintained between the wire (anode) and tube (cathode). This causes the liberated charge to drift to the wire, creating a current that can be read out. The amount of charge created will depend on the sagitta of the chord created by the charged particle's trajectory.

The straws in the TRT have a diameter of 4 mm, a length of 144 cm, and were initially filled with a mixture of gas that is 70% Xenon. Due to leaks, some fraction of tubes have been filled with an Argon-based gas, which is significantly cheaper [95]. The anode wire is gold-plated tungsten. There are 52,544 straws in the barrel region and 122,880 straws in the endcap, arranged such that a charged particle with $|\eta| < 2$ should pass through 35-40 straws. A TRT straw can determine a charged particle's distance of closest approach to the anode wire with a resolution of 130 μm.

The TRT derives its name from the phenomenon of "transition radiation". This phenomenon occurs when a relativistic particle passes between two media with different dielectric constants [96]. There are polypropylene fibers between the straws in the barrel region and foils between the straws in the endcap. As the particles pass between these materials and the straws, photons are emitted. For electrons, these photons tend to have energy of around 6 keV, which is about an order of magnitude higher than the typical energy deposited by a charged particle ionizing the straws' gas. Therefore, detecting higher-energy deposits can improve electron identification. Unfortunately, the use of an Argon-based gas mixture has

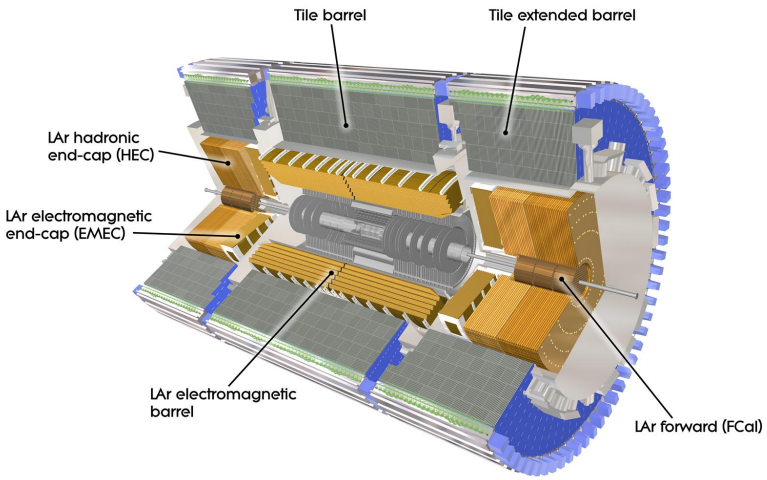


Figure 3.9: Layout of the ATLAS Electromagnetic and Hadronic Calorimeters, including barrel and endcap regions [98]

effectively taken away this aspect of the TRT, as Argon is a poor absorber of transition radiation. However, it is still adequate for tracking purposes.

3.2.3 Electromagnetic Calorimeters

Immediately outside of the Inner Detector is the ATLAS central solenoid [97], which provides the z-directional, 2 T magnetic field in the inner detector. It is a cylindrical barrel, with an inner radius of 1.23 m, an outer radius of 1.28 m, and a length of 5.8 m, contributing about 0.66 radiation lengths to the material in the detector at normal incidence.

Beyond the solenoid magnet is the ATLAS Electromagnetic (EM) Calorimeter system, which is a liquid Argon (LAr) sampling calorimeter. The primary goal of this component of the detector is to fully capture electrons and photons and to measure their energies. The layout of the EM Calorimeter, including the barrel and endcap regions is shown in the indicated components of Figure 3.9.

The EM Calorimeter consists of a barrel region, two endcaps (EMEC), and two Forward Calorimeters (FCal), one each on either side of the detector, providing full coverage in ϕ. Combining all components of the EM Calorimeter, there is coverage for particles with $|\eta| < 4.9$, where coverage for $3.1 < |\eta| < 4.9$ is provided by the FCal.

The barrel region and endcaps are constructed from interleaved lead and LAr, arranged in an accordion layout, as shown in Figure 3.10. When a high-energy photon or electron interacts with the electrons of the lead component of the detector, it causes a shower of additional (lower energy) photons, electrons, and positrons. The shower creates ionization in

Layer	$\Delta\eta \times \Delta\phi$
1	0.003×0.1
2	0.025×0.025
3	0.05×0.025

Table 3.1: Granularities of the three layers of the ATLAS Electromagnetic Calorimeter's barrel and endcap regions. There is no third layer for the region $2.5 < |\eta| < 3.2$.

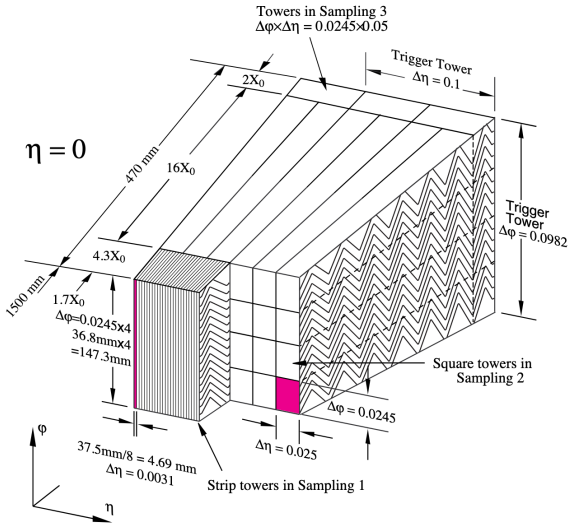


Figure 3.10: Cross-sectional image of the ATLAS Electromagnetic (EM) Calorimeter as would be seen in the barrel region, exhibiting the accordion layout of interleaved lead and liquid Argon [99]. The three layers of decreasing granularity in η are typical of the barrel region of the EM Calorimeter.

the LAr, which creates a current in the associated electronics. The "precision measurement" region covers particles with $|\eta| < 2.5$. Here the EM Calorimeter has three layers, as seen in Figure 3.10. In the region $2.5 < |\eta| < 3.2$, there are two layers. There is a LAr presampler before the EM Calorimeter in the region $|\eta| < 2.5$ to measure energy lost in front of the EM Calorimeters. The barrel region of the EM Calorimeter has an inner radius of 1.4 m and an outer radius of 2 m. The granularities of the three layers of the EM Calorimeter are presented in Table 3.1. There are about 164,000 total readout channels in the EM Calorimeter.

The FCal consists of one layer of interleaved LAr and copper and two layers of interleaved LAr and tungsten. The LAr-copper component provides primary EM calorimetry for this component. The FCal will be discussed more in Section 3.2.4.

The total number of radiation lengths traversed by particles before and within the EM Calormeters is shown in Figure 3.11, as a function of the particles' η.

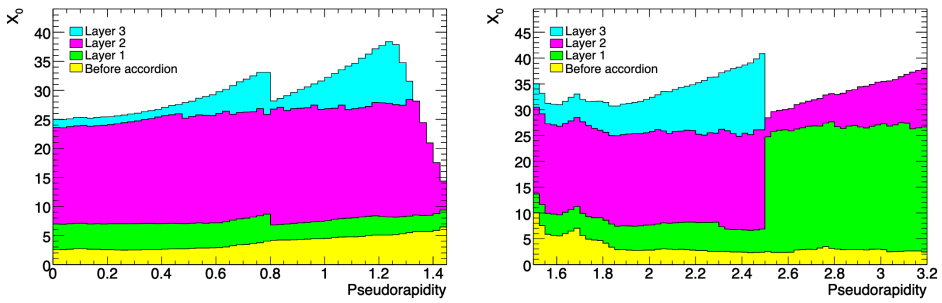


Figure 3.11: Total number of radiation lengths traversed by particles prior to and through the Electromagnetic Calorimeter in the barrel region (left) and endcap (right) [76]

3.2.4 The Hadronic Calorimeters

The ATLAS Hadronic Calorimeter [76] has three components: the Tile Calorimeter [100], the Hadronic Endcap (HEC), and the LAr FCal [99]. The layout of these components is shown in Figure 3.9.

All three of these components are sampling calorimeters, like the EM Calorimeters. However, where an EM calorimeter detects photons, electrons, and positrons through their interactions with the absorbing material's electrons, a hadronic calorimeter detects hadrons via their interactions with the absorber's nuclei. For EM calorimeters, a typical measure of depth is radiation lengths (X_0), but for hadronic calorimeters, the corresponding value is nuclear interaction lengths (λ), which is the mean distance travelled by a hadron before a nuclear interaction.

The Tile Calorimeter

In the Tile Calorimeter, the absorbing material is steel and the active material is scintillating tiles. The layout of the steel and scintillator is shown in Figure 3.12. The inner radius of the Tile Calorimeter is 2.28 m, extending to an outer radius of 4.25 m, with a 5.8 m long barrel region that covers a region of $|\eta| < 1.0$ and 2.6 m long extended barrels that cover the regions with $0.8 < |\eta| < 1.7$. The depth of Tile Calorimeter in interaction lengths is exhibited in Figure 3.13. There are three layers in the Tile Calorimeter; the first two layers have a $\Delta\eta \times \Delta\phi$ granularity of 0.1×0.1, and the last layer has a granularity of 0.2×0.1. The light created by the scintillating tiles are transmitted to photomultiplier tubes using wavelength shifting optical fibers. There are 5076 readout channels in the barrel component and 4092 channels in each of the extended barrels.

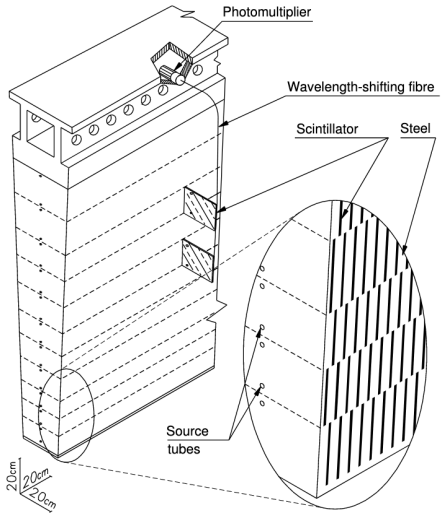


Figure 3.12: Layout of the Tile Calorimeter [76]. Steel and scintillating tiles are interleaved to created and measure hadronic showers.

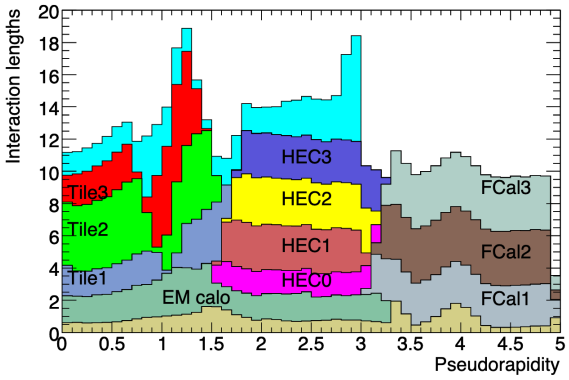


Figure 3.13: Total amount of material (in interaction lengths) encountered by a particle up to and including the Hadronic Calorimeters as a function of η [76]. The contributions of the different components of the detector are labelled.

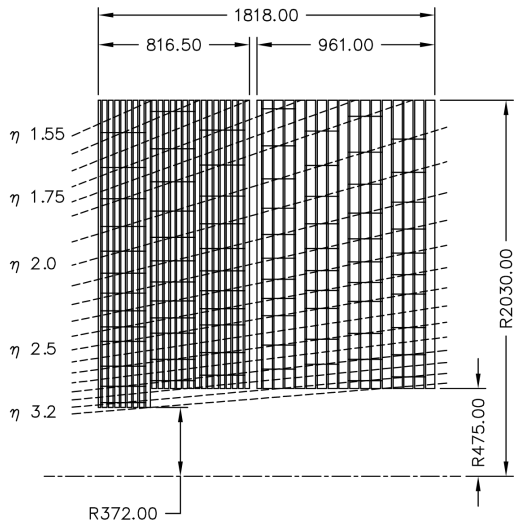


Figure 3.14: Layout of the Hadronic Endcap Calorimeter [76]. Copper and LAr are interleaved to created and measure hadronic showers. The distance measurements are in mm.

The Hadronic Endcap (HEC)

The HEC uses LAr as the active material, similar to the EM Calorimeters, but it uses copper as an absorber. It also uses a flat-plate design rather than the accordion layout, as shown in Figure 3.14. The HEC covers particles with $1.5 < |\eta| < 3.2$, and specifications of its design are included in Figure 3.14, where the dashed line at the bottom indicates the beamline. In the z-direction, each endcap is divided into four sections, where the inner two sections are more granular than the outer two. The depth in interaction lengths is presented in Figure 3.13. For particles with $1.5 < |\eta| < 2.5$, the $\Delta\eta \times \Delta\phi$ granularity is 0.1×0.1, and for $2.5 < |\eta| < 3.2$, it is 0.2×0.2. There are 5632 readout channels on both sides.

The Forward Calorimeters (FCal)

The ATLAS FCal was first introduced in Section 3.2.3. It provides coverage for particles with $3.2 < |\eta| < 4.9$. Some of the design details are shown in the left-hand image of Figure 3.15, including its radius, length, and positioning relative to the HEC and EMEC. It has three layers, where the first is interleaved LAr and copper, and the second and third are interleaved LAr and tungsten. The first layer is primarily for EM calorimetry, while the second and third are for hadronic calorimetry. The depth in interaction lengths is presented in Figure 3.13. The FCal is laid out as a matrix of tubes with their axes parallel to the beamline as shown in right-hand image of Figure 3.15. The LAr forms a ring along the inside the tube, and there is a rod of absorber in the tube. There are 3542 FCal readout channels on both sides of the detector.

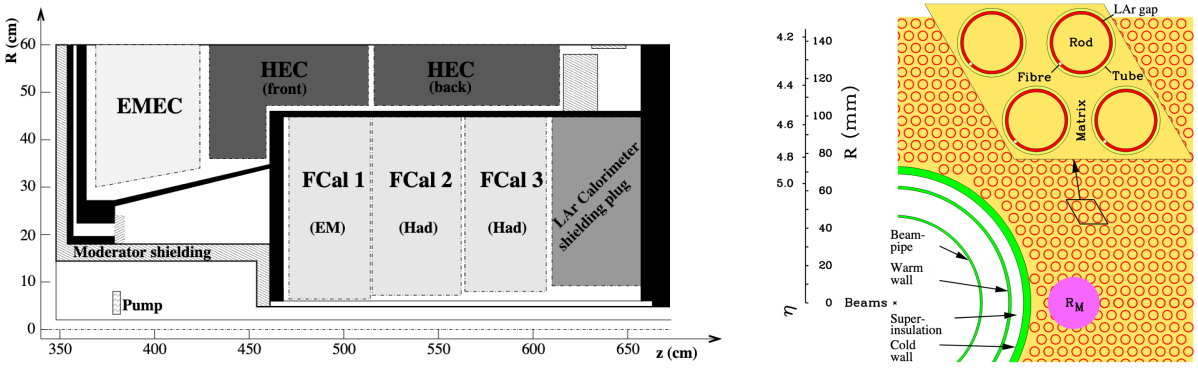


Figure 3.15: Layout of the Forward Calorimeter (FCal) [76]. Copper and LAr are interleaved in the first layer, and Tungsten and LAr are interleaved in the second and third layers. (Right) The positioning relative to the Hadronic Endcap (HEC) and Electromagnetic Endcap (EMEC) is shown. (Left) The matrix of absorber tubes and LAr is shown.

3.2.5 Muon Spectrometer

The outermost section of the ATLAS detector is the multi-component Muon Spectrometer (MS) [76, 101, 102]. Muons are the most penetrating long-lived particle that we know of (other than neutrinos), which is why this system is the farthest from the interaction point. Few other types of particles should reach the MS. The four main active components are the Monitored Drift Tubes (MDT), the Cathode Strip Chambers (CSC), the Resistive Plate Chambers (RPC), and the Thin Gap Chambers (TGC). There are three large air-coil toroidal magnets that create a magnetic field for the spectrometer: a barrel component and two endcaps. In fact, the ATLAS collaboration derives its name from these toroids, as originally ATLAS stood for "A Toroidal LHC ApparatuS". The layout of the active components and magnet systems are shown in Figure 3.16.

Magnet System

The barrel toroid and toroidal endcaps create a magnetic field that primarily curls around the detector in the ϕ-direction. Because of this, a muon travelling exactly transverse to the beamline would be deflected in the z-direction. Thus a muon has a relatively complicated trajectory in the ATLAS detector, being bent in the ϕ-direction by the solenoid magnet and then in the z-direction by the toroids. The bending power of the magnets results from the integral $\int$ B dl. The barrel toroid covers particles with $|\eta| < 1.4$ and provides 1.5 to 5.5 Tm of bending power. The endcaps cover particles with $1.6 < |\eta| < 2.7$, providing 1 to 7.5 Tm of bending power. Particles with $1.4 < |\eta| < 1.6$ do encounter a magnetic field, but with reduced bending power. The non-uniform magnetic field created by this system is simulated as the superposition of the Biot-Savart contributions of all of the magnet windings. A system

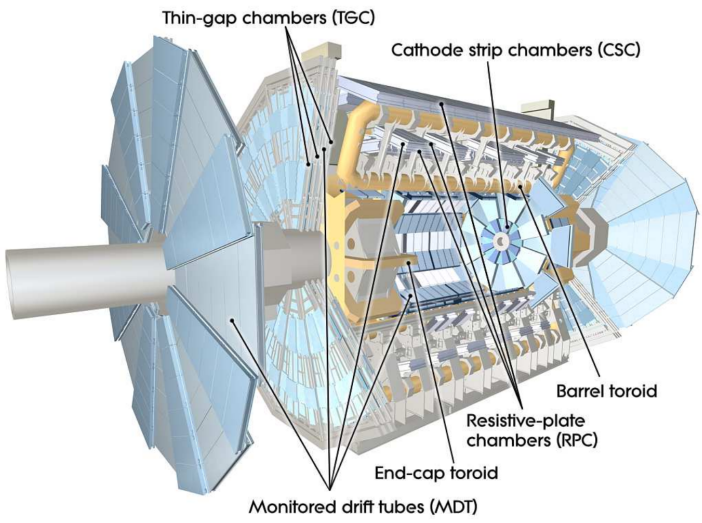


Figure 3.16: Layout of the ATLAS Muon Spectrometer, including the active components and toroidal magnet system [76].

of 1720 Hall cards are positioned throughout the MDT chambers to measure the magnetic field in practice.

Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC)

The MDT and CSC components of the MS are designed to provide accurate muon tracking information, with the bulk of the performance coming from the MDTs. In the barrel region, the MS is arranged in three concentric shells, and in the endcap region, there are four layers in z, as illustrated in Figure 3.17. The MDTs cover a region of $|\eta| < 2.7$, though there is a crack in between the two halves of the detector (split in z), such that particles with $|\eta| < 0.04$ might not pass through three MS layers. Similarly, there are some gaps in ϕ coverage due to support structures.

The principle of operation of the MDTs, illustrated in Figure 3.18, is similar to that of the TRT (see Section 3.2.2). The MDTs are constructed from 3 cm-diameter Aluminum tubes that are filled with pressurized Ar/CO_2 gas at 3 bar, and ionization electrons are collected with a 50 μm, gold-plated tungsten-rhenium wire. The central axis of the tubes is oriented in the ϕ-direction so that bending in the z direction can be measured well. The average position resolution is about 80 μm. Each of the approximately 1,100 MDT chambers is composed of two multilayers of tubes, where a multilayer has three or four layers of tubes. There are about 350,000 readout channels in the MDTs.

As can be seen in Figure 3.17, the CSC contributes to muon tracking only in the innermost endcap layer, for particles with $2 < |\eta| < 2.7$. The CSCs are multiwire proportional chambers with cathode strip readouts. The layout of a CSC is illustrated in Figure 3.18. A multiwire

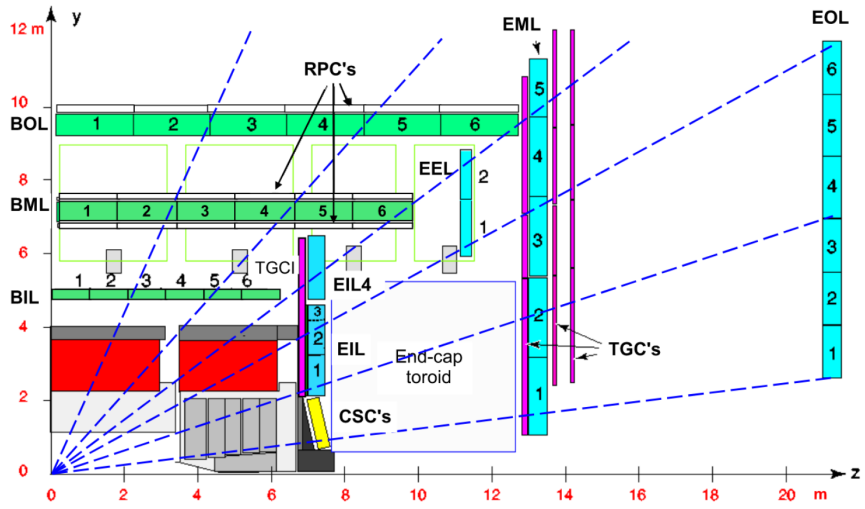


Figure 3.17: Cross-sectional layout of the ATLAS Muon Spectrometer [76]. The green and blue regions are Monitored Drift Tubes.

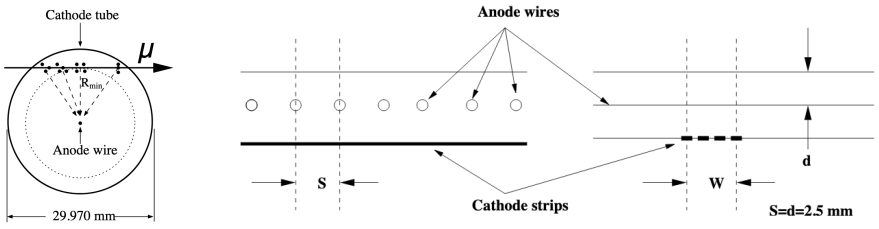


Figure 3.18: Illustration of the operation principle of a Monitored Drift Tube (left), and an illustration of the layout of a Cathode Strip Chamber [76].

proportional chamber is filled with gas that is ionized when a charged particle passes through. An electic field is maintained using central anode wires and outer cathode strips, so ionization charge drifts onto these strips, where it is collected, giving a 1D position measurement. The wires are in the radial direction, and the strips on either side are perpendicular, so they can be combined to give a 2D measurement. CSCs are used in this region due to their higher rate capability and time resolution. They have a resolution of about 40 μm in the bending plane and about 5 μm in the transverse plane. There are about 31,000 total CSC channels.

Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC)

Muons are a common feature in "interesting" proton-proton collisions at the LHC and are a feature in events used for the $\gamma\gamma \to W^+W^-$ analysis. It is important for ATLAS to

efficiently store these events, which is why high-quality data-taking triggers are important to the collaboration (see Section 3.3). The RPC and TGC components of the MS are primarily used to quickly identify muons, determine their multiplicty and estimate their energy for triggering purposes. They also contribute to the precision measurement goal of the MS.

The RPCs are used only in the barrel section of the MS as indicated in Figures 3.16 and 3.17, providing triggering for muons with $|\eta| < 1.05$. RPCs are attached on the inner and outer side of the Barrel Middle Layer and on the outside of the Barrel Outer Layer, meaning that every muon that escapes the MS should pass through at least three RPCs. An RPC is a gaseous parallel electrode-plate detector: two resistive plates are placed parallel to each other, 2 mm apart, and an electric field of 4.9 kV/mm causes avalanches to form from ionising tracks. The avalanche is detected with a signal width of about 5 ns. There are about 370,000 total channels in the RPCs.

The TGCs are only used in the MS endcaps, as shown in Figures 3.16 and 3.17. They cover particles with $1.05 < |\eta| < 2.7$ but provide triggering information for $1.05 < |\eta| < 2.4$, as higher pseudorapidity muons travel through fewer TGCs. There are seven layers of TGCs on the middle layer of the endap and two in the innermost layer. TGCs are multi-wire proportional chambers, like the CSCs, and the principle of operation is therefore the same, though the spacing of the wires is reduced to 1.8 mm and the wire-to-cathode distance is reduced to 1.4 mm. There is a 99% probability that signals are recorded within a time window of 25 ns.

3.2.6 ATLAS Forward Proton Detector (AFP)

While the ATLAS Forward Proton detector (AFP) [103, 104] is not used in the $\gamma\gamma \to W^+W^-$ analysis, it was designed to be used in photon-fusion analyses, and therefore bears mentioning. The AFP is not pictured in Figure 3.6, as it is actually positioned about 200 m away from the interaction point, on both sides of the main detector. The positioning of the AFP is shown in the left-hand image of Figure 3.19. Physically, the AFP is *inside* the LHC beampipe, attached to Roman Pots [105], on bellows which allow the detector to be positioned closer to or further from the actual beamline as desired.

The AFP consists of a silicon pixel tracking component and a Time of Flight component. The tracking component consists of five layers on both sides of the ATLAS detector. The pixels are similar to those used in the IBL, with a size of $50 \times 250\mu\text{m}^2$. Typically, the AFP keeps the closest edge of the pixel module at a distance of 2-3 mm from the beamline. The tracking component of the AFP is critical for measuring the energy of a deflected proton. If a proton does not interact at all as it passes through the beamline, then the magnet system of the LHC will keep it in the beamline. However, if the proton interacts and loses energy (by emitting a high-energy photon, for example), then it will be deflected outside of the beamline. The AFP measures the physical deflection of the proton, which can be converted into a measurement of the proton's energy when combined with detailed knowledge of the magnetic field in the LHC's beampipe. For example, the components in the left-hand image in Figure 3.19 labelled as Q and D are quadrupole and dipole magnets that the protons will

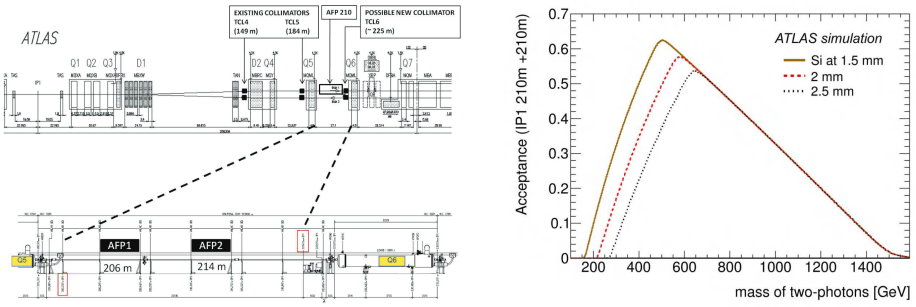


Figure 3.19: (Left) Physical positioning of the ATLAS Forward Proton detector (AFP) [104]. It occupies two stations on both sides of the detector, both more than 200 m from the interaction point. (Right) Expected acceptance of forward proton system as a function of the center of mass energy of the two-photon system in an *elastic* photon-fusion event.

pass through. If an elastic photon-fusion event occurs (see Section 2.2.2) and both protons remain intact, then the energy lost by both protons can be measured using the AFP, giving a precise measurement of the center of mass energy of the two-photon state. The acceptance for such protons is given as a function of the photon center of mass energy in the right-hand image of Figure 3.19. The timing component of the AFP is constructed from quartz radiator bars coupled to micro-channel plate photomultiplier tubes. It is primarily used for pileup suppression.

The AFP was not used in this thesis because data was taken incorrectly with this detector component in Run 2. While information *was* read out, it was for different bunch crossings than those that triggered readout in the main ATLAS detector. Only about 10 fb^{-1} of good data was collected in Run 2, which was enough data for a $\gamma\gamma \to \ell^+\ell^-$ analysis [21], but the $\gamma\gamma \to W^+W^-$ cross-section is too low to make an observation in such a small dataset.

3.3 Triggers

During nominal running, proton bunches pass through each other at the ATLAS interaction point every 25 ns, or with a frequency of 40 MHz. Each crossing where a proton-proton interaction occurs is called an "event". The amount of information collected for a single event is on the order of 1 MB, so the data rate associated with recording every event at the LHC would be about 40 TB/s. Not only is transmitting that much data out of the detector almost impossible given current technology, but recording LHC collisions for a 24 hour period would use up over a million TB of storage space. However, depending on how you look at it, the situation is not actually so dire. The collisions in most events are not very "interesting". In fact, a popular statistic to throw around about the LHC is that a Higgs Boson is produced in about one out of a billion proton-proton collisions [106]. If only one out of a billion events must be recorded, then the data rate drops below 1 MB/s and the

daily storage requirements below 100 GB, both of which are easily achievable. Of course, ATLAS cares about more than Higgs events alone, but it is still necessary to be selective about which events get recorded. Deciding on these events is the the job of the ATLAS Trigger system [76, 107, 108].

The ATLAS trigger system has two components: a hardware-based first level trigger (L1) [109] and a software-based High Level Trigger (HLT) [110]. The L1 trigger reduces the event rate down to about 100 kHz, making decisions to retain events in about 2.5 μs. The HLT further reduces the rate to about 1 kHz, with a processing time of about 200 ms. The trigger makes its decisions about keeping or rejecting events by reconstructing objects in the event, such as a muon (see Section 3.2.5), and determining if the objects pass certain criteria. For example, an event might be triggered if there is a single muons with $p_{\mathrm{T}} > 20$ GeV. Additionally, there might be a prescale factor applied to randomly accept or reject events; if an event type has a prescale factor of n, then there is a $1/n$ chance that the event is kept. In Run 2, about 1500 event selections were included to build up a trigger menu. If an object or combination of objects in an event appears on the menu, then the event may be saved.

Level-1 (L1) Trigger

The L1 Trigger consists of two components: a L1 Calorimeter trigger and a L1 Muon trigger. The calorimeter system uses reduced-granularity objects to search for electrons, photons, taus, jets, and missing energy, and the muon system uses the RPCs and TGCs to search for muons. The custom electronics of the L1 trigger identify regions of interest (RoIs) in the detector which should be investigated by the HLT.

The ultimate L1 trigger decision is made by the Central Trigger Processor, which combines the calorimeter and muon information, as well as information from additional subsystems that can be used to trigger readout on random events. The L1 trigger also applies dead time to the trigger, enforcing a minimum time between subsequent L1 triggers to avoid overlapping readout. At the end of Run 2, the dead time was four bunch crossings, which caused a performance inefficiency of about 1%. When a event passes the L1 trigger, all of the ATLAS detector systems are read out, and their data is processed and temporarily stored until requested by the HLT. Rates of the various objects that triggered L1 readout for an LHC fill in July 2016 are shown in Figure 3.20. A fill is the period between when protons are injected into the LHC and when they are dumped, corresponding to a time period of about 24 hours. The instantaneous luminosity decreases over this period as protons collide or are otherwise deflected out of the beamline, leading to a decrease in the trigger rates. The occasional increases in rates are when the prescale factors are adjusted.

The High Level Trigger (HLT)

The software-based HLT operates by using fast trigger algorithms to provide early rejection if possible and more sophisticated reconstruction algorithms for events where rejection is

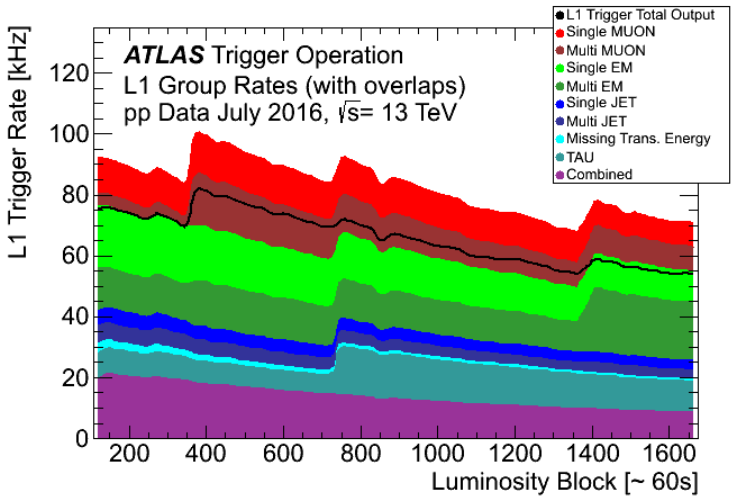


Figure 3.20: Rates of the various objects that triggered L1 readout for an LHC fill in July 2016 [111]. Rates decrease as a function of time because the instantaneous luminosity decreases over this period.

ambiguous. It only considers events that have passed the L1 trigger. The more sophisticated algorithms are similar to those discussed in Chapter 4 and are used for high-quality particle reconstruction. The efficiency of the HLT is improved by performing reconstruction only in the RoI passed to it by the L1 trigger, though for some events, information from the full detector can be used. If the event is not rejected at any point, then it is accepted. The total data output of events that pass the HLT is about 1.2 GB/s; these events are sent to permanent storage on magnetic tapes. Events that do not pass are removed from the temporary buffer that stored events that passed the L1 trigger. The rates of various objects that pass the HLT for an LHC in July 2016 are shown in Figure 3.21.

3.3.1 Single Electron and Muon Triggers

Events that are saved by ATLAS are categorized by the triggers they satisfy. As a specific example, the $\gamma\gamma \to W^+W^-$ analysis uses events that pass single electron [112] or single muon [113] triggers. Events that are triggered by the presence of a high-energy jet are less interesting in this analysis, for example.

The electron trigger at the L1 level uses EM calorimeter information. To pass L1 selection, an event must have a cluster of calorimeter energy that passes a minimum E_{T} threshold, where $E_{\mathrm{T}} = E \sin\theta$ for the energy deposit. There are a variety of potential E_{T} cuts, depending on the HLT trigger for which the event will be considered. There are also possible isolation cuts, rejecting an event if there is either too much energy behind the EM calorime-

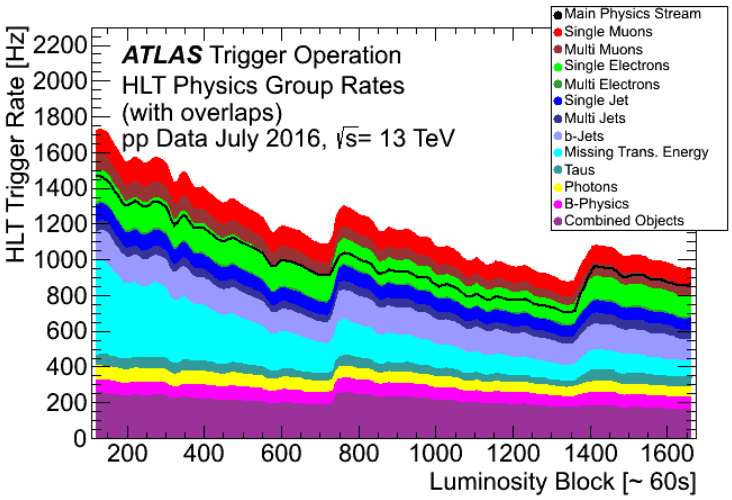


Figure 3.21: Rates of the various objects that passed the HLT for an LHC fill in July 2016 [111].

ter cluster in the hadronic calorimeter or if the energy in the EM calorimeter is too spread out from the center of the cluster. At the HLT level, fast tracking is performed using inner detector information in the RoI created around the EM calorimeter deposit. To identify a particle as an electron at trigger level, the $|\Delta\eta|$ and $|\Delta\phi|$ of the track and energy cluster must be smaller than 0.05.

For the single electron triggers used in the $\gamma\gamma \to W^+W^-$ analysis, the minimum E_{T} was 24 GeV in 2015 and 26 GeV in 2016-2018. The thresholds are set to ensure that the trigger rate remains within the L1 bandwidth. The rates of these triggers are shown as a function of instantaneous luminosity in the left-hand plot of Figure 3.22. The efficiency of correctly triggering on electrons is shown in Figure 3.23 as a function of E_{T} and η. The efficiency was determined using the tag-and-probe method with $Z \to ee$ events [114], and it plateaus above 90% for high-E_{T} electrons.

The single muon trigger at the L1 level uses the RPCs and TGCs to measure how distinct the hit pattern is from what would be expected for a muon with infinite momentum. The L1 trigger is operational for about 99% of the endcap, but only about 80% of the barrel region of the MS. This is due to the gaps around $\eta = 0$ and around the MS support structures, as were mentioned in Section 3.2.5. At the HLT level, MDT information is used to perform more precise muon tracking in the RoI indicated by the L1 trigger. The muon track is then extrapolated back to the interaction point to be combined with a track in the inner detector. The isolation of the muon can be factored into the trigger decision by summing up the p_{T} of the inner detector tracks that point to a ± 6 mm region in z around the muon (a ± 2 mm window was used in 2018 due to higher pileup). The rates of single muon triggers are shown

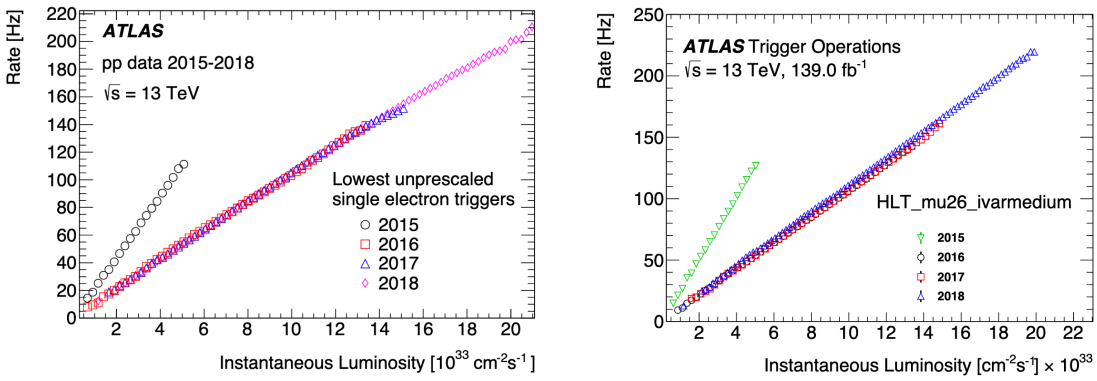


Figure 3.22: (Left) Trigger rates of the lowest threshold, isolated single electron trigger as a function of instantaneous luminosity for the four years of LHC Run 2 [112]. (Right) Trigger rates for the lowest threshold single muon trigger [113].

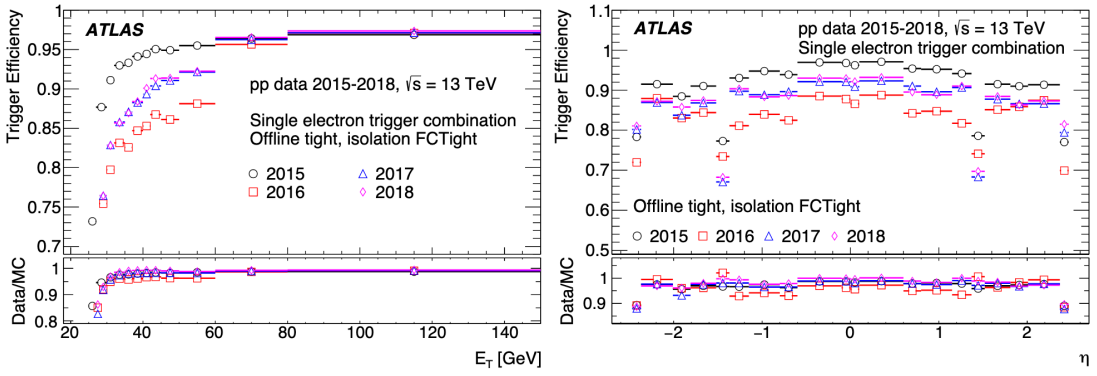


Figure 3.23: Efficiency for single electron trigger as a function of the offline (fully processed) E_{T} (left) and η (right) [112]. The electrons considered for this plot must be "tight" and satisfy the "FCTight" isolation requirement, which indicate high confidence that they are well-isolated, real electrons (see Section 4.2).

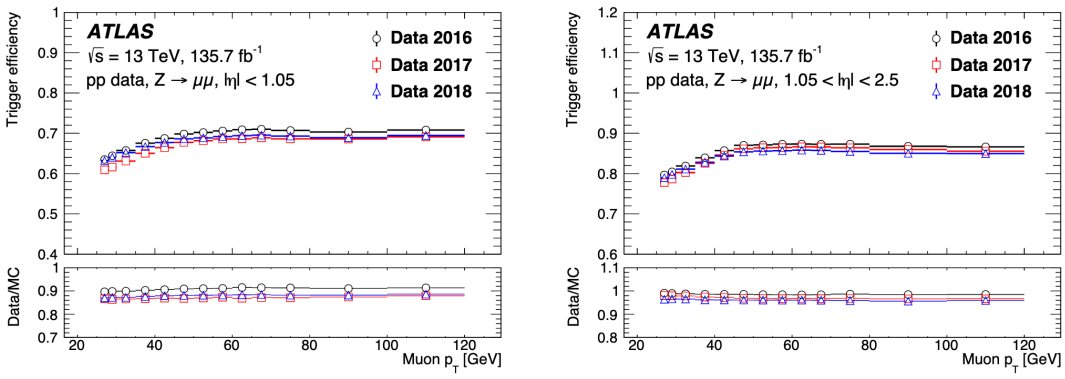


Figure 3.24: Efficiency for single muon trigger as a function of p_{T} in the barrel region (left) and the endcap region (right) [113].

as a function of instantaneous luminosity on the right-hand plot in Figure 3.22. Similar to the electron case, the efficiency was determined using the tag-and-probe method with $Z \to \mu\mu$ events [113]. The efficiency is shown as a function of p_{T} separately for the barrel and endcap regions. In the barrel region, the efficiency plateaus at around 70%, and in the endcap region the plateau is at about 90% for high-p_{T} muons.

For the single muon triggers used in the $\gamma\gamma \to W^+W^-$ analysis, the minimum p_{T} was 20 GeV in 2015 and 26 GeV in 2016-2018.

Chapter 4

Particle Reconstruction and Performance

The ATLAS detector is formed from many components, and these components work in concert to allow for particle reconstruction and measurement. A few examples of this concept were introduced in Section 3.3.1, when discussing how EM calorimeter hits and inner detector (ID) tracks are combined to find electrons and how MS tracks and ID tracks are combined to find muons. Rather than working with a collection of, say, MS tracks and EM calorimeter clusters for each event, analyses are typically performed using collections of all of the fully reconstructed photons, electrons, muons, jets (which result from the scattering of color-charged particles), and tracks in each event. Each of these objects has a host of potential associated properties, such as momentum and η for a muon, the number of ID hits along a track, or the number of tracks in a jet.

Detecting a particle is inherently a noisy process, which is to say that it is impossible to measure a feature like energy or momentum with infinite precision. Understanding resolution effects is an important feature of any analysis, as it will have a strong impact on any measurement or search. The resolution goals of ATLAS are laid out in Figure 4.1.

Detector component	**Required resolution**	**η coverage**	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\%\ p_T \oplus 1\%$	± 2.5	
EM calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic calorimetry (jets)			
barrel and end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2	± 3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	σ_{p_T}/p_T=10% at p_T = 1 TeV	± 2.7	± 2.4

Figure 4.1: Resolution goals of the ATLAS Experiment for track p_{T}, electron and photon energy, hadronic jet energy, and muon p_{T} [76].

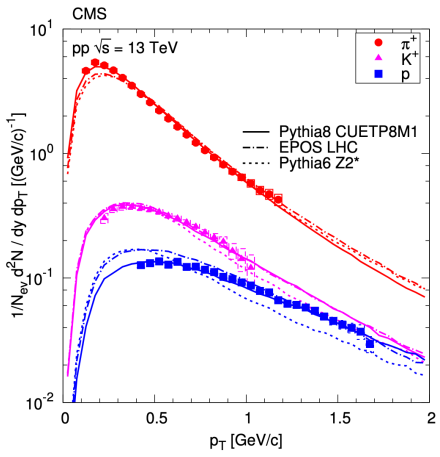


Figure 4.2: Distributions of pion, kaon, and proton yields vs p_{T} as observed by the CMS collaboration [115]. The distributions are normalized such that the relative yields of the particle species are preserved.

4.1 Charged Particle Tracking

Charged particle tracking plays a central role in this thesis[1]. At high pileup, there can be on the order of 1000 trackable charged particles passing through the detector per event. Accurately and efficiently reconstructing the trajectories of these particles is highly non-trivial, requiring both sophisticated hardware and software.

Over 90% of the particles of the charged particles created at the LHC are pions, most of which come from pileup collisions. Kaons and protons are the next most likely particles to be produced, and all of these particles are most likely to have p_{T} below 500 MeV as can be seen in Figure 4.2 [115]. While the low-momentum pileup hadrons are not relevant for most ATLAS analyses, they play a large role in this thesis (see Part III).

Tracking in ATLAS proceeds algorithmically as follows [116, 117, 118]:

1. Individual hits in the pixel layers and SCT layers are combined to form clusters [119].
2. These clusters are interpreted as space points and are stored in a space point collection.
3. Combinations of three clusters (on three different inner detector layers) are combined to form a track seed. Three SCT hit combinations are most favored, then three pixel hit combinations, then mixed SCT+pixel combinations. The seed can be extended to include a fourth confirmation hit.

[1]Throughout this thesis I will use the word "tracking" to refer to the determination of charged particles' trajectories using the ID.

4. Seeds are extended into track candidates with a Kalman Filter [120] that iteratively updates track parameters and next-layer position estimates based on newly acquired hits. A collection of track candidates is created. Track candidates may have holes in silicon layers or share hits with other candidates (a "hole" occurs when a track's trajectory is extrapolated through an active silicon layer, but the expected hit there is not found).

5. Ambiguity solving is performed on the track candidate collection. A score is assigned to each track based on quality. Score is reduced for issues such as having too many holes, too few hits, or too many shared hits with accepted higher-quality tracks. Candidates with scores above a certain threshold are accepted, and candidates with low scores are rejected.

6. Tracks are extended into the TRT and parameters are further refined.

7. A final track collection is created.

Appendices B and C discuss work I performed to improve ATLAS tracking by identifying SCT clusters that *should* be shared between tracks and by identifying single tracks created by charged particles with similar trajectories.

After the inside-out tracking, a second tracking pass is performed that starts with TRT "segments" and proceedes inwards, which is called "back tracking" or outside-in tracking. The hits used in the first set of tracks are removed for consideration in this round. The primary goal of this tracking pass is to find tracks associated with secondary vertices - sometimes particles decay in flight (such as a K_S) or interact with the detector material, creating a set of two or more particles that might not leave enough hits in the silicon layers to create a traditional seed.

An illustration of the hits used for tracking in a simulated $t\bar{t}$ event with no pileup is shown in Figure 4.3. The red points indicate hits that were used in the first round of tracking, and the darker highlighted TRT hits show those that were associated in the back tracking step.

Because charged particles are travelling in a magnetic field, their trajectories will be helical. A helix is completely described by five parameters, such as the radius of curvature, tilt, and a starting point. The set of parameters chosen by ATLAS for each track are:

- q/p_{T}, which is the charge divided by the transverse momentum. The p_{T} is related to the radius of curvature, and the charge dictates which way the particle bends.

- d_0, which is the transverse distance of the closest approach of the track to the beamline (signed depending on the particle's charge and whether the closest approach is in front of or behind the beamline relative to the center of the helix).

- z_0, which is the longitudinal (z-direction) distance of closest approach of the track to the primary vertex in the event [121] at the point were d_0 is measured. Alternatively, a z coordinate can be defined for each track using the center of the detector as the location of $z = 0$.

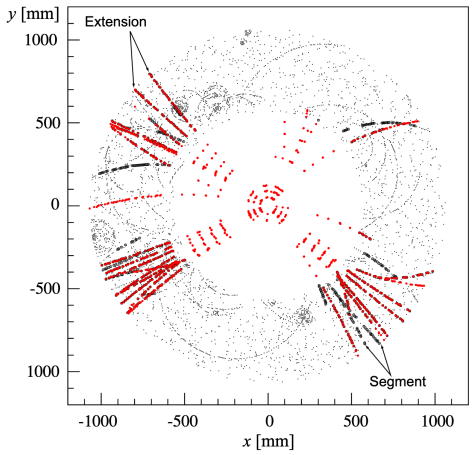


Figure 4.3: Hits used in tracks associate with a simulated $t\bar{t}$ event with no pileup [116]. Red points indicate hits that were used in the first round of tracking, and the black TRT hits show those that were associated in the back tracking step.

- ϕ_0, which is the initial angle of the trajectory in the ϕ-direction where d_0 and z_0 are measured.
- η, which is the track's pseudorapidity.

Track selection

An important thing to note is that tracks with p_{T} below a certain threshold are rejected outright. In 2015 - 2016, this threshold was 400 MeV, and in 2017 - 2018, it was 500 MeV. The minimum threshold was set at these values for two primary reasons. First, it is set to reduce storage space requirements for each event. Approximately half of the charged particles produced in a proton-proton collision at $\sqrt{s} = 13$ TeV will have p_{T} below 500 MeV [115, 87]. Second, the processing time is reduced by not having to consider seeds that would be extended into low-p_{T} tracks. Seed making is primarily a problem in combinatorics. The more possible combinations, the longer the required processing time. It is also possible that ID hits can be combined to form fake tracks; if noise hits or hits left by different charged particles are arranged in a way that looks roughly helical, the algorithm may associate them into a track. While this is not exclusively a feature at low-p_{T}, the problem is exacerbated in that regime.

Beyond the minimum-p_{T} cut, there are other quality control cuts placed on tracks from the first tracking pass:

- $|\eta| < 2.5$. This is the intended range of track coverage.

- Minimum of 7 pixel and SCT clusters. Leaving one hit in each layer would result in 12 hits. A track might leave fewer hits due to inefficiencies in the silicon or a dead module, for example.
- Track can share a maximum of either one pixel hit or two SCT clusters on the same layer with another track.
- Track can have a maximum of 2 holes in the combined pixel and SCT layers.
- A maximum of 1 hole in the pixel layers.
- $|d_0| < 2.0$ mm and $|z_0 \sin\theta| < 3.0$ mm. $|z_0 \sin\theta|$ is a good proxy for for the track's resolution in the z-direction.

Tracks that fail these quality cuts are rejected. Tracks that pass these selections are labelled as "loose". There is a "tight-primary" track selection working point that is used to reduce fake rate at the expense of tracking efficiency. Tight-primary tracks must satisfy all of the criteria above and have:

- At least 9 silicon hits if $|\eta| \leq 1.65$.
- At least 11 silicon hits if $|\eta| > 1.65$.
- At least 1 hit in the IBL or B-Layer.
- No pixel holes.

The expected tracking efficiency as a function of charged particle p_{T} and η is shown in Figure 4.4. Tracking efficiency is the likelihood that a charged particle will be reconstructed as a track. For tight primary tracks, tracking efficiency is generally above 80%, though it is worse at higher η.

Before the application of the tight selection, the fake rate in events of interest for this thesis is generally below 1%. After application of the tight selection, the fake rate is reduced to below 0.2%. More details on fake rate are given in Chapter 17.

Track resolution

Another particularly relevant track feature in this thesis is the d_0 and z_0 resolution. Plots of these resolution values as a function of p_{T} and η are shown in Figure 4.5. The resolution of both d_0 and z_0 improve as a function of p_{T} and are worse for higher η. For central, higher-p_{T} tracks, the d_0 resolution is about 50 μm, and the z_0 resolution is about 100 μm. For higher-η tracks, the resolution can be on the order of 0.5 mm or more. As indicated in Figure 4.1, the resolution of track p_{T} gets worse with increasing p_{T}, with a minimum uncertainty of about 1%.

Tracking is extremely important in the $\gamma\gamma \to W^+W^-$ analysis for signal vs. background discrimination purposes. The primary backgrounds to the $\gamma\gamma \to W^+W^-$ process are QCD

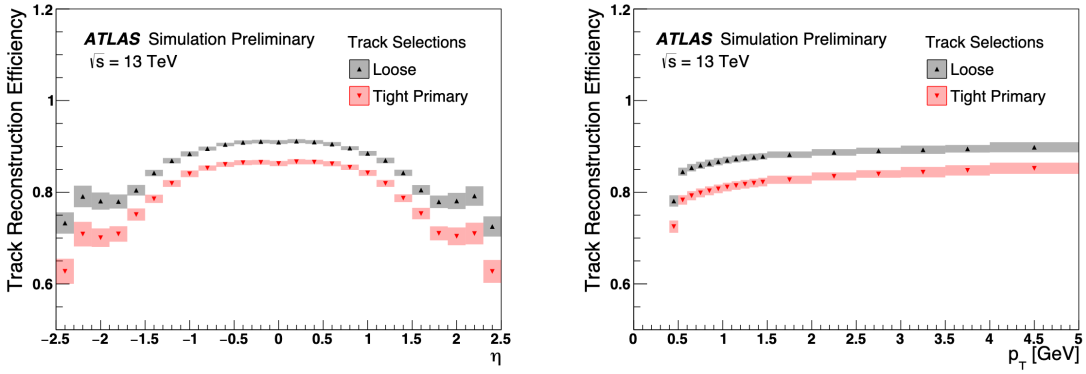
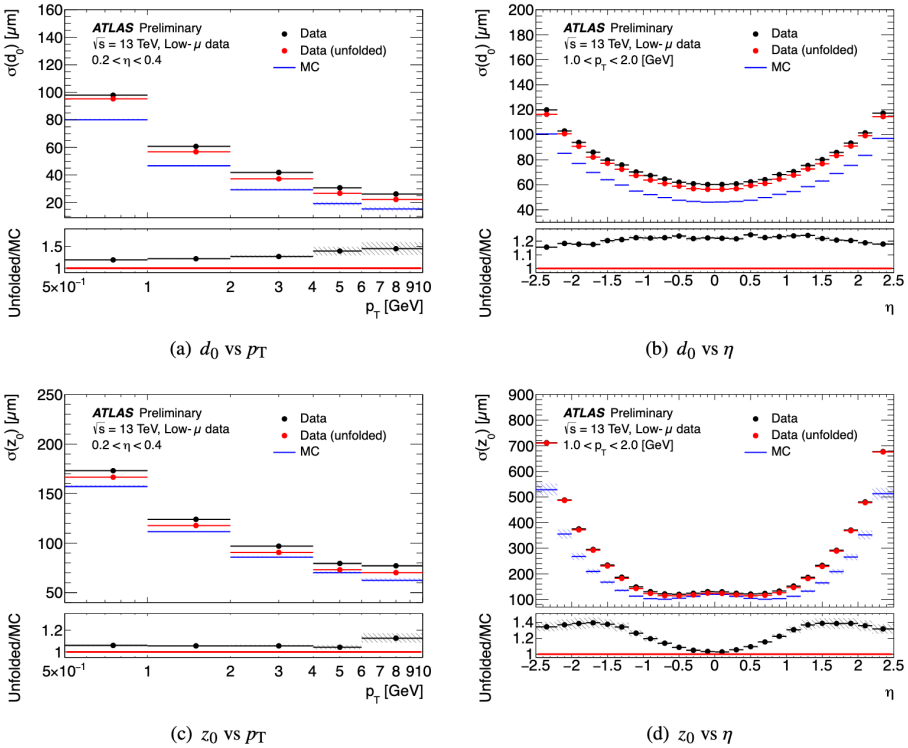


Figure 4.4: Simulated tracking efficiency as a function of η (left) and p_{T} (right) [122]. The "tight primary" and "loose" requirements are track selection working points.



(a) d_0 vs p_{T}

(b) d_0 vs η

(c) z_0 vs p_{T}

(d) z_0 vs η

Figure 4.5: Resolution of track impact parameters as a function of p_{T} and *eta* [123].

induced proton-proton interactions, which have underlying events (UEs). These UEs are primarily soft[2] pions, which manifest as tracks pointing to where the proton-proton interaction occurred. Photon-induced events will not have UEs, so in a $\gamma\gamma \to W^+W^-$ event, there should be no additional tracks pointing to the interaction location, other than those coming from the W decays.

An additional purpose of tracking for generic ATLAS analyses is vertex finding [121]. Vertices are defined as locations where proton-proton interactions occur along the beamline. A main goal in vertex finding is the association of tracks with a hard-scatter vertex, where a hard-scatter is loosely-speaking defined as a proton-proton event that results in an interesting interaction. For example, a QCD induced W^+W^- interaction would be a hard-scatter, but pileup interactions are not hard-scatters. If all of the tracks are assigned correctly, then the kinematic information associated with the interaction can be more fully reconstructed.

Vertex reconstruction is performed in two steps in ATLAS: vertex finding and vertex fitting. Vertex finding is the process of associating tracks together into vertices, and vertex fitting is the process of assigning a specific location to the vertex. Vertex finding proceeds iteratively. First a seed position is selected as the mode of the z-coordinate of tracks in the event. The vertex position is then determined through an iterative χ^2-minimization process, with tracks being assigned a weight based on their compatibility with the vertex. Once the fit converges, tracks that are deemed incompatible with the vertex are removed, and the remaining tracks are considered to be associated with the vertex. The process starts again with a new seed using all un-assigned tracks, and ends once all tracks have been assigned to a vertex. After the finding process ends, all of the vertices' positions are refitted and finalized. This vertex finding process is how the beamspot size information is determined, as shown in Figure 3.2.

The vertex with the highest sum of the p_{T} of its associated tracks ($\sum p_{\mathrm{T}}$) is called the "primary vertex". The primary vertex usually coincides with a hard-scatter in the event.

4.2 Electrons and Photons

Electron[3] and photon reconstruction relies on both the EM calorimeters and the ID [124, 125, 126]. Electron reconstruction plays a more significant role in this thesis than photon reconstruction, as W bosons can decay into electrons. There are some background processes that involve photons, but these are primarily a problem if the real photon is incorrectly determined to be an electron.

The first step of electron and photon reconstruction is the same for both species of particle. The EM calorimeter is divided into $\Delta\eta \times \Delta\phi$ towers of 0.025×0.025, and the total energy in each tower is computed. Electromagnetic energy clusters are then found by combining 3×5 towers into segments, forming seeds, and eventually applying a clustering

[2] "Soft" means relatively low energy or momentum

[3] Here and elsewhere "electron" refers to both electrons and positrons

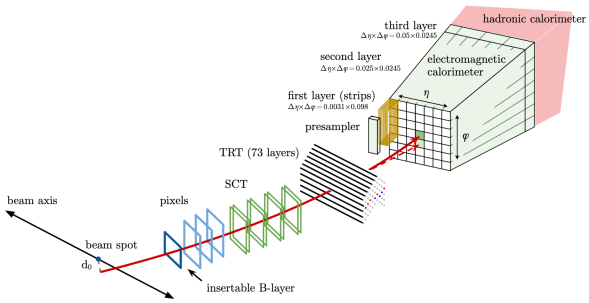


Figure 4.6: Illustration of the path that an electron takes through the ATLAS detector [125].

algorithm [127]. For electrons and photons with $E_{\mathrm{T}} > 20$ GeV, the cluter-finding efficiency is 99%.

Both electrons and photons rely on ID track reconstruction, which was described in Section 4.1. A Gaussian Sum Filter [128] is applied to the tracks to account for energy loss of electrons in the detector material, and the tracks are extrapolated to the second layer of the EM calorimeter. They are then matched to EM clusters, with the requirements that $|\eta_{\text{EM cluster}} - \eta_{\text{track}}| < 0.05$ and either $-0.20 < -q * (\phi_{\text{EM cluster}} - \phi_{\text{track}}) < 0.05$ or $-0.10 < -q * (\phi_{\text{EM cluster}} - \phi_{\text{track}})_{\text{res}} < 0.05$. In those equations, q is the particle charge as derived from the direction of bending, and "res" means that the momentum of the track has been rescaled to the energy of the EM cluster.

Using the EM cluster and tracking information, it is determined whether each EM cluster corresponds to an electron, a converted photon, or an unconverted photon. A converted photon occurs when a photon interacts with the detector material, splitting into an electron and positron. Information about the cluster's energy, the potential track's momentum, the number of silicon-layer hits on the potential track, and secondary-vertex information is used. An EM cluster matched to a high quality track with pixel hits generally indicates an electron, whereas a cluster with no track or with a track stemming from a secondary vertex is generally a photon. An illustration motivating the above electron reconstruction algorithm is given in Figure 4.6; an electron created at in a proton-proton interaction in the beamline should create a complete track before leaving its energy in the EM calorimeter. Because final-state photons are not used in this thesis, only electron selection working points will be discussed further.

Electron identification

Similar to the situation with tracks, different working points are chosen for electrons to balance reconstruction efficiency and fake rate. For each reconstructed electron, an electron-likelihood value is computed using a variety of discriminating variables [125], such as the ratio of a cluster's energy in the EM calorimeter to that in the hadronic calorimeter just

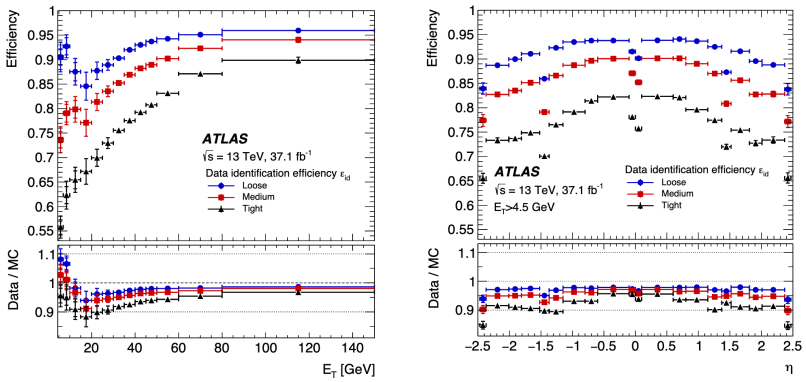


Figure 4.7: Electron identification efficiency as a function of E_{T} (left) and η (right) for the loose, medium , and tight electron identification working points [125].

behind it, relative energy of the cluster in the last EM calorimeter layer, EM cluster width, number of silicon hits on the associated track, and the angular agreement between the track and cluster. Four working points are chosen based on the outcome of the likelihood score: very loose, loose, medium, and tight. The electron identification efficiency for the loose, medium, and tight working points are given in Figure 4.7 as a function of the electron's E_{T} and η.

Electron isolation

An additional criterion is applied to electrons is that of isolation, which as no analog in the case of tracks. A well-isolated electron is likely to have come directly from a hard-scatter event, but a poorly-isolated electron is likely to have originated in a heavy-flavor hadron decay, such as a B-hadron decay[4]. Two types of isolation are considered for electrons. First, there is calorimeter isolation, which determines the ratio of the energy in the electron's EM cluster to the energy in the EM calorimeter in topological clusters in a cone around the electron [129]. The other isolation definition is track-based. For track-based isolation, the sum of the p_{T} of tracks within a cone of the electron is computed, with a few cuts on the tracks, such as requiring a p_{T} of > 1 GeV and that they point to where the electron originated in z. Three classes of working points are established: one that seeks uniform electron reconstruction efficiency in both E_{T} and η, one that seeks E_{T}-dependent efficiency that must still be uniform in η, and a last one that has efficiency that varies with both E_{T} and η, but has fixed cuts in the isolation variables. The first class is called "Loose", the second class is called "Gradient", and the last class is called "Fix". The "Fix" type of isolation is used in the $\gamma\gamma \to W^+W^-$ analysis. This class has several working points: loose, tight, tight (track only), Calo only, and Track $R_{\mathrm{max}} = 0.4$. The efficiencies of the "Fix" isolation criteria

[4]There are some isolation requirements in the Electron-based triggers

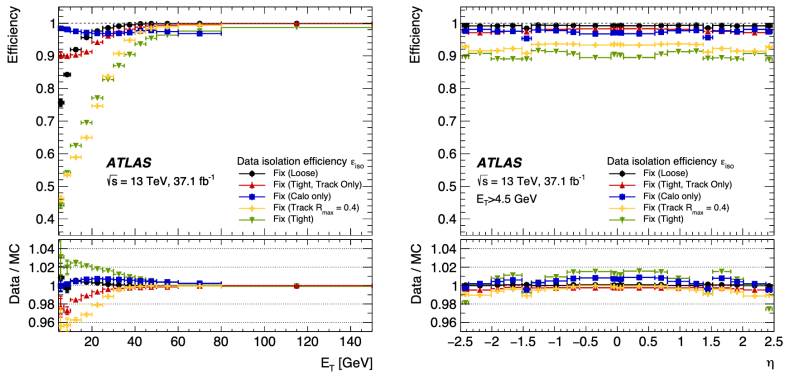


Figure 4.8: Electron isolation efficiencies for the "Fix" isolation working points plotted against electron E_{T} (left) and η (right) [125].

against electron E_{T} and η are shown in Figure 4.8. The loose version of the "Fix" working point is used in the $\gamma\gamma \to W^+W^-$ analysis, which requires that the track-based isolation variable be less than 15% of the electron p_{T} and that the calorimeter-based variable be less than 20% of the electron p_{T}.

The cumulative efficiency to find an electron for the purposes of an analysis will come from a convolution of the electron reconstruction efficiency, the identification effienciency, and the isolation efficiency. The reconstruction efficiency itself plateaus above 95% for electrons with E_{T} above 10 GeV. As can be seen in Figure 4.1, the electron energy resolution actually improves with increasing energy and is about 6% for an electron with an energy of 20 GeV.

4.3 Muons

Muon reconstruction by ATLAS involves both the ID and the MS [130]. The ID component of a muon is reconstructed just as any other charged particle would be, using the algorithm described in Section 4.1. In the MS, each chamber is initially treated separately, with a Hough transform [131] used to search for track segments. These segments are then combined across chambers to form complete MS tracks.

By combining ID and MS information, four types of muons are created: combined (CB), segment-tagged (ST), calorimeter-tagged (CT), and extrapolated (ME). For CB muons, tracking is done separately in the ID and MS, and then a global refit is performed to improve muon parameter resolution and potentially add MS segments. Normally this is done by extrapolating an MS track inwards. For ST muons, an ID track is extrapolated into the MS and matched to a segment in the CSCs or MDTs. This is used in the case that the muon's energy is too low to pass through all layers of the MS. For CT muons, a track in the ID is matched to a calorimeter energy deposits consistent with the passage of a minimum ionizing particle. The muon loses energy by ionizing detector material throughout the detector, not

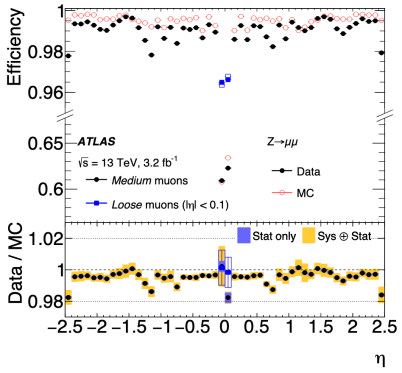


Figure 4.9: Reconstruction efficiency for medium tagged muons with $p_{\mathrm{T}} > 10$ GeV as a function of η [130]. The efficiency is determined using the tag-and-probe method in $Z \to \mu\mu$ events.

just in the ID and MS. The MS does not have good coverage for muons with $|\eta| < 0.1$, so CT tracks are useful in that region. Lastly, ME muons are created using MS information alone. At least two MS chambers must be traversed for this type of muon, which is used for $|\eta| > 2.5$, where there is no ID coverage. An overlap removal procedure is performed such that ID tracks and MS segments can only be used by one track, with the assignment preference being the order of presentation in this paragraph

Muon identification

As with electrons and tracks, muons have identification likelihood working points that seek to balance reconstruction efficiency and fake rate. For muons, there are four working points: loose, medium, tight, and high-p_{T}. These working points are formed using information such as the muon type, the agreement between the p_{T} measured in the ID and MS for a candidate muon, the χ^2 of the combination fit, and the numbers of hits in the ID or MS. For muons with 20 GeV $< p_{\mathrm{T}} <$ 100 GeV, the reconstruction efficiencies of the four working points are 98.1%, 96.1%, 91.8%, and 80.4%, in order from loose to high-p_{T}. The reconstruction efficiency for medium muons as a function of η is shown in Figure 4.9.

Muon isolation

Muons have isolation working points analogous to the electron isolation working points discussed in Section 4.2. Cuts on isolation are designed to separate muons produced directly in the hard-scatter from those resulting from a heavy-flavor hadron decay in a jet. As in the electron case, two types of isolation are defined: calorimeter-based and track-based. For muons, the track-based isolation variable is the sum of the p_{T} of tracks with $p_{\mathrm{T}} > 1$ GeV within $\Delta R = \min(10\mathrm{GeV}/p_{\mathrm{T}}^{\mu}, 0.3)$ of the muon. The calorimeter-based definition is the sum

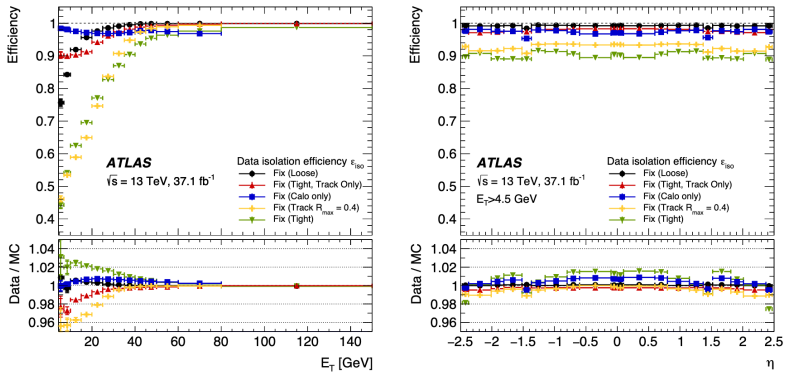


Figure 4.10: Muon isolation efficiency for the FixedCutLoose isolation working point [130].

of topological cluster energies within $\Delta R = 0.2$ of the muon, above what is expected from pileup alone. There are seven isolation working points, divided into three classes. The first class has constant efficiency in both muon p_{T} and η: LooseTrackOnly, Loose, and Tight. The second class has constant efficiency in η, but is p_{T} dependent: Gradient and GradientLoose. The last class has both η- and p_{T} - dependent efficiencies, but uses fixed cuts in the isolation variables: FixedCutTightTrackOnly and FixedCutLoose. The FixCutLoose working point is used in the $\gamma\gamma \to W^+W^-$ analysis, which requires that the track-based isolation variable be less than 15% of the muon p_{T} and that the calorimeter-based variable be less than 30% of the muon p_{T}. The isolation efficiency for this working point is shown in Figure 4.10.

Because muon reconstruction is tracking-based, the p_{T} resolution for muons gets worse with increasing p_{T}. The resolution is relatively position dependent due to the different components of the MS and magnetic field inhomogeneity. As seen in Figure 4.1, the goal for muon reconstruction is to have a resolution better than 10% for muons with $p_{\mathrm{T}} < 1$ TeV. At a p_{T} of 100 GeV, the resolution is generally better than 5%.

4.4 Jets and Missing Energy

At particle-level, jets are collimated sprays of hadrons, and are one of the most common features of events at the LHC. Roughly speaking, jets are caused by the creation of high-energy quarks and gluons. For example, if a $q\bar{q} \to q\bar{q}$ hard-scattering process takes place, then the observed final state will likely be two jets. Quarks and gluons cannot be observed as independent particles due to the running of the strong coupling constant, which results in hadronization (see Section 2.1.5). As hadronization occurs, many hadrons are created, and these are collimated by the initial boost of the underlying quark or gluon. The physics of jets is a rich and evolving subject, but they only appear in the $\gamma\gamma \to W^+W^-$ analysis as backgrounds when they are accidentally reconstructed as electrons or give rise to a non-isolated electron or muon. W bosons can decay into jets, but due to experimental considerations dis-

cussed in Chapter 6, no specific attempt to find jets is made. Because of this, the discussion of jets here is rather cursory and is primarily included for the sake of completeness.

At reconstruction-level, there is no single definition of a jet. This contrasts somewhat strongly with the rest of the experimental objects discussed above in this chapter, which essentially seek to reconstruct single particles. Traditionally, jets are found by grouping together clusters of energy in the calorimeter, and nothing in the standard model dictates exactly which clusters to include or leave out. Many grouping schemes have been developed, with context-dependent advantages and drawbacks to each. The most common jet-clustering algorithm used in ATLAS [132, 133] is the anti-k_t algorithm [134] with a distance parameter of either $R = 0.4$ or $R = 1.0$. This algorithm attempts to group energy deposits around the highest energy calorimeter clusters roughly into circles in $\eta \times \phi$ space with radii of $\Delta R = 0.4$ or $\Delta R = 1.0$. Larger radii jets capture a higher fraction of the energy of a single jet, but are more susceptible to noise, particularly from pileup. Large-R jets are often groomed to make them less sensitive to noise [135].

In ATLAS, the calorimeter energy deposits are combined into topological clusters [129], which are then the inputs for the jet-clustering algorithm [132, 133]. A jet energy scale calibration is performed to reduce the effects of pileup and potential position-dependent effects within the detector [136]. The jet energy resolution improves with energy as indicated in Figure 4.1 and can be as good as 1% for high-p_{T} central jets, though it can reach 6% for 20 GeV jets. Attempts to improve this resolution include particle-flow algorithms [137] that match tracks to calorimeter deposits and then use tracking momentum measurements rather than calorimeter energy measurements. This can improve the energy resolution for low energy jets to about 3%.

Flavor tagging

In many analyses, it is useful to try to identify the "flavor" of a jet, or the species of quark or gluon that the jet originated from. For example, identifying if a jet comes from a b quark is necessary in analyses looking for Higgs bosons decaying to $b\bar{b}$ pairs. Some of the most common types of jet tagging are described below. None of the methods are perfect, and it is not uncommon for jets to be tagged incorrectly; there is always a trade off between tagging efficiency and fake rate.

- B-tagging is the processes of identifying when a jet comes from a b quark [138]. Because b quarks form relatively long-lived B hadrons [2], b-jets often have displaced vertices. These occur because a B hadron can travel a distance of a few mm's before decaying, and the tracks of the decay products will point to where the decay occured, rather than to the primary vertex. ATLAS uses several multivariate approaches to b-tag jets.
- Charm tagging is harder than b-tagging, as charmed hadrons typically have shorter lifetimes, leading to less displaced vertices. C-jets can be identified using similar algorithms to those used for b-jets, though the efficiency is lower [138]. C-jets can also be identified by fully reconstructing charmed hadrons inside a jet.

- Using variables such as jet mass [139] can allow for jets from heavy particle decays, such as W or Higgs bosons, to be distinguished from generic QCD jets.
- Various substructure variables, such as the number of particles in a jet allow for some degree of quark vs. gluon jet discrimination [133]. Appendix E describes a few of these variables in more detail.
- τ-leptons most often decay into either one or three charged hadrons plus up to two neutral hadrons. Particle flow techniques allow for the reconstruction of these dominating decay modes [140].

Top quarks are unique among the quarks in that they do not have time to hadronize before decaying. They almost exclusively decay into a b quark and a W boson, so reconstructing these two decay products allows for the identification of top quarks.

Missing energy

Missing transverse energy, or $E_{\mathrm{T}}^{\mathrm{miss}}$, allows for the identification of unobservable particles [132, 141]. For example, neutrinos do not interact with the detector, and some models for dark matter expect dark matter particles to be produced in LHC collisions that would be undetectable after their production. These particles can be "found" if they recoil off a detectable particle. Momentum must be conserved in the plane transverse to the beamline (or in any direction for that matter), so if there is an imbalance in the measured momentum, that indicates the presence of some undetected particle.

Missing energy is the last reconstructed object mentioned in this chapter because it is determined using all of the other objects. In general, the missing energy in the x (y) direction is given by:

$$E_{x(y)}^{\mathrm{miss}} = -\sum_{i\in\{\mathrm{hardobjects}\}} p_{x(y),i} - \sum_{j\in\{\mathrm{softobjects}\}} p_{x(y),j}.$$

Here, the hard objects are electrons, photons, hadronically-decaying τ-lepton, muons, and jets, and the soft objects are tracks not used by the hard objects.

Missing transverse energy is itself treated as a vectorial quantity:

$$\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} = (E_x^{\mathrm{miss}}, E_y^{\mathrm{miss}}).$$

The magnitude of the missing transverse energy is given by

$$\sqrt{(E_x^{\mathrm{miss}})^2 + (E_y^{\mathrm{miss}})^2},$$

with direction given by

$$\phi^{\mathrm{miss}} = \tan^{-1} E_y^{\mathrm{miss}}/E_x^{\mathrm{miss}}.$$

Like jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ is not used in the $\gamma\gamma \to W^+W^-$ analysis. While the analysis *does* look for final states with neutrinos, $E_{\mathrm{T}}^{\mathrm{miss}}$ has relatively poor resolution and introduces systematic

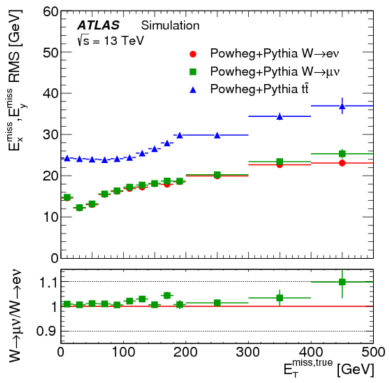


Figure 4.11: Expected $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in simulated events with a single neutrino resulting from a W boson decay [141]. The resolution is given as a function of $E_{\mathrm{T}}^{\mathrm{miss,true}} = p_{\mathrm{T},\nu}$.

uncertainties to an analysis. The expected resolution of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the case of an event with a single neutrino from a leptonic W decay is shown in Figure 4.11. The resolution metric here is the root-mean-square of the distribution of the difference between the neutrino p_{T} and the $E_{\mathrm{T}}^{\mathrm{miss}}$. The neutrinos in the $\gamma\gamma \to W^+W^-$ analysis generally have p_{T} well below 100 GeV, so the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution could range from 20 - 50%. While, $E_{\mathrm{T}}^{\mathrm{miss}}$ itself is not used in the $\gamma\gamma \to W^+W^-$ analysis, a proxy based on charged lepton kinematics is used as described in Chapter 6.

Chapter 5

Simulating ATLAS Events

To find the $\gamma\gamma \to W^+W^-$process, we must not only understand what that process will "look like" in the ATLAS detector, but we must also know what other processes could mimic the $\gamma\gamma \to W^+W^-$signal. In this context, the processes that mimic our signal would be considered as "backgrounds". To predict how any particle physics process, either signal or background, will look like in the detector, we must put together both Standard Model calculations and calculations of particle-detector interactions.

The Standard Model will tell us how often a process will occur and the probability distribution functions for particles' energies and directions. For example, if we know the initial energies and momenta of two initial state photons, then the Standard Model can tell us how likely it is for two W bosons to be produced, and how likely it is for one of the W bosons to have $0.5 < \eta < 0.55$ and $0.0 < \phi < 0.05$. An even more complicated calculation could tell us how likely it is for both of the W bosons to decay to muons and the angular and energy distributions for the muons.

Understanding how a particle will interact with the detector takes a combination of calculation and calibration. For example, it is possible to calculate a probability distribution function for the amount of energy a particle deposits in a particular segment of the detector due to ionization. It is also necessary to know how that energy would get read out from the detector; if an electrical signal is created, what is the energy to signal-height conversion?

There is also a complication from the fact that not all quantities can be computed exactly. For example, due to the strength of QCD at low energies, it is not easy to calculate the the number of of hadrons produced due to the underlying event of a proton-proton collision or the number of particles created in the hadronization process leading to jet formation. In these cases, a model is used in order to make a well-informed guess about such quantities. Similarly, parton distribution functions (PDFs) are not known with perfect precision, so different choices can be made as to the exact compositions of protons at different energy scales.

Of course, in one event at the LHC, 1,000 or more particles may be produced, and each will interact with many meters worth of both active and inactive detector material. To fully predict what a signal or background event would look like in data, each particle would

have to be completely characterized. Performing such a complicated series of calculations by hand would be practically impossible, so to make meaningful predictions, physicists turn to computer simulations [142, 2]. Sophisticated programs can generate large numbers of events using the Monte Carlo (MC) method, which can compute integrals or determine the likelihood of sequential probabilistic processes by randomly sampling from distributions. For example, for one simulated $\gamma\gamma \to W^+W^-$event, the direction of one of the Ws would be selected from a probabilty distribution function, its decay muon's direction from a different distribution, the amount of energy the muon leaves in the IBL from another distribution, etc.

In general, there are three steps involved in event simulation, though these steps can overlap. First there is a physics generator that simulates the hard-scatter process and subsequent decays that can be calculated perturbatively. Then non-perturbative physics is simulated, such as a potential underlying event or jet formation. This step relies on information from the hard-scatter. Lastly, the interactions of the final-state particles with the detector are simulated.

Of course, simulations are not perfect. Even for perturbative processes, calculations are only computed at finite order, leading to uncertainties on cross sections or distribution shapes. The models for non-perturbative physics also carry occasionally large uncertainties and may systematically mis-model distributions such as the number of particles in an underlying event. Because of this, different generators are used and compared to determine reasonable ranges of what to expect in the detector and to set numerical uncertainties. There is often a need to correct simulation by using data. For example, it is often possible to define a "control region" which contains data events that do not appear in the signal region. Then by comparing the simulation and data in the control region, a correction for the simulation can be derived that can then improve simulation predictions for signal region yields. This concept will be applied repeatedly in Part II.

Frequently, the correction takes the form of MC reweighting. When generating a simulation sample, it is not actually necessary to generate events by performing a truly random sampling from distributions. For example, if a distribution has a peak and a tail, then a random sampling from the distribution will select mostly events from the peak with few events from the tail. If the events from the tail are most interesting for an analysis, then it would be necessary to generate a large number of events to populate the tail segment of the distribution well. To avoid having overly-large simulation samples, it is possible to sample from the tail more often than from the peak. In this case, each generated event will be given a weight. Events from the peak will have a higher weight, as that part of the distribution was sampled less frequently, and events from the tail will have smaller weights. The total cross section squared would then be

$$\sigma^2 = \sum_i w_i^2, \tag{5.1}$$

where the sum is over each event, and w_i is the weight for the i-th event.

In light of the fact that simulation is not perfect, the weights generated by the MC may not be correct *a priori*. Events can be reweighted to make the distributions in MC match those in data.

In the $\gamma\gamma \to W^+W^-$ analysis, a variety of specialized corrections will be applied to the simulation, which are described in more detail in Chapters 7 - 10. Plots illustrating the agreement of data and simulation both before and after correction will be shown throughout these chapters, and the cumulative agreement of fully corrected photon-induced (Section 5.1) and QCD-induced (Section 5.2) simulation with data is shown in Chapter 11.

5.1 Simulation of Photon-Induced Processes

The primary photon-induced processes that need to be considered in this analysis are $\gamma\gamma \to W^+W^-$, $\gamma\gamma \to e^+e^-$, $\gamma\gamma \to \mu^+\mu^-$, and $\gamma\gamma \to \tau^+\tau^-$, where the first is our signal and the photon-induced dilepton processes can be backgrounds by mimicking the signal. These processes are occasionally referred to as "exclusive". As discussed in Section 2.2.3, three scenarios can occur in photon-induced events: both protons remain intact ("elastic"), one proton remains intact and the other dissociates ("single-dissociation"), or both protons dissociate ("double-dissociation"). Feynman diagrams for exclusive WW and di-lepton production are shown in Figures 2.4, 2.12, and 2.16. Because these different scenarios are sufficiently different physically and because the modelling of photon-induced physics is itself a somewhat niche field, different generators are used for the different scenarios.

In general, modelling photon-induced physics is difficult, so correcting the simulations is an important part of this analysis[1]. The corrections applied to these samples will be discussed in Chapter 10. A summary of the different generators used for photon-induced physics is given in Table 5.1.

There are some photon-induced hadron production processes, and photon-induced four-lepton processes that could act as minor backgrounds in this analysis. These were modelled with Pythia8.244 and MadGraph5_AMC@NLO [143] and found to be negligible.

For all Pythia8 and Herwig7 samples that involve parton showers, the decays of bottom and charm hadrons was performed with EvtGen 1.2.0 [144]

5.1.1 Elastic Production

Elastic production of the four considered processes was modelled at leading order (LO) using Herwig 7.1.5 [145, 146]. To model the initial state photons, the BudnevQED photon flux [147] was used, having been interfaced with Herwig using ThePEG software [148].

MG5_AMC@NLO was used as an alternate sample for the four elastic processes, with its default flux for photons radiated from protons. However, the initial state photons in this generator have no p_{T}, which is not necessarily physically accurate.

[1]For example, the only relevant generator that includes the Survival Factor effects discussed in Section 2.2.4 is SuperChic [19], which is not currently used in this analysis.

Process	Generator (+ Shower)	PDF	Use
El. WW	Herwig 7.1.5	-	Main
El. ll	Herwig 7.1.5	-	Main
SD WW	MG5_AMC@NLO+Pythia8.243	CT14QED	Alternate
SD ll	LPAIR	(built-in)	Main
SD ll	MG5_AMC@NLO+Pythia8.243	CT14QED	Alternate
DD WW	MG5_AMC@NLO+Pythia8.243	CT14QED	Alternate
DD ll	Pythia8.240	NNPDF3.1NLOluxQED	Main
DD ll	MG5_AMC@NLO+Pythia8.243	CT14QED	Alternate

Table 5.1: Summary of the different generators used for photon-induced processes considered in the $\gamma\gamma \to W^+W^-$analysis. Here, "El." indicates elastic production, and "SD" and "DD" indicate single- and double-dissociative production, respectively.

5.1.2 Single-Dissociative Production

The single-dissociative component of the four processes was modelled with MG5_AMC@NLO interfaced with Pythia8.243 for the dissociation of the proton. The default photon flux was used for the non-dissociating proton, and the CT14QED PDF [149] was used to model the photon flux from quarks for the dissociating proton. The multiplicity of particles resulting from the proton shower is not expected to be accurate in this simulation, because the Pythia showering stage is tacked on after the emission of the photon. However, the hadronization step can help correct the issues caused by the lack of initial-state photon p_{T}.

The primary generator for single-dissociative di-lepton production was LPAIR [150]. This is a generator specifically used for photon-induced physics, with its own proton showering model. However, LPAIR cannot generate the $\gamma\gamma \to W^+W^-$process, so it is not used for that component of the signal.

5.1.3 Double-Dissociative Production

Double-dissociative production can be modelled using the same MG5_AMC@NLO setup discussed for single-dissociative production, though of course both of the protons are showered using Pythia8.243, rather than just one.

For double-dissociated dilepton production, Pythia8.240 was used, with the photon fluxes given by the NNPDF3.1NLOluxQED PDF set [13].

5.2 Simulation of QCD-Induced Processes

In the $\gamma\gamma \to W^+W^-$ analysis, several QCD-induced processes can act as background. In general, these processes will be referred to as "inclusive". In the following, for all Pythia8

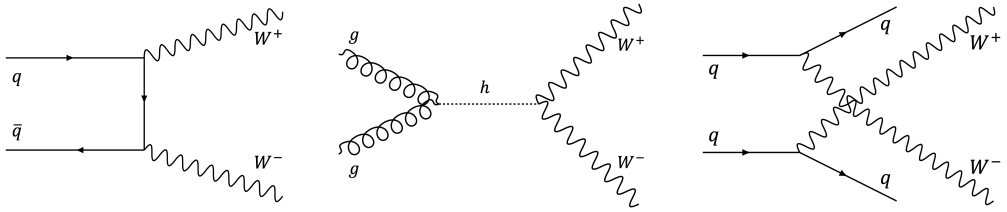


Figure 5.1: Feynman diagrams for inclusive WW production, which is the primary background in the $\gamma\gamma \to W^+W^-$ analysis. The lefthand diagram depicts the $q\bar{q} \to W^+W^-$ process; the middle diagram depicts gluon-induced WW production with an intermediate Higgs boson; and the righthand diagram depicts a possible Vector Boson Scattering diagram.

and Herwig7 samples that involve parton showers, the decays of bottom and charm hadrons were performed with EvtGen 1.2.0 [144]. A summary of the different generators used for QCD-induced physics is given in Table 5.2. The simulations of inclusive processes receive several corrections, primarily described throughout Chapters 8 and 9.

5.2.1 Inclusive WW

The primary background process is inclusive WW production, which itself is primarily $q\bar{q} \to W^+W^-$. Inclusive WW production has a total cross-section that is about 1,000 times higher than exclusive WW, and some inclusive events are expected to pass even a very efficient signal vs. background discrimination scheme (see Chapter 6). The Feynman diagram for this process is depicted in the left-most diagram of Figure 5.1. Inclusive WW can also have a qg or a gg initial state, with an intermediate Higgs boson possible in the gg case, and there is also a Vector Boson Scattering (VBS) mechanism for W^+W^- production. Feynman diagrams for the $gg \to h$ and VBS processes are also given in Figure 5.1.

The $q\bar{q} \to W^+W^-$ process was simulated at next-to-leading (NLO) order using the Powheg-Box v2 generator [151, 152, 153, 154, 155], which uses the CT10 PDF [156]. Parton showering and hadronization were performed with Pythia 8.212 as a baseline, using the AZNLO tune[2] [157] and the CTEQ6L1 PDF [158]. An alternate parton showering/hadronization simulation was provided by Herwig 7.1.6, employing the H7UE tune [159] and the MMHT2014LO PDF set [160].

Sherpa 2.2.2 [161, 162] was used as an alternate generator for the $q\bar{q} \to W^+W^-$. Sherpa performs calculations at NLO if there is up to one extra parton emitted in the hard scatter and LO if there are up to three additional partons. Sherpa's own parton shower and hadronization models were used [163, 164, 165, 166, 167, 167], though some QCD corrections

[2]The models used to simulate non-perturbative physics generally have a variety of parameters that can be set by experimenters. Designated sets of parameters are known as "tunes".

were performed with the OpenLoops 1 library [168, 169, 170]. The Sherpa sample used the NNPDF3.0NNLO PDF set [171].

The $gg \to h \to W^+W^-$ process was modelled at next-to-next-to leading order (NNLO) order using Powheg NNLOPS program [172, 173], along with the NNPDF3.0NNLO PDF set. Showering was performed with Pythia 8.240.

Vector-boson scattering induced WW production was modelled with Sherpa 2.2.2, also with the NNPDF3.0NNLO PDF set. Generic $gg \to WW$ production was modelled with Sherpa 2.1.1 and the CT10 PDF set.

5.2.2 Inclusive Di-Boson Production

The two-lepton decay of a Z boson can mimic a W decay if a lepton is missed at reconstruction level; similarly, photons can be incorrectly reconstructed as electrons. Because of this, inclusive di-boson production processes, such as WZ, ZZ, and $W\gamma$, can act as backgrounds.

The primary generator for the WZ, ZZ, $W\gamma$, and $Z\gamma$ backgrounds was Sherpa 2.2.2, calculated at NLO with the NNPDF3.0NNLO PDF set. Here again, Sherpa's own showering model was used. The minimum p_{T} for the photons in the $W\gamma$ and $Z\gamma$ was set to 7 GeV.

The Powheg-Box v2 setup, interfaced to Pythia 8.212 was also used for WZ and ZZ production, calculated at NLO, using the same settings as used for the $q\bar{q} \to W^+W^-$ simulation discussed in Section 5.2.1.

5.2.3 Drell-Yan Production

The Drell-Yan (DY) di-lepton production process, which is $q\bar{q} \to Z/\gamma* \to l^+l^-$ can also mimic signal in the case where the W bosons decay to leptons. The two l's must be the same flavor in Drell-Yan production, but if two taus are produced, the taus are not constrained to decay in the same way.

The Drell-Yan process was modelled with the Powheg-Box v1, using the CT10 PDF set. The parton showering and hadronization was performed with Pythia 8.210, once again with the AZNLO tune and CTEQ6L1 PDF. The Herwig 7 setup for showering described in Section 5.2.1 was used as an alternate. Another alternate sample was provided by the Sherpa 2.2.1 generator, using the NNPDF3.0NLO PDF set and its own showering model.

Dedicated $\tau\tau$ samples were created where the taus were forced to decay leptonically to improve statistics of this important background. These were created with the Powheg generator interfaced with Pythia 8.186 and the NNPDF3.0NLO PDF set.

5.2.4 Top Background Production

Top quark decays involve the production of an on-shell W boson, so $t\bar{t}$ and Wt processes can act as background. Both were modelled at NLO with the Powheg-Box v2 generator [174, 175] using the NNPDF3.0NLO PDF set. A diagram removal scheme was applied to prevent

Process	Generator (+ Shower)	PDF	Use
$q\bar{q} \to WW$	Powheg-Box v2+Pythia 8.212	CT10	Main
$q\bar{q} \to WW$	Powheg-Box v2+Herwig 7.1.6	CT10	Alternate
$q\bar{q} \to WW$	Sherpa 2.2.2	NNPDF3.0NNLO	Alternate
$gg \to h \to WW$	Powheg NNLOPS+Pythia 8.240	NNPDF3.0NNLO	Main
VBS WW	Sherpa 2.2.2	NNPDF3.0NNLO	Main
$gg \to WW$	Sherpa 2.1.1	CT10	Main
WZ, ZZ, $W\gamma$, $Z\gamma$	Sherpa 2.2.2	NNPDF3.0NNLO	Main
WZ, ZZ	Powheg-Box v1+Pythia 8.212	CT10	Alternate
DY ll	Powheg-Box v1+Pythia 8.210	CT10	Main
DY ll	Powheg-Box v1+Herwig 7.1.6	CT10	Alternate
DY ll	Sherpa 2.2.1	NNPDF3.0NLO	Alternate
DY $\tau\tau$	Powheg-Box v1+Pythia 8.186	NNPDF3.0NLO	Main
$t\bar{t}$, Wt	Powheg-Box v2+Pythia 8.230	NNPDF3.0NLO	Main

Table 5.2: Summary of the different generators used for QCD-induced processes considered in the $\gamma\gamma \to W^+W^-$analysis.

overlap between the two samples [176]. Pythia 8.230 was used for showering, employing the A14 tune [177] and NNPDF2.3LO PDF set.

5.3 Simulation of Pileup

Pileup interactions, or additional proton-proton interactions in an event other than the hard-scatter, are simulated using Pythia8.186 [178] and the . The decays of bottom and charm hadrons was performed with EvtGen 1.2.0 [144]. The "A3" tune of Pythia 8 [179] was used, which is a tune optimized for collisions where little momentum is transferred between the colliding protons. Lastly, the NNPDF2.3LO set of PDFs [180] was used for pileup interactions. This PDF set used Run 1 LHC data to refine predictions about proton constituents at higher energies than had previously been experimentally accessible. Pileup is included in the simulation by overlaying several events.

The number of pileup interactions and their spatial distributions are simulated based on what is expected at the LHC prior to running. For example, the majority of samples used in this analysis had beamspots with σ_z values of 42 mm, while a σ_z value of about 35 mm was actually observed for the majority of Run 2 (see Figure 3.2). It is standard to reweight the simulation to make the distribution of the number of pileup interactions in simulation events match what was actually observed in data. In the $\gamma\gamma \to W^+W^-$analysis, the spatial distribution of pileup interactions is also corrected (see Chapter 7).

5.4 Modelling the ATLAS Detector and Pileup

In ATLAS, the final step of simulation is the same for all generated events, both signal and background. Events are passed through a GEANT4 [181] simulation of the ATLAS detector [182] or a combination of GEANT4 and parameterized calorimeter response for the sake of increasing simulation speed [183]. This simulation involves a detailed spatial layout of the entire detector, with the chemical composition of both active and inactive material specified. This step also produces simulated activity in the detector, such as hits in the ID or MS. Using these simulated hits, events and their corresponding objects are reconstructed using the same software that is used to process data.

Part II

Observation of $\gamma\gamma \to W^+W^-$

Chapter 6

Analysis Strategy

The $\gamma\gamma \to W^+W^-$ analysis can essentially be classified as a "cut-and-count" analysis, which is a straightforward way to establish an observation. These generally proceed as follows.

1. A "signal region" (SR) is defined, which is a set of event selections designed to emphasize signal with minimal background contamination. The event selections are based on the objects reconstructed in the events and their kinematic properties, with cuts that are optimized using simulations as discussed in Chapter 5. Cut-and-count analyses are typically "blinded", meaning that experimentalists avoid looking at data events in the SR to avoid biasing how they develop the cuts.

2. Having defined a SR, the number of background events in the SR is estimated using a combination of simulation and data in "control regions" (CR). Control regions are similar to, but do not overlap with, the SR, instead emphasizing backgrounds. They allow for the validation of simulations.

3. The SR is then unblinded, and the number of signal events in the SR ($N_{\text{sig,SR}}$) is taken to be
$$N_{\text{sig,SR}} = N_{\text{data,SR}} - N_{\text{bkgd,SR}}. \tag{6.1}$$
Here, $N_{\text{data,SR}}$ and $N_{\text{bkgd,SR}}$ are the number of data and background events in the SR, respectively.

4. The significance of the observation is determined based on the statistical and systematic uncertainties on the data yield and background estimates in the SR. The cross-section of the signal is given by the formula
$$\sigma = \frac{N_{\text{sig,SR}}}{\mathcal{L} \cdot \varepsilon_S \cdot A}. \tag{6.2}$$
Here, $\mathcal{L}$ is the integrated luminosity of the data, ε_S is the expected efficiency of selecting the signal, and A is the the expected kinematic acceptance of the signal. The rate of signal production can be compared to the Standard Model prediction based on simulations of the signal.

This chapter details how data and simulation events are selected in the $\gamma\gamma \to W^+W^-$ analysis. In particular, the question of how to discriminate photon-induced processes from QCD-induced processes is addressed. At the end of this chapter, a road map is given for the rest of Part II, describing the various corrections that will need to be made to ensure accurate estimates of $N_{\text{bkgd,SR}}$. Precise definitions of the main SR and CRs will be given in Chapter 11.

6.1 Data Used

The data used in this analysis comes from Run 2 of the LHC (see Chapter 3). In Run 2, protons were collided at a center-of-mass energy of $\sqrt{s} = 13$ TeV, and the ATLAS collaboration recorded a total integrated luminosity of $\mathcal{L} = 139.0 \pm 2.4$ fb^{-1} after data quality selections were applied to remove data when detector components were non-functional[1] [83]. The length of the beamspot in the z direction and the number of pileup interactions are both important features for this analysis; these are described in more detail in Sections 3.1.2 and 3.1.4, respectively. In Run 2, the beamspot length was about 40 mm long, with an average of about 34 pileup interactions per event.

Photon-induced physics can also be studied in Lead-Lead (PbPb) collisions data [35, 38] (see Section 2.3.1). In Run 2, ATLAS collected an integrated luminosity of $\mathcal{L} = 1.76$ nb^{-1} of PbPb data at a nucleon-nucleon center-of-mass energy of 5.02 TeV [82]. There was essentially no pileup in this dataset. While the total integrated luminosity is much lower than that for proton-proton running, the per-hadron photon flux is much higher. In EPA (see Section 2.2.2), each hadron adds a factor of Q^2 to the photon flux, where $Q = 82$ is the atomic number of a Lead ion. However, the energy of the photons produced by Lead ions at the LHC tends to be lower than those produced by photons. As can be seen from Figure 6.1, in PbPb collisions, there are not many photon-photon interactions that occur with center-of-mass energy above 80 GeV, which is 80 GeV shy of the center-of-mass energy needed to produce two W bosons [39]. Because of the lack of *photon* luminosity at high-energy during PbPb data-taking, this dataset is not used.

6.2 Event Selection

The defining features of the elastic $\gamma\gamma \to W^+W^-$ process are the two final state W bosons with no additional underlying event. Even in the single- and double-dissociative cases, little central activity is expected in the detector from the dissociating protons. Each of the W bosons can either decay into a lepton and a neutrino or into two quarks, which will result in either one or two jets per W.

In the case when both of the W's decay into leptons, the only central activity expected from the $\gamma\gamma \to W^+W^-$ process are the two charged leptons. To re-iterate, that is the unique signature of two leptons emerging from the same point of origin with no other tracks nearby.

[1] Please see Section 3.1.3 for additional details on the luminosity.

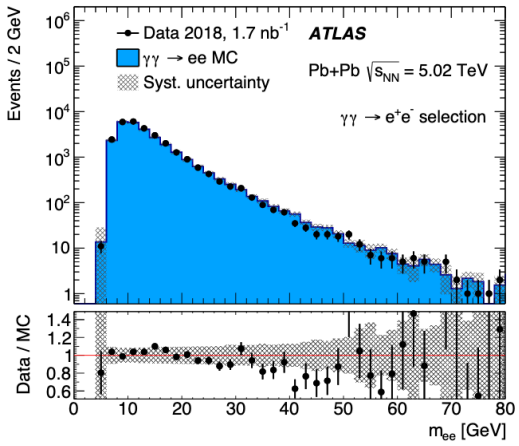


Figure 6.1: Number of photon-induced e^+e^- events observed as a function of m_{ee} in 1.76 nb^{-1} of PbPb data collected by ATLAS during LHC Run 2. There are few events with a photon-photon center-of-mass energy above 80 GeV, making WW production unobservable in this dataset.

Furthermore, these leptons are not expected to be back-to-back, as each lepton must balance its respective neutrino in the W decays. Another unique possibility is that one W can decay into an electron and the other into a muon. Experimentally, an isolated $e + \mu$ final state should only occur for processes with two final state W bosons, two final state τ leptons, or when leptons are missed or reconstructed incorrectly. For each W boson, there is a 10.8% chance that it decays into an electron, a 10.6% chance that it decays into muon, and an 11.3% chance that it decays into a tau [2]. Therefore, there is a 10.7% chance that both W bosons decay leptonically and a 2.3% chance that the WW decay results in an $e + \mu$ final state.

In the case where one or both of the W bosons decays into jets, the experimental signature is less distinct. Jets are comprised of many hadrons, making it hard to distinguish underlying event particles from W-decay particles. Jets are the most commonly produced objects at the LHC, so backgrounds would be significant if W-to-jet decays were used in this analysis.

The case where both W bosons decay leptonically is favored in this analysis, as it allows for a cleaner distinction between photon-induced and QCD-induced processes. And while the $e + \mu$ final state is relatively rare, the low background rates make it the most promising channel to observe the $\gamma\gamma \to W^+W^-$ process. However, ee and $\mu\mu$ events are also usedful in this analysis to study background modelling. The selection of relevant data events is detailed in the following subsections. Simulated events are required to pass the same selections.

Selection	Electrons	Muons
E_{T} or p_{T}	$E_{\mathrm{T}} > 20$ GeV	$p_{\mathrm{T}} > 20$ GeV
η	$\lvert\eta\rvert < 1.37$ or $1.52 < \lvert\eta\rvert < 2.47$	$\lvert\eta\rvert < 2.4$
Identification	Medium	Medium
Isolation	Fix: loose	FixCutLoose
$\lvert d_0\rvert/\sigma_{d_0}$	< 5.0	< 3.0

Table 6.1: Summary of the selections that electrons and muons must pass to be considered in the $\gamma\gamma \to W^+W^-$ analysis.

6.2.1 Triggers

The single electon and single muon triggers, which were discussed in more detail in Section 3.3.1, are used to select events in this analysis. For the single electron triggers, the minimum E_{T} was 24 GeV in 2015 and 26 GeV in 2016-2018. In 2015, trigger electrons were required to pass a medium identification criteria[2] if the E_{T} was below 120 GeV or a loose requirement if above 120 GeV. In 2016-2018, a tight requirement was employed below 60 GeV, a medium requirement between 60 and 140 GeV, and a loose requirement above 140 GeV. For the single muon triggers, the minimum p_{T} was 20 GeV in 2015 and 26 GeV in 2016-2018. For all years, the trigger muon had to be considered isolated[3] if its p_{T} was below 50 GeV.

A trigger matching scheme was employed to ensure that at least one of the fully reconstructed leptons in the event was the object that triggered the event readout.

6.2.2 Lepton Selections

Electrons are reconstructed as detailed in Section 4.2. They are required to have $E_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.47$. They are also rejected if they fall in the region $1.37 < |\eta| < 1.52$, which is the transitional region between the barrel and endcaps of the EM calorimeter. For this analysis, electrons are required to pass the medium identification criteria and the loose version of the "Fix" isolation criteria. The significance of the d_0 value, $|d_0|/\sigma_{d_0}$, must be less than 5.0 for electrons.

Muons are reconstructed as detailed in Section 4.3. They are required to have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$. They are also required to pass the medium or high-p_{T} identification criteria and the FixCutLoose isolation criteria. The significance of the d_0 value, $|d_0|/\sigma_{d_0}$, must be less than 3.0 for muons.

The selection criteria for leptons are summarized in Table 6.1.

[2]similar to that discussed in Section 4.2

[3]using criteria similar to those discussed in Section 4.3

6.2.3 Additional Selections

Beyond the generic single-lepton selections, additional kinematic selections are used to create a final event-level pre-selection:

- Exactly two leptons in the event which pass the criteria discussed in Section 6.2.2.
- The two leptons must have opposite charge.
- Both leptons must have a measured z_0 value within 0.5 mm of the "lepton vertex" (z^{ll}_{vtx}). Because photon-induced do not produce charged particle tracks from an underlying event, generic ATLAS vertex reconstruction is insufficient for identifying where in z the interaction occurred. The algorithm tends to associate pileup tracks with the lepton vertex, biasing the position measurement. Additionally, using the standard method, the vertex associated with the leptons is occasionally not considered the primary vertex, as it does not always have the highest $\sum p_{\text{T}}$ in the event. A selection based on the primary vertex would therefore result in lost efficiency. To address these issues, a custom vertex definition was created, which only depends on the lepton kinematics:

$$z^{ll}_{\text{vtx}} = \frac{z_0(l_1)\sin^2\theta(l_1) + z_0(l_2)\sin^2\theta(l_2)}{\sin^2\theta(l_1) + \sin^2\theta(l_2)}. \tag{6.3}$$

  Here, the θ values are the angles between the leptons and the beamline, as discussed in Section 3.2.1. The use of $\sin^2\theta$ serves as a proxy for the resolution of the z_0 of the leptons, giving leptons with a better resolution (those with θ closer to $\pi/2$) a higher weight.

- The lepton with higher p_{T}, or the "leading lepton", must have $p_{\text{T}} > 27$ GeV.
- The lepton with lower p_{T}, or the "subleading lepton", must have $p_{\text{T}} > 20$ GeV.
- The invariant mass of the di-lepton system must be greater than 20 GeV ($m_{ll} > 20$ GeV). Modelling di-lepton final states with lower invariant mass is difficult due to the presence of hadronic resonances that can decay to two leptons.

Tracks considered in this analysis must pass the "tight-primary" selection discussed in Section 4.1. They must additionally have $|d_0| < 1$ mm measured with respect to the beamline. This reduces likelihood of considering tracks from secondary vertices or material interactions.

There are no selections in this analysis based on jets or missing energy. The use of these objects would introduce uncertainties to the analysis due to calibration and pileup. While a jet veto could eliminate some backgrounds and neutrinos from W decays result in $E^{\text{miss}}_{\text{T}}$, proxies for both can be implemented using lepton kinematics and tracking information. In lieu of a jet veto, a track veto is implemented as described below, and the p_{T} of the di-lepton system, $p_{\text{T},\ell\ell}$, can be used instead of $E^{\text{miss}}_{\text{T}}$.

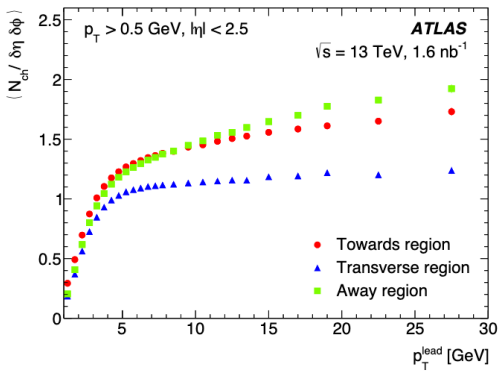


Figure 6.2: Average angular density of charged particles with $p_{\mathrm{T}} > 500$ MeV and $|\eta| < 2.5$ produced in proton-proton collisions in LHC Run 2 as a function of the leading charged particle's p_{T} [184]. The data was collected using ATLAS's minimum bias triggers [108], which is a small dataset that does not rely on the presence of a particular object. The densities are given per unit of η and ϕ in radians, so the total area covered is 5 units of η and 2π units of ϕ. The densities are broken down into those within $\pm 30°$ of the leading charged particle (Towards region), those within $\pm 30°$ of the direction exactly opposite of the leading charged particle (Away region), and the remaining $240°$ of the transverse plane (Transverse region).

6.3 Discriminating Photon-Induced and QCD-Induced Physics

The primary means to distinguish a photon-induced event from a QCD-induced event is the presence of an underlying event (UE) or lack thereof. The UE will produce charged particles, and therefore tracks that point to the interaction location. The UE will also produce neutral particles, though these do not result in pointing information.

The number of charged particles that are produced in a typical proton-proton collision depends on the momentum transferred between the two protons and therefore on the p_{T} of the leading charged particle. The typical angular density of charged particles as a function of the leading charged particle's p_{T} is illustrated in Figure 6.2 [184]. For collisions with a leading charged particle with $p_{\mathrm{T}} > 27$ GeV, such as those required by the $\gamma\gamma \to W^+W^-$ analysis, more than one charged particle is expected per unit of η and ϕ, representing an average of over 30 charged particles throughout the tracking region.

A photon-induced interaction will have 0 UE tracks in most cases, though there is a small chance of having a track if at least one proton dissociates. This suggests that a powerful means of photon-induced vs. QCD-induced collision discrimination is simple track counting. If there are UE tracks (i.e. those not matched to a lepton), then the event can be classified as background-like, and if there are no tracks, then the event can be classified as signal-like.

Of course, this classification scheme is not perfect. While the average number of tracks in a QCD-induced event is significantly more than 0, some small fraction of events[4] will have no reconstructed tracks.

Additionally, there is an important factor that has been ignored so far in this section: pileup (PU). The tracks in the events considered for the $\gamma\gamma \to W^+W^-$ analysis will not all be from the collision that yields the WW pair, as PU collisions will also create tracks. This is illustrated in Figure 6.3 [185]. It should be noted that this simulated event only had 23 simultaneous collisions, which is below average for a data event in Run 2. The presence of PU tracks, along with the non-zero z-resolution for tracks (illustrated in Figure 4.5), complicates the track counting metric for signal vs. background discrimination. If there were no PU collisions in the events considered for the $\gamma\gamma \to W^+W^-$ analysis, then one could simply count all tracks in the event for discrimination purposes. However, given the presence of PU, no data event would pass that all-track inclusive selection. While this means the background rate would be 0, which is a good thing, it also means that the signal selection efficiency would be 0, which makes an observation impossible. Because of this, tracks must only be counted in a restricted region, or an "exclusivity window". However, we must determine how wide this window should be.

The truth-level charged particles produced in a collision will originate in a space that is about the size of a proton. Of course, tracking resolution is not so fine, being about 100 μm at best, so most UE tracks will point to within about ± 100 μm from the truth-level interaction location. If a very small window size is used, say on the order of ± 1 μm away from the leptons, then a large fraction of QCD-induced events would pass the selection, leading to overwhelming background rates and, once again, no observation.

Setting a window size is a balancing act: too small a window increases background rates[5], and too large a window reduces signal acceptance rates[6]. A window centered on the lepton vertex, z^{ll}_{vtx}, that has a size of ± 1 mm has been used in other photon-induced physics analyses with ATLAS in proton-proton events [41, 21], providing roughly optimal expected signal significance. The ± 1 mm window size is adopted in this analysis, though attempts to optimize this selection are discussed further in Appendix A. An illustration of how PU can can cause photon-induced events to fail a window-based exclusivity selection is given in Figure 6.4.

It is worth noting here that one of the key appeals of performing photon-induced physics analyses with PbPb collision data is a lack of pileup in such events [38, 35]. Track counting is still the metric used to discriminate signal and background events, but no exclusivity window is needed. The PbPb dataset is not used in the $\gamma\gamma \to W^+W^-$ analysis for the reasons highlighted in section 6.1.

[4] About 0.1% in the case of $q\bar{q} \to W^+W^-$.

[5] When the window is too small, UE tracks may point outside of the window and therefore not aid in discrimination.

[6] PU tracks will point into the window if it is too large, leading to signal rejection.

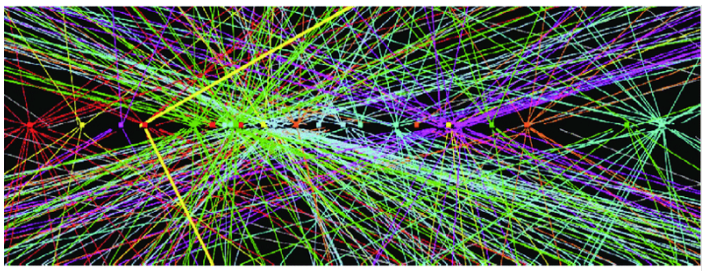


Figure 6.3: Illustration of the pileup tracks in a simulateed event with 23 simultaneous proton-proton interactions [185]. The hard scatter in this event is a Drell-Yan $\mu\mu$ event, which occurs where the two bold yellow lines are pointing. The thin red lines also emerging from that vertex are considered UE tracks in this context, and all additional tracks are considered pileup tracks.

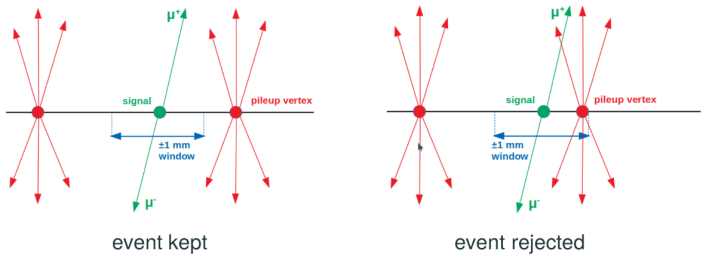


Figure 6.4: Illustration of how pileup (PU) can spoil the ±1 mm exclusivity window. On the left, all PU tracks are sufficiently far from the lepton vertex that the event passes the exclusivity selection. On the right, however, PU tracks fall within the exclusivity window, causing the photon-induced event to fail the exclusivity selection.

6.4 Overview of Simulation Shortcomings

At this point, the path forward should be relatively clear. We have simulations of our signal and background processes and have defined event and object selections that are used in both simulated and real data events. What remains is to define a signal region, estimate the expected yields in that region using our simulation, and then see if the data agrees with our expectation. The exact definition of the signal region will be informed by expectation from simulation, and it will depend on the number of tracks counted in the exclusivity window and on the $p_{\mathrm{T},\ell\ell}$ of the di-lepton pair. Before unblinding the analysis, it will also be useful to define control regions (also by using the number of tracks and $p_{\mathrm{T},\ell\ell}$) to check that the simulations accurately predict yields in regions that are similar to, but not exactly the same as, our signal region.

However, before proceeding, we must address known flaws in the simulation. If the simulation is taken as-is, the expected yields will be incorrect, making steps 3 and 4, as laid out in the beginning of this chapter, inaccurate. Chapters 7-11 will detail the procedures used to correct simulation or otherwise address its shortcomings, but a brief summary of the needed corrections is given here to contextualize and motivate the next chapters.

- Pileup is not modelled correctly *a priori*; the density of interactions can be incorrect, as can the number of tracks produced per interaction. Understanding the ε_S in Equation 6.2 relies on an accurate estimation of how likely it is for PU to cause a signal event to fail the exclusivity selection (see Figure 6.4). At its core, the correction applied to address this issue seeks to make the distribution of the number of tracks in the ± 1 mm window around the leptons correct.

- Underlying events in QCD-induced events are not modelled correctly. The multiplicity of tracks produced in a typical UE depends on physics in the non-perturbative regime of QCD. This makes modelling this distribution challenging. However, we need to know what fraction of QCD-induced events have 0 UE tracks in order to estimate $N_{\text{bkgd,SR}}$.

- The p_{T} distribution of the Z-boson in Drell-Yan di-lepton samples is not modelled correctly. We will eventually put a cut on the $p_{\text{T},\ell\ell}$ of the di-lepton system; mismodelling of the Z-boson's p_{T} will influence how many such events pass that cut.

- The rate at which photons or jets mimic electrons or muons is not necessarily modelled correctly. For example, we need to know what fraction of $W + \gamma$ events will fall into our signal region because the photon was accidentally reconstructed as an electron. If the $\gamma \to e$ fake rate is not correct in simulation, then the estimate of the $W + \gamma$ events in the signal region will be incorrect.

- Modelling of dissociative photon-induced events is not accurate. As in the case of UE track multiplicity, proton dissociation relies on non-perturbative QCD physics, making calculations difficult. Additionally, the single- and double-dissociative WW process has not been implemented in many physics generators, so it is not easy to cross-validate simulation. Because of this, simulation is not used to directly estimate the single- and double-dissociative WW rates in the signal region. We instead rely on a data-driven estimate of the rates of dissociation.

- Cross sections in data are not necessarily exactly accurate. Calculations are not performed at all orders, leading to uncertainty on the final cross-sections of major backgrounds. Our main control regions will be used to check the yields of background processes in simulation vs. data.

In all of these cases, data-driven methods are used to address the shortcomings of simulation. By definition, the data can never be wrong.

An additional correction is applied for the p_{T} of the WW system in $q\bar{q} \to W^+W^-$ samples. This is not a data-driven method, but rather will rely on a higher-order calculation of the process. Similar to the Z-boson p_{T} issue highlighted above, this correction is relevant due to the $p_{\mathrm{T},\ell\ell}$ cut that will eventually be applied to the final state di-lepton system.

Furthermore, standard ATLAS scale factors and reweighting are applied to the simulation samples to improve the data/simulation agreement due to lepton trigger rates, lepton reconstruction efficiency, and lepton identification and isolation criteria [130, 186, 113, 112].

Chapter 7

Pileup Modelling Corrections

Tracks from pileup (PU) interactions can fall within the $\Delta z = \pm 1$ mm window around the lepton vertex either due to the simple fact that the PU interaction occurs within the window or due to track z-resolution effects. Regardless of *why* the PU track is in the window, it will cause the event to fail the exclusivity selection, reducing our signal-selection efficiency.

Correctly modelling the multiplicity and spatial distribution of PU tracks is important to the success of the $\gamma\gamma \to W^+W^-$ analysis; it improves our estimate of the number of expected photon-induced events in the signal region. It also improves our estimate of $N_{\text{bkgd,SR}}$, as some QCD-induced events with no UE tracks will fail the exclusivity selection due to PU. While an estimate of the expected number of signal events in the signal region is not strictly needed in order to measure the cross-section, it *is* needed in order to compare the measured cross-section to the SM expectation

The critical feature here is PU track density, which is related to PU interaction density. To correct the PU in simulation, the following steps are performed:

1. The simulation is reweighted such that the distribution of the number of PU interactions per event matches the distribution observed in data.

2. The tracks in the simulation are shifted such that the longitudinal beamspot size will match that observed in data. Together with the first point, this will make the density of PU interactions in the simulation match data.

3. A last reweighting is performed to make the distribution of the number of tracks counted in the ± 1 mm window match the distribution in data. This correction is necessary because the number of tracks produced per PU interaction is not modelled correctly in the simulation. Therefore, the first two points, which should get the density of *interactions* correct, are not sufficient.

The correction indicated in the first point above is standard to ATLAS analyses. The end result is that a plot of the mean number of interactions per bunch crossing in the simulation would look like that in Figure 3.4 on a per-year basis. The second two points are unique to this analysis and are detailed further in the following sections.

7.1 Beamspot Length Correction

The number of PU interactions and their longitudinal distribution in simulation is not correct out of the box because simulation is normally generated before data is taken. This means that those distributions in the MC are set by expectations rather than reality. The positional distribution of interactions in the z-direction is taken to be Gaussian, so the main metric used to describe the beamspot length is σ_{BS}. The simulation used in this analysis had beamspot widths, or $\sigma_{\text{BS}}^{\text{MC}}$, of 35 mm or 42 mm depending on the sample. The values observed in data are shown on a run-by-run basis in Figure 3.2. In 2015, the width, or $\sigma_{\text{BS}}^{\text{data}}$, was centered around 44 mm, and in 2016-2018 the value was centered around 34-38 mm depending on the run. The beamspot size changes from run-to-run due to changes in the LHC beam optics.

In data, each run has a run number, and the $\sigma_{\text{BS}}^{\text{data}}$ for each run is a known quantity. In the simulation, each event is randomly assigned a pseudo-run number, such that distribution of the number of events in each run matches between the simulation and data. Because of this internal assignment, each simulated event has both a known $\sigma_{\text{BS}}^{\text{MC}}$, which is either 35 or 42 mm, and a known $\sigma_{\text{BS}}^{\text{data}}$, which corresponds to the beamspot width actually observed in that particular run.

When correcting the beamspot length, our goals are:

- To make the z-distribution of hard-scatter (HS) interactions match what is expected in data.
- To preserve the simulated spatial distribution of UE tracks[1] relative to the HS in events with a UE.
- To correct the density of PU tracks.

The steps to accomplish these goals are as follows:

1. In each simulated event, the lepton vertex, z_{vtx}^{ll}, is multiplied by the value $\sigma_{\text{BS}}^{\text{data}}/\sigma_{\text{BS}}^{\text{MC}}$. This rescales the HS distribution. The value $\Delta z = z_{\text{new}} - z_{\text{old}}$ is recorded.
2. Reconstructed tracks that are matched [119] to charged particles originating in the HS UE are shifted by the same amount as z_{vtx}^{ll} was shifted, Δz. This avoids squeezing or broadening the distribution of UE tracks around the HS.
3. For tracks that aren't matched to the HS UE, their measured z value is multiplied by $\sigma_{\text{BS}}^{\text{data}}/\sigma_{\text{BS}}^{\text{MC}}$. The density of PU interactions based on the new z information of tracks should match that in data.

All of these rescalings introduce uncertainties into the analysis[2]. For example, the efficiency of track and lepton reconstruction is mildly position dependent, and this correction does not seek to fine tune the simulation to account for such effects.

[1] Which should not be confused with PU tracks.

[2] See Chapter 12 for more details

7.2 Pileup Track Multiplicity Correction

Having applied the reweighting based on the mean number of pileup interactions and rescaled the tracks' and HS's location within each event, the physical distribution of PU interactions in the simulation should closely match that in data. What is still not necessarily correct at this point is the number of tracks resulting from each PU interaction. This will affect the distribution of the number of tracks counted in the ± 1 mm window around the lepton vertex, $n_{\text{trk}}^{\text{PU}}$. A data-driven correction is derived to reweight events based on the number of PU tracks counted within the exclusivity window. The derivation proceeds as follows.

First, it is important to note that the position of a HS interaction, either signal or background, is essentially random with respect to the PU distribution. This is because the PU interactions and the HS interaction are totally independent of each other. Because of this, the distribution of $n_{\text{trk}}^{\text{PU}}$ in a window around the leptons should be the same as that found in a window around a random location in the beamspot[3]. This distribution can be measured in a straightforward way.

An auxiliary measurement is performed using data events that pass all of the preselections described in Section 6.2.3, with the additional selections that both leptons have the same flavor and 70 GeV $< m_{\ell\ell} <$ 105 GeV. This corresponds to the Z-peak region in data, where the vast majority of events will have resulted from the DY di-lepton process with a nearly on-shell intermediate Z-boson. For each data event, multiple randomly centered, non-overlapping ± 1 mm windows are placed along the beamline, and $n_{\text{trk}}^{\text{PU}}$ is recorded for each window. The distribution is recorded as a function of the position of the window center, as the density of PU tracks is position depended. Further, the window must be centered at least 10 mm away from z_{vtx}^{ll} to avoid accidentally counting UE tracks. An illustration of how random windows would be positioned around the lepton vertex is given in Figure 7.1.

The $n_{\text{trk}}^{\text{PU}}$ distribution observed in data for random windows is shown in Figure 7.2, averaging over all positions and numbers of pileup interactions. On average, 52.6% of random windows have no PU tracks, or $n_{\text{trk}}^{\text{PU}} = 0$. The figure also shows the distributions as predicted in simulation, both before and after the beamspot length correction. The simulation before the correction, presented in solid lines, has clear mismodelling in all bins of $n_{\text{trk}}^{\text{PU}}$, especially in the case of a 42 mm beamspot for the simulation. We can note here that the majority of the simulated samples was produced with the 42 mm beamspot, so the importance of this correction should be clear. After the beamspot length correction, the $n_{\text{trk}}^{\text{PU}} = 0$ bin is modelled relatively well, indicating that the density of PU interactions has been corrected, but the higher $n_{\text{trk}}^{\text{PU}}$ bins still show mismodelling. The ratio of the data to the corrected simulation is taken on a bin-by-bin basis. This ratio is the reweighting that will make the $n_{\text{trk}}^{\text{PU}}$ distribution agree between data and simulation. The distributions resulting from the application of this correction are also shown in Figure 7.2. For the sake of visualization, these distributions have been integrated over the all window positions. The density of PU

[3]Assuming the center of the random window in each event is selected based on the same distribution that determines the lepton vertex location

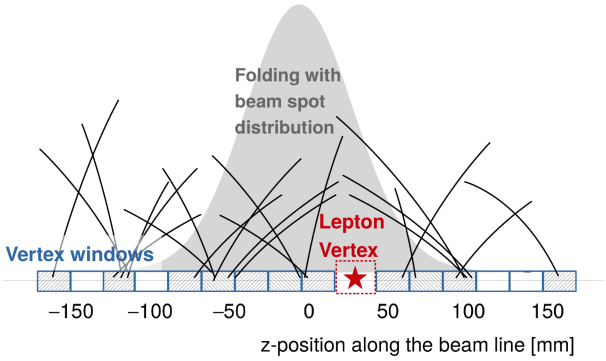


Figure 7.1: Illustration of the method used to measure the expected number of pileup tracks in a random ±1 mm window around a lepton vertex [187]. The correction is derived in a window-center position dependent way.

tracks depends on the location within the beamspot, so the reweighting is actually derived on a window-position dependent basis.

7.3 Exclusivity Window Selection Efficiency

To conclude this chapter, it is worth quantifying the impact that pileup tracks will have on the the signal selection efficiency. A photon-induced event should have no UE tracks, and after accounting for the effect of our various preselections on efficiency, the only other effect that would cause a signal event to fail an $n_{\text{trk}} = 0$ selection[4] would be the presence of a PU track in the ±1 mm window around the leptons. From Figure 7.2, it can be seen that about 50% of randomly placed windows should have no tracks in them, meaning that 50% of signal events should pass the window-based exclusivity selection. This value is averaged over the beam conditions throughout the entirety of LHC Run 2.

Of course, the number of PU interactions per event will strongly affect this selection efficiency. This dependence is illustrated in Figure 7.3. At the lowest PU rates observed in Run 2, the efficiency is as high as 87% in data, and at the the highest PU rates, it drops to 25%. The average efficiency over Run 2 is 52.6%.

[4]Which will be applied in our signal region

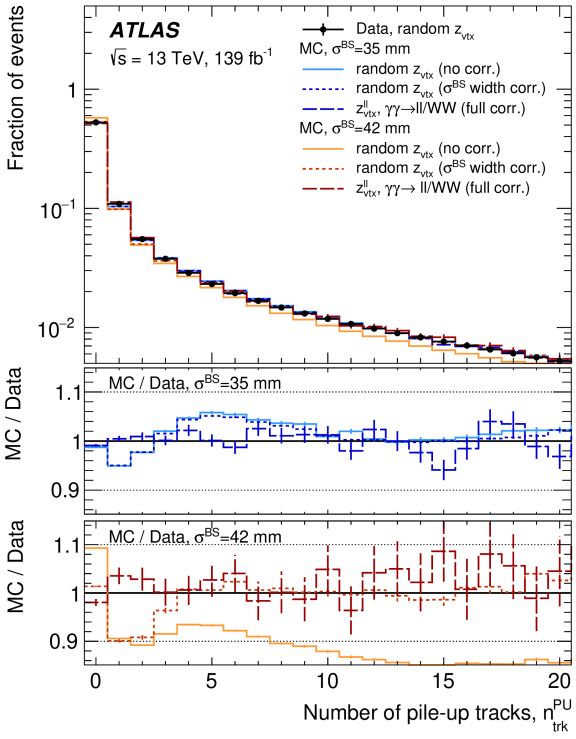


Figure 7.2: Distribution of the number of pileup tracks counted in random ±1 mm windows in both data and simulation [187]. The data is presented in black cicles, and the simulation is shown with either solid or dashed blue or orange/red lines. The blue lines comes from simulation where a 35 mm beamspot length was used, and the orange/red lines come from simulation where a 42 mm beamspot length was used. The solid lines indicate the simulation distributions in the case where no beamspot length correction was applied. The shorted dashed lines represent the case where the beamspot length correction is applied, but the $n_{\text{trk}}^{\text{PU}}$ distribution correction is not applied. The longer dashed lines correspond to the cases where the $n_{\text{trk}}^{\text{PU}}$ correction *is* applied, so by definition the MC/Data ratio plots should be flatter and closer to 1 for these distributions.

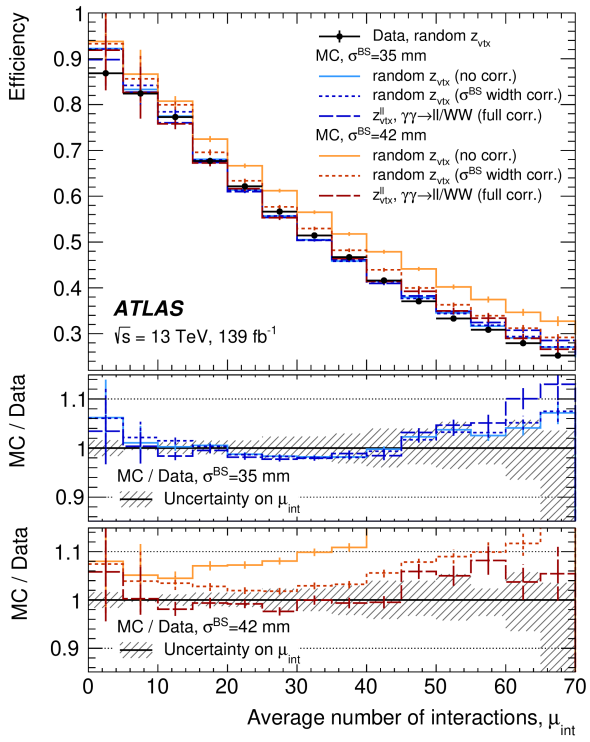


Figure 7.3: Efficiency for a randomly placed ±1 mm window to pass the exclusivity selection, having $n_{\text{trk}}^{\text{PU}} = 0$, as a function of the number of simultaneous proton-proton interactions in the event [187]. The data and simulation presentation is the same as described in the caption for Figure 7.2.

Chapter 8

Underlying Event Modelling Corrections

For QCD-induced processes, mismodelling of the distribution of the number of charged particles, and therefore tracks, due to the underlying event (UE) is a known problem [188, 184, 189, 190]. In fact, the modelling of the distribution is particularly poor at lower multiplicities, which is the region of interest for the $\gamma\gamma \to W^+W^-$ analysis[1]. QCD-induced events with few UE charged particles can end up in our signal region and control regions, so to make accurate predictions of the number of background events, the simulation must be corrected. The difficult question, of course, is how to derive this correction.

The first thing to understand is how UE particles are produced. The majority of these particles are created in multi-parton interactions at relatively low energy scales. In contrast, the $q\bar{q} \to W^+W^-$ process that leads to our primary background occurs at a high energy scale, as the quarks involved in this interaction must have at least enough energy to create two on-shell or nearly on-shell W bosons. The $q\bar{q} \to W^+W^-$ process, which results in a color-neutral final state, is effectively decoupled from the rest of the proton-proton interaction due to the asymptotic freedom of QCD and due to the different time scales of the $q\bar{q} \to W^+W^-$ process and the UE. Therefore, we can derive our correction using a different quark-induced process that results in a color neutral final state – the UE track multiplicity should be similar in all such processes that have similar momenta transfers [190].

However, it should be noted that the track multiplicity will depend on the p_{T} of the initial $q\bar{q}$ system, or equivalently the p_{T} of the color-neutral final state system. Such dependence is illustrated in Figure 8.1, and discussed in [188] and [189]. Therefore, the correction should also be derived as a function of the p_{T} of the hard-scattering system.

The best candidate for a region rich in $q\bar{q}$ induced processes is the Z-peak region. Here, the $q\bar{q} \to Z$ process dominates, giving ample statistics for a derivation of a UE charged particle correction with relatively little data grooming required. That being said, several

[1]This discrepancy can be seen in Figure 8.3, though the derivation of the distributions in this figure are described throughout Sections 8.2, 8.3, and 8.4.

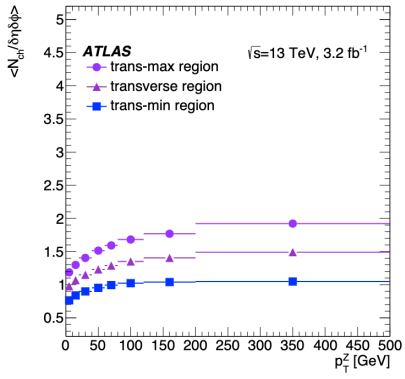


Figure 8.1: Dependence of the density of charged particles produced in a $q\bar{q} \to Z$ event on the p_{T} of the Z boson [188]. The density is presented as the number of charged particle per unit of pseudorapidity and ϕ in radians. The densities are given in the transverse plane regions defined in the caption of Figure 6.2.

steps are still required to produce a correction:

1. The background from photon-induced processes must be subtracted[2]. Photon-induced di-lepton events are expected to comprise about 5% of the total yield of events in the Z-peak region with n_{trk}= 0 (where n_{trk} is the number of non-lepton tracks counted in the $\pm$1 mm window around the leptons).

2. Pileup tracks must be subtracted from the window. To do this in data, we must determine the expected number of pileup tracks in a given random window.

3. Unfolding is performed to probabilistically determine the number of truth-level charged particles that caused the observed number of tracks. Due to track reconstruction inefficiencies and resolution, the counted number of tracks will likely be less than the actual number of truth particles created in the UE.

4. A reweighting is then derived for the simulation based on the truth-level number of charged particles and the truth-level p_{T} of the Z boson. The reweighting is derived separately for the different generator and shower model samples.

5. The reweighting derived in the Z-peak region can then be applied to other simulated samples, such as the $q\bar{q} \to W^+W^-$ sample.

[2]Other backgrounds to the Z production process here are negligible.

8.1 Data and Simulation Samples

Similar to the pileup correction discussed in Chapter 7, data events used to derive the UE modelling correction must pass all of the preselections described in Section 6.2.3. The dataset corresponding to the Z-peak has the additional requirement that both leptons have the same flavor (e^+e^- or $\mu^+\mu^-$) and that 70 GeV $< m_{\ell\ell} <$ 105 GeV. The vast majority of these events will have resulted from the DY di-lepton process with an on-shell intermediate Z boson.

All of the simulated samples discussed in Tables 5.1 and 5.2 were considered, though events were required to pass the same pre-selections, flavor requirements, and mass cuts applied to data. The DY $\ell\ell$ samples comprise the majority of events that pass the selections. The next most important contribution comes from single top and $t\bar{t}$ production, though this is less than 1% of the total yield in the Z-peak region and less than 0.1% of events with $n_{\text{trk}} < 5$. The only other major contributors are the $\gamma\gamma \to \ell^+\ell^-$ processes in the $n_{\text{trk}} = 0$ bin. Because of this, the top and other backgrounds are neglected, and the correction is derived using the POWHEG-BOX v1+PYTHIA 8, POWHEG-BOX v1+HERWIG 7, and SHERPA 2 DY $\ell\ell$ samples.

8.2 $\gamma\gamma \to \ell^+\ell^-$ Subtraction

Rather than taking the expected yields of the elastic, single-dissociative, and double-dissociative $\gamma\gamma \to \ell^+\ell^-$ processes in the $n_{\text{trk}} = 0$ bin directly from the simulation in Table 5.1, a control region is used to check whether these predictions are sensible. The control region should be one where the photon-induced di-lepton process is more prominent than in the Z-peak.

A natural control region for this involves events with $m_{\ell\ell} > 105$ GeV and $n_{\text{trk}} = 0$. This region is away from the Z peak, and the $n_{\text{trk}} = 0$ cut greatly reduces any other backgrounds. Data which passes the preselections described in Section 6.2.3, the same-flavor lepton requirement, and the new $m_{\ell\ell}$ and n_{trk} requirements are compared to the simulation yields that pass the same selections. The simulation samples considered here are the POWHEG+PYTHIA 8 DY $\ell\ell$ and the elastic, single-, and double-dissociative $\gamma\gamma \to \ell^+\ell^-$ processes. The $\gamma\gamma \to W^+W^-$ process contributes less than 1% of the events in this region and is therefore neglected. A template fit is performed based on the $p_{\text{T},\ell\ell}$ spectrum of events in this control region. That is, the shapes of the DY $\ell\ell$ and elastic, single-, and double-dissociative $\gamma\gamma \to \ell^+\ell^-$ processes are set by the simulation, but their overall normalization is allowed to float to best fit the data. The results of this fit are shown in Figure 8.2. Four scale factors result from this fit, and those corresponding to the photon-induced processes are then applied to their appropriate samples in the Z-peak region. Of course, this assumes that the modelling of the $m_{\ell\ell}$ shape is accurate for photon-induced events, but that comes from relatively well understood QED.

For context, after correction, the photon-induced di-lepton processes correspond to 5.5% of the $n_{\text{trk}} = 0$ events in the Z-peak region, 0.5% of the $n_{\text{trk}} = 1$ events, and 0.2% of the $n_{\text{trk}} = 3$ events. However, while those values are averaged over all $p_{\text{T},\ell\ell}$ values, there is

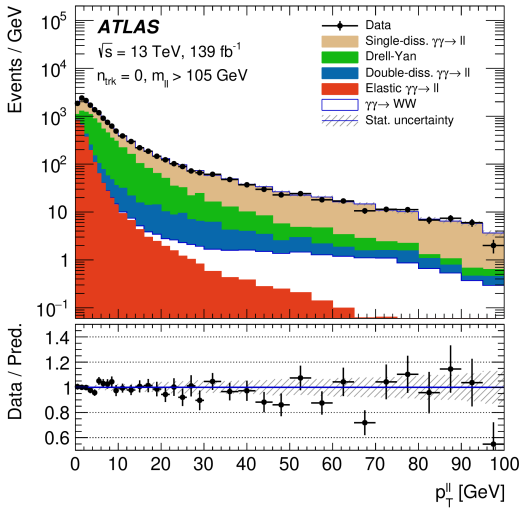


Figure 8.2: Post-fit data and simulation yields in the $\gamma\gamma \to \ell^+\ell^-$ control region [187]. These same flavor lepton events must have $m_{\ell\ell} > 105$ GeV and $n_{\text{trk}} = 0$. The Drell-Yan sample is Powheg+Pythia 8. The $\gamma\gamma \to W^+W^-$ process is plotted here, though it was not included in the fit.

$p_{\text{T},\ell\ell}$-dependence to the photon-induced contribution. In the higher $p_{\text{T},\ell\ell}$ bins, the photon-induced contribution is greater as the leptons resulting from Z production are relatively back-to-back, but leptons produced in dissociative photon-induced processes often have a greater transverse boost.

8.3 Pileup Track Subtraction

For a given track multiplicity counted in the ±1 mm window around the lepton vertex, we would like to determine what fraction of the tracks came from an UE and what fraction came from PU. To do this, we need to determine the distribution of $n_{\text{trk}}^{\text{PU}}$ in the window, which is found using the same process described in Section 7.2.

The relation between PU and UE tracks is constrained by the relation:

$$n_{\text{trk}} = n_{\text{trk}}^{\text{UE}} + n_{\text{trk}}^{\text{PU}}.$$

For a given n_{trk}, the probability of each valid $n_{\text{trk}}^{\text{PU}}$ is given by the distribution shown in Figure 7.2 (when averaging over all of Run 2). For example, if $n_{\text{trk}} = 1$, then there is about a 10% chance that the track is PU, giving a 90% chance that the track is from the UE. If $n_{\text{trk}} = 2$, there is about a 10% chance that 1 track is PU and a 5% chance that both tracks are PU, meaning an 85% chance that both tracks are from the UE, a 10% chance that 1

is, and 5% chance that neither is. We can populate a matrix giving the likelihood of each $n_{\text{trk}}^{\text{UE}}$ value for each value of n_{trk}: $P(n_{\text{trk}}^{\text{UE}} \mid n_{\text{trk}})$.

Our ultimate goal is to determine the probability distribution function (PDF) for $n_{\text{trk}}^{\text{UE}}$. We have the PDF of n_{trk} and $n_{\text{trk}}^{\text{PU}}$ in data, which means that we can get to the PDF for $n_{\text{trk}}^{\text{UE}}$. This is done using the matrix described in the previous paragraph. The probability of a single $n_{\text{trk}}^{\text{UE}}$ value, say k, will be the sum:

$$P(n_{\text{trk}}^{\text{UE}} = k) = \sum_n P(n_{\text{trk}}^{\text{UE}} = k \mid n_{\text{trk}} = n) P(n_{\text{trk}} = n),$$

where the sum runs over all values contained in the PDF for n_{trk}. Because $\text{P}(n_{\text{trk}})$ and $P(n_{\text{trk}}^{\text{UE}} \mid n_{\text{trk}})$ are known, this means that each $P(n_{\text{trk}}^{\text{UE}} = k)$ value can be determined, giving the complete pdf for $n_{\text{trk}}^{\text{UE}}$, which is exactly what we wanted.

In truth, this section described a method that was more of a PDF determination than a subtraction scheme, though the PDF we derived now has PU subtracted. This PDF will be used in the truth-level charged particle unfolding described in the next section.

8.4 Track to Charged Particle Unfolding

Subtracting the presence of the $\gamma\gamma \to \ell^+\ell^-$ process in the Z-peak left us with a data sample that is almost entirely the $q\bar{q} \to Z$ process, and using the method described in Section 8.3, we have found the PDF for the number of UE *tracks* alone in the ± 1 mm window around the lepton vertex in data. The "pileup subtraction" was entirely data-driven. However, in the simulation, we have information about the number of truth-level *charged particles* created in the UE, which is an even more fundamental quantity that the number of UE tracks. While we don't have the truth-level charged particle information in data, we can use a process similar to that in Section 8.3 to derive the PDF for the number of UE charged-particles found in data. This process is known as "unfolding".

The unfolding here will rely on our simulation of the tracking efficiency, which is to say that we use Monte Carlo to derive the conditional probability $P(n_{\text{trk}}^{\text{UE}} \mid n_{\text{ch}})$. This is the probability of observing $n_{\text{trk}}^{\text{UE}}$ UE tracks, given that there were n_{ch} truth-level charged particles. Due to tracking inefficiencies, $n_{\text{trk}}^{\text{UE}}$ should be less than or equal to n_{ch}.

The relationship between the PDFs for $n_{\text{trk}}^{\text{UE}}$ and n_{ch} is:

$$P(n_{\text{trk}}^{\text{UE}} = k) = \sum_n P(n_{\text{trk}}^{\text{UE}} = k \mid n_{\text{ch}} = n) P(n_{\text{ch}} = n). \tag{8.1}$$

To be precise, what we know from data is $P(n_{\text{trk}}^{\text{UE}} = k)$, and what we have from simulation is $P(n_{\text{trk}}^{\text{UE}} = k \mid n_{\text{ch}} = n)$. What we *want* to know is $P(n_{\text{ch}} = n)$. If we express Eq. 8.1 as a matrix multiplication, we can see that to solve for $P(n_{\text{ch}} = n)$ algebraically, we would have to invert the matrix representing $P(n_{\text{trk}}^{\text{UE}} = k \mid n_{\text{ch}} = n)$. Such an inversion is not always possible or well-defined.

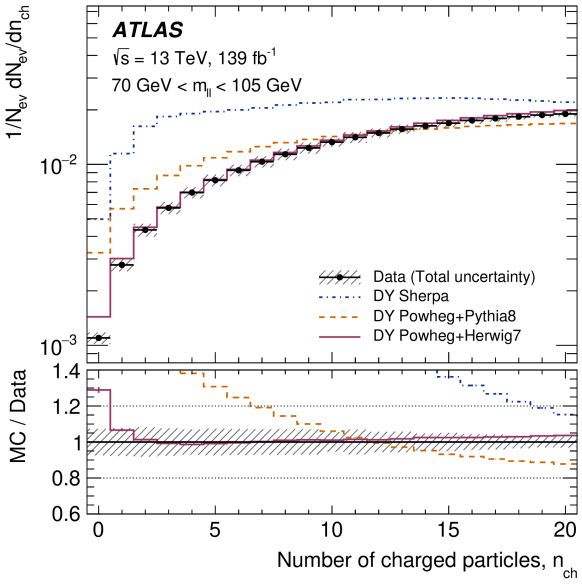


Figure 8.3: The probability distribution functions of the number of underlying event charged particles, $P(n_{\text{ch}})$, for data (shown in black dots) and for simulation (shown in dashed lines) in the Z-peak region [187]. The simulated PDFs are shown for the POWHEG-BOX v1+PYTHIA 8, POWHEG-BOX v1+HERWIG 7, and SHERPA 2 DY $\ell\ell$ samples separately.

To find $P(n_{\text{ch}} = n)$, we use the method known as "Iterative Bayesian Unfolding" [191, 192] with four iterations[3].

8.5 Reweighting the Number of Charged Particles

Having followed the steps outlined in Sections 8.2, 8.3, and 8.4 using the Z-peak region data, we now have the PDF: $P(n_{\text{ch}})_{\text{data}}$. From simulation we also have $P(n_{\text{ch}})_{\text{MC}}$. These PDFs are shown in Figure 8.3.

The PDFs are binned in integer values of the number of charged particles, so the weight that should be applied to each event falling in the k-th bin in the simulation is

$$\frac{P(n_{\text{ch}} = k)_{\text{data}}}{P(n_{\text{ch}} = k)_{\text{MC}}}.$$

The application of this reweighting is illustrated in Figure 8.4. In truth, PDFs and weights are derived for intervals of $p_{\text{T},\ell\ell}$ (which corresponds to $p_{\text{T},Z}$) separately, though the figure

[3]There is no change in the result after the fourth iteration.

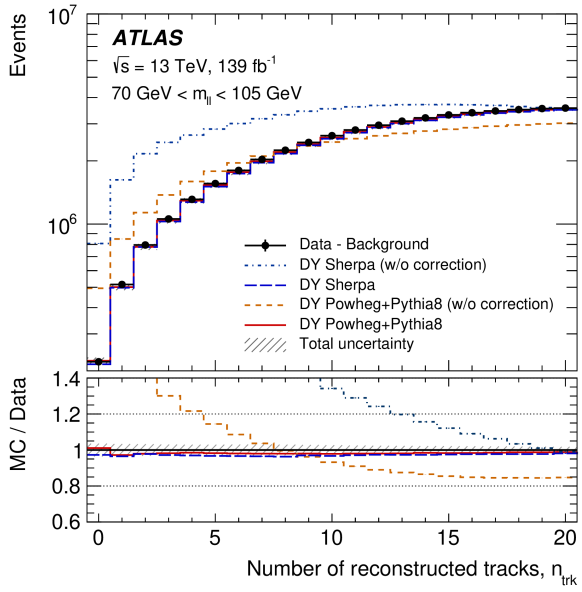


Figure 8.4: The probability distribution functions of the number of underlying event charged particles, $P(n_{\text{ch}})$, for data (shown in black dots) and for simulation (shown in dashed lines) in the Z-peak region [187]. The PDFs are shown both before and after reweighting for illustrative purposes. The post-reweighting PDFs in simulation agree with the data within uncertainty.

has been integrated over these bins. The $p_{\text{T},\ell\ell}$ bins have a width of 5 GeV. The weights are also derived for the three types of generators separately. Systematic uncertainties associated with this reweighting are discussed in Chapter 12.

After applying the correction for the number of UE particles described in this chapter and the correction for the number of PU tracks described in Chapter 7, the simulation should be able to predict the fraction of background events that will pass the window-based exclusivity selection relatively accurately. It should also be able to predict the fraction of photon-induced events that will fail the selection due to PU tracks.

While the PU track correction is applied to all samples, the UE particle correction is only applied to the samples in Table 5.2 with $q\bar{q}$ initial states. The particular set of weights used relies on the generator. For the WW, WZ, and ZZ samples, the weights are applied as a function of the truth-level p_{T} of the di-boson system.

8.6 Reweighting of Z-boson p_{T} in Drell-Yan Simulation

Drell-Yan $\tau\bar{\tau}$ production can contaminate the signal region and major control regions, as it can give an observed $e^{\pm}\mu^{\mp}$ final state. Similar to the $n_{\mathrm{trk}}^{\mathrm{UE}}$ spectrum, the p_{T} spectrum of the Z-boson produced in DY samples is known to be mismodelled in Powheg.

A correction for this mismodelling is derived by comparing the $p_{\mathrm{T},\ell\ell}$ spectra of ee and $\mu\mu$ events in data and simulation. Here, no cut on the n_{trk} in the event needs to be applied. The region used to define this correction is ee and $\mu\mu$ events that pass our lepton selections and have 75 GeV $< m_{\ell\ell} <$ 105 GeV. The background is subtracted from the data in this region using default simulation, and a weight is derived as the ratio of data events to simulated DY events in 5 GeV bins of $p_{\mathrm{T},\ell\ell}$.

This correction is applied *before* the derivation of the UE correction described in the previous sections of this chapter.

8.7 Reweighting of Simulated WW p_{T} in $q\bar{q} \to W^+W^-$ Events

While the Powheg and Sherpa $q\bar{q} \to W^+W^-$ simulations are performed with next-to-leading order (NLO) precision, calculations have recently been performed at next-to-next-to-leading order (NNLO), using resummation of soft gluon emissions up to next-to-next-to-next-to-leading-logarithm ($\mathrm{N^3LL}$) accuracy. This calculation was performed with MATRIX+RadISH [193, 194, 195, 196, 197, 198, 199, 200, 201].

Weights are derived by taking the ratio of the MATRIX+RadISH yield to that of either Powheg or Sherpa, in bins of $p_{\mathrm{T},WW}$. Two sets of weights are found; one for each generator separately. This is not a data-driven correction, so the weights are derived at truth-level.

This correction is also applied *before* the derivation of the UE correction described in the previous sections of this chapter.

Chapter 9

Estimation of Events with Fake Leptons and Top Backgrounds

After applying the pileup (PU) correction of Chapter 7 and the underlying event (UE) correction from Chapter 8, the dominating backgrounds for the $\gamma\gamma \to W^+W^-$ analysis should be modelled well by the samples listed in Table 5.2, at least in terms of the number of tracks reconstructed within the ± 1 mm exclusivity window. We should note that the samples in that table will all give final state leptons at truth level. However, with the end goal of estimating $N_{\text{bkgd,SR}}$ correctly, we must make an estimate of all backgrounds, even sub-dominant ones, which may comprise 5% or less of the final yield in the signal region for example.

The estimation of two such sub-dominant backgrounds is described in this chapter:

- An estimation of the yield of events where one of the leptons is "fake"[1]. Fake leptons are due primarily to two phenomena. The first is the case when the decay of a hadron within a jet produces a real lepton. Occasionally these leptons will be incorrectly considered well-isolated at reconstruction-level, giving the false impression that the lepton comes from the underlying hard-scatter of the event, rather than a secondary decay. The second phenomenon is when a non-lepton object is incorrectly identified as a lepton at reconstruction level. For example, a jet or photon can be reconstructed as an electron. The primary processes that would lead to this type of background are W+jets and $W + \gamma$. Unfortunately, the modelling of the fake-lepton phenomena in simulation is not perfect, leading to large uncertainties if attempting to make an estimate from simulation alone. Futhermore, performing such an analysis in simulation would require a dedicated sample with large statistics for events with 0-4 tracks.

- An estimation of the yield due to backgrounds that involve top quarks. While these processes have a large cross-section at the LHC, only a very tiny fraction of these events should pass the exclusivity selection. The UE correction is not directly applicable to

[1] Also referred to as "non-prompt"

these samples, as the primary production mechanism involves initial-state gluons [202], making it hard to estimate the expected yield with few tracks in the ±1 mm window.

9.1 Data-Driven Method for Fake Lepton Estimation

To make an estimation of the yield of events due to fake leptons, we employ the "fake factor" method, which proceeds as follows:

1. First, we note that our identification and isolation cuts are designed precisely to eliminate these backgrounds. It is much more common for events with fake leptons to fail those cuts, so a region defined with the cuts reversed would be enriched in the backgrounds we are trying to study.

2. We also note that the ratio of fake leptons failing the cuts to fake leptons passing the cuts should be roughly process-independent for events will similar values of n_{trk}. To find this ratio, we can define a region that is orthogonal to the SR and main CRs used for the $\gamma\gamma \to W^+W^-$ analysis, and further subdivide that region into one with leptons that pass our original identification and isolation cuts and one where one of the leptons fails.

3. Using the subdivided region, we find the ratio of fake leptons that pass the cuts to those that fail: R_{ff} ("ff" here is for "fake-factor").

4. We can then define regions that are the same as our SR and CRs, but with one lepton that fails the identification or isolation cuts. If the yield of fake leptons in this region is multiplied by R_{ff}, what results is an estimate for the yield of events due to fake leptons in the main SR and CRs.

This process is visualized in Figure 9.1.

9.1.1 Regions for Fake Factor Derivation and Application

As illustrated in Figure 9.1, two new regions are required for the derivation of the fake factor. Furthermore, for each of the SR and CRs used in the analysis[2], a new region is required, which is where the fake factor will be applied.

Throughout this section, leptons will be described as either "passing" or "failing" their respective selections. For both electrons and muons, if the lepton is said to pass the selection, that means that it satisfies all of the criteria laid out in Section 6.2.2. If a muon fails the selection, that means that it satisfies all of the criteria of Section 6.2.2, except that it must fail the FixCutLoose isolation requirement. Similarly, a failing electron is one that satisfies all of the criteria of Section 6.2.2, except that it fails *either* the medium identification criteria

[2]see Chapter 11

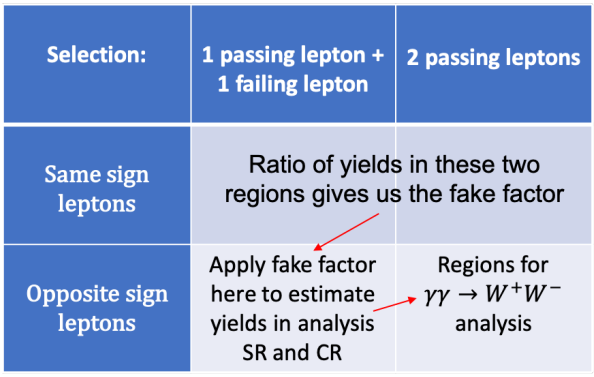


Figure 9.1: Illustration of the logic behind the fake-factor method. The fake factor ratio is derived using two regions that are orthogonal to the SR and CRs used in the $\gamma\gamma \to W^+W^-$ analysis. The orthogonality is enforced by the presence of same-*sign* leptons. The fake factor can then multiplied by the yield in regions that are the same as the SR or CRs in all respects except for the presence of a lepton that fails identification and isolation requirements.

or the loose version of the "Fix" isolation criteria. In either failure mode, electrons must pass the loose identification criteria.

To measure the fake factor, we want to define regions with no overlap with events used elsewhere in the $\gamma\gamma \to W^+W^-$ analysis. Because the fake factor should not depend on the charge of the lepton, we can therefore define regions that have two leptons of the same charge but opposite flavor. All other events used in this analysis have *opposite* sign leptons. The region enriched in fake leptons (the top left bin of Figure 9.1) has one lepton that passes the nominal selections described in Section 6.2.2 and one that fails. If the electron is the failing lepton, then the object that triggered event readout is required to be the passing muon in the event. If the muon is the failing lepton, then the event must have been triggered by the passing electron. A requirement of $m_{\ell\ell} > 20$ GeV is placed on the di-lepton pair. The events are also required to have $1 \leq n_{\text{trk}} \leq 4$ in the ± 1 mm window around the lepton vertex.

The top right region in Figure 9.1 should have fewer fake leptons. This region is defined in exactly the same way as the region defined in the preceding paragraph, except both leptons must pass the isolation and identification criteria.

In both of these regions, there should be events where both leptons are not fake. The expected contribution from these events is subtracted using simulation. What is left is four numbers: $N_{e,\text{fake,fail}}$ (the number of events where the electron was fake and fails the cuts - from top left region of Figure 9.1), $N_{e,\text{fake,pass}}$ (the number of events where the electron was fake and passed the cuts - from top right region of Figure 9.1), $N_{\mu,\text{fake,fail}}$ (the number of events where the muon was fake and fails the cuts - from top left region of Figure 9.1), $N_{\mu,\text{fake,pass}}$ (the number of events where the muon was fake and passed the cuts - from top

right region of Figure 9.1).

The fake factor is taken separately for electrons and muons. The electron fake factor is

$$R_{\mathrm{FF,e}} = \frac{N_{e,\mathrm{fake,pass}}}{N_{e,\mathrm{fake,fail}}}, \tag{9.1}$$

and the muon fake factor is

$$R_{\mathrm{FF},\mu} = \frac{N_{\mu,\mathrm{fake,pass}}}{N_{\mu,\mathrm{fake,fail}}}, \tag{9.2}$$

To produce an estimate for the number of data events with fake leptons, new regions are defined for the application of the fake factors. Duplicate regions for each of the SR and CRs are created, with the important difference that one of the leptons in the event must fail the identification/isolation criteria. For reference, the SR and CRs are laid out in detail in Chapter 11. In these failing-lepton versions of the SR and CRs, the expected number of events with two genuine leptons is estimated using simulation. After this expected yield is subtracted from the data yield, what remains in data is the contribution from events with fake leptons that fail the lepton selection. The yield of events with a failing electron are multiplied by $R_{\mathrm{FF,e}}$, giving an estimate for the number of events with fake electrons that pass the lepton selection in the appropriate region. Similarly, multiplying the yield with a failing muon by $R_{\mathrm{FF},\mu}$ gives an estimate for the number of events with fake muons.

The actual yields of events with fake leptons will be discussed in Chapter 11.

9.2 Studies of Top Backgrounds

A small contribution in both the SR and CRs comes from $t\bar{t}$ events, where the W-bosons from the top quarks decay into an opposite-sign, opposite-flavor pair of leptons. A smaller contribution is expected from $W + t$ events, due to its significantly lower cross-section at the LHC. In either case, for such events to enter the SR or a CR, there must be few extra reconstructed tracks in the event, which will be rare given that high-energy QCD objects are involved in the final state.

The estimate of the yield of events in the SR and CRs from the $t\bar{t}$ and $W + t$ processes will be taken from the simulations in Table 5.2. However, because these are sub-dominant backgrounds, no attempt is made to reweight such events using a process similar to that described in Chapter 8. Because of this, the estimate from simulation is not expected to be extremely accurate. This section describes a determination of the modelling uncertainty on top events with 0-4 tracks.

Two control regions are considered, with definitions summarized in in Table 9.1. The two control regions are similar, differing only in the jet selection criteria. The first control region, called "2 general jets CR", requires at least two jets with no limit on the jets' $|\eta|$s, while the second control region, called "2 forward jets CR", requires at least two jets with $|\eta| > 3$. Charged particles in these high-$|\eta|$ jets will not cause reconstructed tracks, so the presence of those jets should have little impact on the number of tracks observed in the event.

Cut	**2 general jets**	**2 forward jets**
$m_{\ell\ell}$	> 20 GeV	
Leptons flavor	$e^{\pm}\mu^{\mp} + \mu^{\pm}e^{\mp}$	
Number of jets	≥ 2	
Jet η	$\lvert\eta\rvert < 4.9$	$3 < \lvert\eta\rvert < 4.9$
Jet p_{T}	> 30 GeV	
$\lvert z_{\mathrm{PV}} - z^{ll}_{\mathrm{vtx}}\rvert$	< 2.0 mm	

Table 9.1: Definitions of the $t\bar{t}$ control regions. The main difference between the two is that the forward jets control region requires at least two jets with $|\eta| > 3$, such that charged particles within the jet would not create reconstructed tracks, whereas the general jet region has no cut on angularity. The lepton selections are the same as those described in Section 6.2.2.

A brief overview of jet reconstruction with ATLAS was presented in Section 4.4. For this top yield estimation, the jets are reconstructed using the anti-k_t jet clustering algorithm [134] with a radius variable of $R = 0.4$ and the particle flow algorithm [137]. To be considered, jets must have $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 4.9$. Additionally, if the jet has $|\eta| < 2.4$ and $p_{\mathrm{T}} < 60$ GeV, the jet must have a jet-vertex-tagger (JVT) [203] fraction greater than 0.5. This selection increases confidence that the jet comes from the event's primary vertex [204], reducing contamination from pile-up interactions. The jets are corrected for contributions arising from pile-up [205], and are calibrated to the particle energy scale prior to interaction with the detector [136]. Overlap removal is applied so that muons, electron, and jets will always be distinct objects. The leptons are selected as described in Sec. 6.2.2.

The cut on the difference in longitudinal beampipe position between the primary vertex and lepton vertex ($|z_{\mathrm{PV}} - z^{ll}_{\mathrm{vtx}}|$) is designed to avoid considering events where the jets in the the event come from a pile-up interaction rather than the interaction that produces the leptons. This is possible when the sum of the p^2_{T} of the tracks from the pileup interaction exceeds that of the lepton-producing interaction. In particular, in events with fewer tracks associated with the lepton interaction, this is more likely to happen. This cut is not necessary in the SR or CRs described in Chapter 11, as jets are not used for event selection there.

Figure 9.2 shows the distributions of the number of tracks within the ± 1 mm window around the lepton vertex for the two control regions. While simulation tends to overpredict the number of tracks near the leptons, the per-bin data/simulation disagreement tends to not exceed 20% for the 2 general jet region and 30% for the forward jet region. It is also clear from this figure that the fraction of $t\bar{t}$ and $W + t$ events with fewer than 5 tracks is very small. While this region suffers from high statistical uncertainties, if we average over the 0-4 track bins, the data vs. simulation discrepancy is less than 30%. Therefore, a conservative uncertainty on the $t\bar{t}$ and $W + t$ yield at low-n_{trk} is placed at 30%.

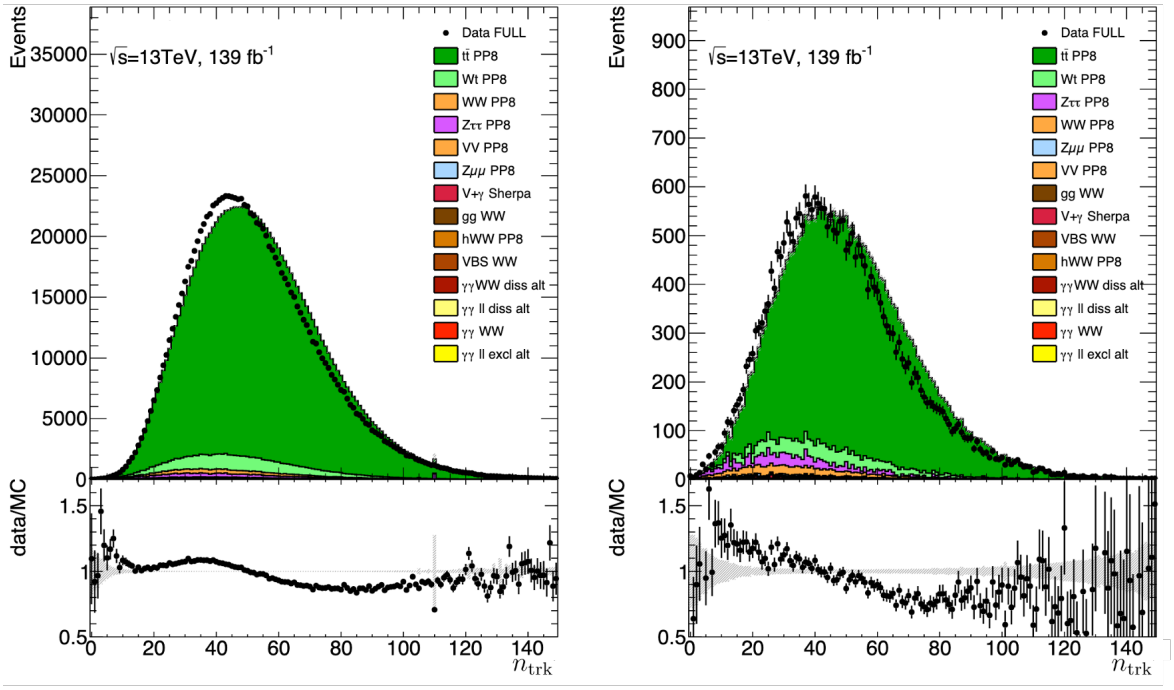


Figure 9.2: Distributions of the number of tracks, n_{trk}, within the ± 1 mm window around the lepton vertex for events in the 2 general jets (left) and 2 forward jets (right) $t\bar{t}$ control regions. There is some mismodelling here, with simulation tending to overpredict the number of tracks, though the overall normalization of the samples is accurate.

Chapter 10

Estimation of Dissociative Production

With corrections for pileup (PU) tracks and QCD-induced processes' underlying event (UE) tracks in hand, and having studied the yields of minor backgrounds, it is now possible to make a prediction for $N_{\text{bkgd,SR}}$. Therefore, at this point it is possible to measure the fiducial cross-section of the $\gamma\gamma \to W^+W^-$ process using Equations 6.1 and 6.2.

However, in Equation 6.2, there are unknown factors of ε_S and A. To estimate these values, we would like to use simulation of $\gamma\gamma \to W^+W^-$ events, both elastic and dissociative. The primary loss of signal efficiency will come from PU tracks entering the ± 1 mm exclusivity window, but additional losses will come from reconstruction inefficiency for the leptons and the occasional tracks resulting from a dissociating proton in single- or double-dissociative $\gamma\gamma \to W^+W^-$ events. Beyond finding ε_S, we want to use signal simulation to predict the signal yield based on SM expectations. This prediction will be compared to the measured value.

Unfortunately, just as it was not possible to take the background simulation at face value when estimating $N_{\text{bkgd,SR}}$, it is not possible to take $N_{\text{sig,SR}}$ directly from simulation. In Section 5.1, problems with the simulation of the $\gamma\gamma \to W^+W^-$ process were explained in some detail. While generators do exist for the estimation of elastic $\gamma\gamma \to W^+W^-$ production, there are no dedicated generators for the single-dissociative (SD) and double-dissociative (DD) versions of the process. Even the dedicated elastic generator used in this analysis does not include soft rescattering effects (see Section 2.2.4), which should suppress the overall cross section. There *are* dedicated generators for the dissociative $\gamma\gamma \to \ell^+\ell^-$ process, but even these carry fairly significant uncertainties due to the fact that the dissociation involves non-perturbative QCD physics.

Therefore, a data-driven approach will be used to derive a scale factor, S_{Excl}, to account for mismodelling of the elastic $\gamma\gamma \to W^+W^-$ process and the lack of dedicated generators for its dissociative components[1]. The scale factor is used to estimate $N_{\text{sig,SR}}$ as

$$N_{\text{sig,SR}} = S_{\text{Excl}} * N_{\text{El.MC,SR}}, \tag{10.1}$$

[1] A similar approach was used in [41] and [42]

where $N_{\text{El.MC,SR}}$ is the simulation-based estimate of the elastic $\gamma\gamma \to W^+W^-$ yield in the signal region (SR). This scale factor will also be used in any other control region (CR) that has 0 tracks in the ± 1 mm window around the lepton vertex. Its derivation is detailed in Section 10.1.

The fact that events in the SR have $n_{\text{trk}} = 0$ implies that there are neither pileup tracks nor tracks resulting from proton dissociation. However, it *is* possible (though rare) for a proton's dissociation to produce a track. Because the elastic $\gamma\gamma \to W^+W^-$ process will never create a track from proton dissociation and because the n_{trk} distribution due to dissociation tracks should be different from the $n_{\text{trk}}^{\text{PU}}$ distribution, it is not well-motivated to use a scale factor for signal yield estimation in CRs with $n_{\text{trk}} \geq 1$.. Simulation will be used to make $N_{\text{sig,CR}}$ estimates, but corrections will be applied to the single-dissociative MG5_AMC@NLO+Pythia8 (MG5) simulation. These corrections are described in Section 10.2.

10.1 Dissociative Scale Factor for 0 Track Regions

The critical insight behind a data-driven derivation of S_{Excl} is that the $\gamma\gamma \to \ell^+\ell^-$ process *also* has elastic, SD, and DD components. It will suffer from any rescattering effects not included in the simulation as well. Therefore, to avoid pre-emptively looking at our SR, we will find S_{Excl} using $\gamma\gamma \to \ell^+\ell^-$ events. To do this, we will need to define a region that is enriched in $\gamma\gamma \to \ell^+\ell^-$, with low background contamination. We would also prefer that the elastic:SD:DD yield ratios in this region mimic those in the SR.

Because of these considerations, we will use events with same-flavor, opposite-sign lepton pairs that have $m_{\ell\ell} > 160$ GeV. We will also require that $n_{\text{trk}} = 0$ in the ± 1 mm window around the lepton vertex. Our SR will involve *opposite*-flavor pairs, so there is no overlap between this new region and the SR. Additionally, the $m_{\ell\ell} > 160$ GeV requirement forces the initial state photons to have enough energy to have created an on-shell W^+W^- pair. This selection assumes that the proton's potential dissociation depends on the energy of the emitted photon, and that it is otherwise decoupled from the photon-induced process. While this is a physically well-motivated assumption, an associated systematic is assigned, as discussed in Chapter 12. The lepton selection requirements are the same as discussed in Section 6.2.2, and the other preselections for events are the same as described in Section 6.2.3.

Having defined a region in which to derive S_{Excl}, we will use the formula

$$S_{\text{Excl}} = \frac{N_{\text{data}} - N_{\text{bkgd}}}{N_{\text{El,MC}}} \tag{10.2}$$

to calculate its value. Here, N_{data} is the number of data events in the region, N_{bkgd} is the expected number of background events in the region, and $N_{\text{El.MC}}$ is the number of elastic $\gamma\gamma \to \ell^+\ell^-$ events in the region as predicted by simulation.

As with the $\gamma\gamma \to W^+W^-$ analysis as a whole, much of the difficulty with this derivation involves the prediction of N_{bkgd}. Rather than taking this directly from simulation, which has shape and normalization uncertainties, a data-driven method is used. To estimate N_{bkgd}

Process	$n_{\text{trk}} = 0$	$n_{\text{trk}} = 2$	$n_{\text{trk}} = 5$
$\gamma\gamma \to \ell^+\ell^-$	53.2%	3.3%	0.7%
$\gamma\gamma \to W^+W^-$	1.4%	0.1%	0.0%
DY $\ell^+\ell^-$	43.9%	93.8%	96.0%
Others	1.5%	2.8%	3.4%

Table 10.1: The expected percentage yields of photon-induced processes and QCD-induced processes in the $m_{\ell\ell} > 160$ GeV region for different values of n_{trk} in the ± 1 mm window around the lepton vertex. These yields are based on predictions from the simulated samples in Tables 5.1 and 5.2. "Others" indicates background processes included in Table 5.2 except for the Drell-Yan di-lepton process.

using data, we will use the template method. This means that we will use data to determine the shape of the $m_{\ell\ell}$ spectrum of the background, which is mostly Drell-Yan di-lepton production.

While photon-induced processes dominate when $n_{\text{trk}} = 0$, they are a small component of the yield when $n_{\text{trk}} > 0$. The expected percentage yields of photon-induced processes for $n_{\text{trk}} = 0$, 2, and 5 are shown in Table 10.1. The predictions are based on the simulations discussed in Chapter 5. For the $n_{\text{trk}} = 5$ bin, fewer than 1% of events are expected to be photon-induced. Because of this, the background template will be based on these events. An assumption of the template method is that the background's $m_{\ell\ell}$ spectrum does not depend on n_{trk}. Uncertainty will be assigned to the template choice by comparing the $n_{\text{trk}} = 5$ template to the $n_{\text{trk}} = 2$ template. Further details are given in Chapter 12.

The template gives us the shape of the background's $m_{\ell\ell}$ distribution when $n_{\text{trk}} = 0$, but it does not tell us N_{bkgd} unless it is normalized correctly. To normalize the template, we must find a place where we know the background's yield even for events with $n_{\text{trk}} = 0$. Luckily, background will be dominant in the Z-peak region, giving us a convenient place to perform our normalization. In this context, the Z-peak region is taken to be events with same-flavor, opposite-sign lepton pairs that pass all preselections and have 83.5 GeV $< m_{\ell\ell} <$ 98.5 GeV. Thus, the proper normalization of the template will be the ratio of events in the Z-peak with $n_{\text{trk}} = 0$ to those with $n_{\text{trk}} = 5$, so we have

$$N_{\text{bkgd}} = N_{\text{high-mll,5}} * \frac{N_{Z-\text{peak,0}}}{N_{Z-\text{peak,5}}}. \tag{10.3}$$

Here, $N_{\text{high-mll,5}}$ is the number of events with $n_{\text{trk}} = 5$ and $m_{\ell\ell} > 160$ GeV, and $N_{Z-\text{peak,0}}$ and $N_{Z-\text{peak,5}}$ are the Z-peak yields with 0 and 5 tracks, respectively.

Because the normalization region will have some photon-induced contributions that are not accounted for, an uncertainty is assigned using a larger Z-peak region of 73.5 GeV $< m_{\ell\ell} <$ 108.5 GeV. Further details are given in Chapter 12.

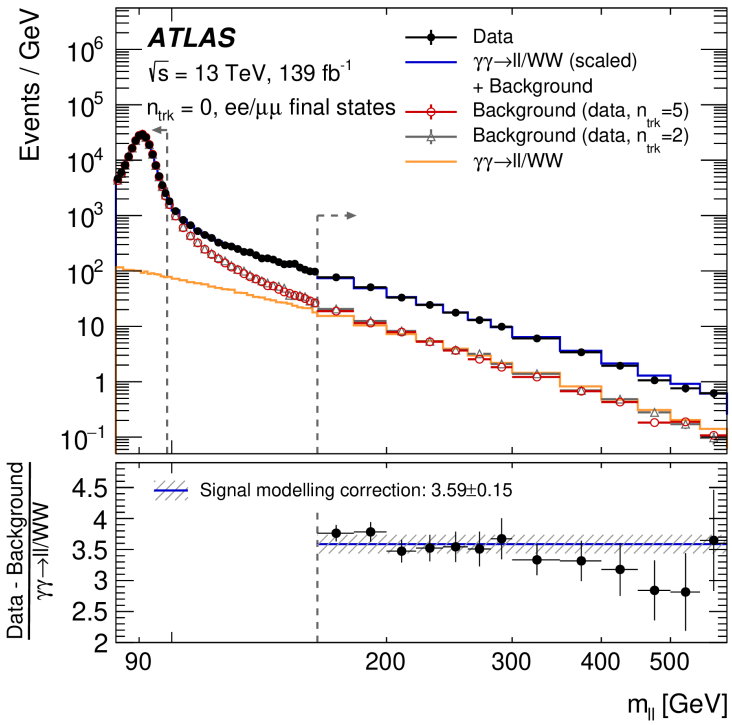


Figure 10.1: The $m_{\ell\ell}$ spectrum of same-flavor, opposite-sign data and simulation events with $n_{\text{trk}} = 0$ [187]. The simulated spectrum of elastic $\gamma\gamma \to \ell^+\ell^-$ and $\gamma\gamma \to W^+W^-$ events is shown before (yellow) and after (blue) application of S_{Excl}. The data templates, which come from events with $n_{\text{trk}} = 5$ (red) and $n_{\text{trk}} = 2$ (grey) are also shown, having been normalized to the $n_{\text{trk}} = 0$ yield in the Z-peak.

With all the needed components in hand, we now apply Equation 10.2, finding[2] that $S_{\text{Excl}} = 3.59 \pm 0.15$. This is the value that will be used to determine the $\gamma\gamma \to W^+W^-$ yield in any region that has $n_{\text{trk}} = 0$ through Equation 10.1. The data/simulation agreement after the application of S_{Excl} is illustrated in Figure 10.1.

10.2 Corrected Simulation for 1-4 Track Regions

While the exclusivity scale factor is applicable for events with $n_{\text{trk}} = 0$, it is not applicable for events with $n_{\text{trk}} > 0$, where the elastic:SD:DD ratio should be drastically different. It is also not possible to derive scale factors for each value of n_{trk} separately, as only $n_{\text{trk}} = 0$

[2]In truth, the high-$m_{\ell\ell}$ region described in this chapter is used as a CR for the fit described in Chapter 13. Through this fit, the yield of $\gamma\gamma \to \ell^+\ell^-$ and $\gamma\gamma \to W^+W^-$ events in the SR and CRs described in Chapter 11 are able to influence the prediction of $N_{\text{El,MC}}$.

events have a sufficiently high photon-induced component relative to backgrounds. Because of these difficulties, simulation will be used to predict the signal yields in CRs with $n_{\text{trk}} > 0$, most of which have a requirement of $1 \leq n_{\text{trk}} \leq 4$.

As discussed in Section 5.1, the track multiplicity of the MG5 simulation is not expected to be correct. When the LPAIR and MG5 samples of the $\gamma\gamma \to \ell^+\ell^-$ process are compared, the MG5 tends to predict higher numbers of tracks from the dissociative protons, $n_{\text{trk}}^{\text{diss}}$. While LPAIR is not expected to be perfect, it is still more physically motivated than the MG5 setup. Because the disassociation is simulated the same way for both the MG5 $\gamma\gamma \to \ell^+\ell^-$ samples and for the MG5 $\gamma\gamma \to W^+W^-$ samples, the $\gamma\gamma \to W^+W^-$ samples are expected to have similarly non-physical track multiplicity spectra. Thus, in order to extract physically reasonable CR yield predictions from the MG5 samples, we must perform a reweighting based on $n_{\text{trk}}^{\text{diss}}$. The goal of the reweighting is to match the MG5 track multiplicity spectrum to that predicted by LPAIR. MG5 and LPAIR samples with complementary di-lepton phase space exist for SD $\gamma\gamma \to \mu^+\mu^-$ production with $m_{\mu\mu} > 18$ GeV.

In order to perform the reweighting in a more event-specific manner, we would like to perform the reweighting as a function of key variables that are most correlated with the track multiplicity. It is physically reasonable to expect that the track multiplicity resulting from the proton disassociation will be related to the kinematics of the photons that the protons radiate, in particular with the photon energy and p_{T}. At the event level, these manifest as the $m_{\gamma\gamma}$ and $p_{\text{T},\gamma\gamma}$ of the di-photon system, which will also correspond to the mass and p_{T} of the resulting di-lepton or WW system. Because the p_{T} of the WW system in the MG5 $\gamma\gamma \to W^+W^-$ sample is generally $\ll 1$ GeV, performing the reweighting as a function of $p_{\text{T},\gamma\gamma}$ will not be useful. Therefore, the reweighting is performed as a function of $m_{\gamma\gamma}$.

In order to extract the weights, histograms of track multiplicity are constructed for bins of $m_{\gamma\gamma}$ for both MG5 and LPAIR SD $\gamma\gamma \to \ell^+\ell^-$. The LPAIR bin acts as the numerator and the MG5 bin as the denominator. The histograms are normalized to the respective cross-sections.

Example histograms of the $n_{\text{trk}}^{\text{diss}}$ yields predicted by MG5 and LPAIR are shown in Fig. 10.2 for events with $m_{\gamma\gamma}$ between 140 and 180 GeV. The per-bin weight would be the ratio of the LPAIR yield to the MG5 yield. These weights are applied to the MG5 SD $\gamma\gamma \to W^+W^-$ samples depending on the event's $m_{\gamma\gamma}$ and the number of reconstructed tracks resulting hard scatter.

Because the $n_{\text{trk}}^{\text{diss}}$ spectrum is modelled slightly better for the MG5 DD $\gamma\gamma \to \ell^+\ell^-$ samples when compared to the Pythia8 default samples, and because DD production is subdominant compared to SD production, no similar weighting is derived for DD simulation. The yield of DD $\gamma\gamma \to W^+W^-$ events in the the 1-4 track CRs is taken directly from simulation.

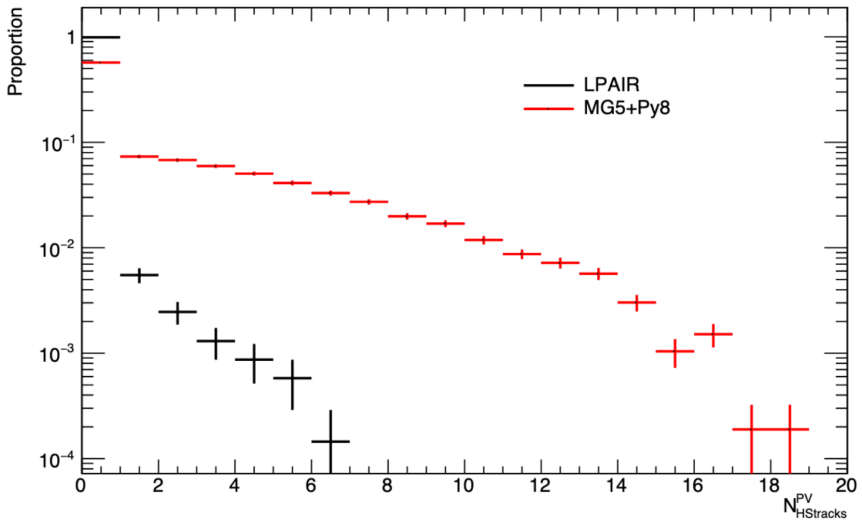


Figure 10.2: Track multiplicity ($N_{\mathrm{HStracks}^{\mathrm{PV}}} = n^{\mathrm{diss}}_{\mathrm{trk}}$) spectra for di-muon events with truth-level $m_{\ell\ell}$ (or equivalently $m_{\gamma\gamma}$) between 140 and 180 GeV for both MG5_AMC@NLO+Pythia8 and LPAIR. Spectra are normalized to have an integral of 1 for visualization purposes.

Chapter 11

Signal and Control Regions

The idea of using a Signal Region (SR) and Control Regions (CRs) for the $\gamma\gamma \to W^+W^-$ analysis was introduced in Chapter 6 without formally defining the regions. This was to avoid making region definitions before fully understanding and correcting our simulation. In this chapter, we will finally define the SR and main CRs and give physical motivations for these definitions. Miscellaneous CRs have been discussed so far in Chapters 8 and 9:

- A CR was used to improve the estimation of the $\gamma\gamma \to \ell^+\ell^-$ yield in the Z-peak region for the derivation of the underlying event (UE) correction[1].
- CRs were introduced to study the uncertainty on the yield of processes involving top quarks with few tracks in the ± 1 mm exclusivity window[2].

In these cases, the use of a CR allows us to gauge the accuracy of simulation compared to data in a well-defined region. In the $\gamma\gamma \to \ell^+\ell^-$ case, we additionally found a scale factor to account for overall normalization discrepancies. Typically a CR is defined to emphasize a particular process relative to others by cutting on variables or objects such as lepton flavors and kinematics or the presence of tracks and jets. Similarly, both the Z-peak region used in Chapter 8 and the high-mass $\ell^+\ell^-$ region used in Chapter 10 are CRs to the analysis as a whole, allowing us to gauge the accuracy of UE modelling and proton-dissociation modelling.

Now, with a complete set of corrections for the simulation, including pileup (PU) corrections, UE correction, and dissociation corrections, we can now define our SR and CRs. The SR should of course be enriched in the $\gamma\gamma \to W^+W^-$ process. However, background events *will* contaminate the SR; after all, one of the main tasks of the analysis is to estimate $N_{\text{bkgd,SR}}$. Therefore, this last and central set of CRs will be defined with the purpose of checking the normalization of the main backgrounds to the $\gamma\gamma \to W^+W^-$ analysis: inclusive WW, DY $\tau\bar{\tau}$, and $\gamma\gamma \to \tau^+\tau^-$.

It is important to note that we will not be able to define the CRs to perfectly isolate only one process. Just as a variety of background processes will contaminate the SR, the signal

[1] Recall that the $\gamma\gamma \to \ell^+\ell^-$ yield was subtracted from the Z-peak when deriving the UE correction

[2] See section 9.2

itself and Drell-Yan (DY) $\tau\bar{\tau}$ process will "contaminate" the CR designed to emphasize inclusive WW, for example. Rather than using the CRs independently to extract normalization corrections for individual processes, a fit will be performed that combines yield information from the SR, the three CRs described in Section 11.2, and the high-$m_{\ell\ell}$ region described in Section 10.1. In this fit, the normalizations of $\gamma\gamma \to W^+W^-$, $\gamma\gamma \to \ell^+\ell^-$, DY $\ell^+\ell^-$, and inclusive WW will be allowed to float in order to maximize the Poisson probability of the yields across the five regions. This fit is described in detail in Chapter 13.

11.1 Signal Region Definition

All events in the SR must pass the triggers, lepton selections, and preselections defined in Sections 6.2.1, 6.2.2, and 6.2.3. Additionally, the leptons must have opposite charge, with one being an electron and the other a muon. The opposite charge requirement stems from the fact that the initial state di-photon system is electrically neutral; the opposite flavor requirement is to reduce contamination from DY events. There is an additional cut requiring that the p_{T} of the $e+\mu$ final state[3], $p_{\mathrm{T},\ell\ell}$, is greater than 30 GeV. This means that the two leptons are not very back-to-back in physical space, accounting for the neutrinos that will also carry away momentum from the W decays. The 30 GeV value is set to reduce DY $\tau\bar{\tau}$ and $\gamma\gamma \to \tau^+\tau^-$ contamination. While there *are* neutrinos present in τ-decays, due to their lighter mass, the τ leptons will generally be more boosted than W bosons, making the charged leptons resulting from their decays more back-to-back. Events in the SR must have 0 tracks in the ± 1 mm window around the lepton vertex.

11.2 Main Control Regions

As in the SR case, all events in the CRs must pass the triggers, lepton selections, and preselections defined in Sections 6.2.1, 6.2.2, and 6.2.3. Three control regions will be used, which together with the SR form a 2×2 grid in $p_{\mathrm{T},\ell\ell}$ – n_{trk} space as shown in Figure 11.1.

1. "CR1": This CR has a $e^{\pm}\mu^{\mp}$ pair with $p_{\mathrm{T},\ell\ell} < 30$ GeV and $n_{\mathrm{trk}} = 0$ within the ± 1 mm exclusivity window. The electron and muon here are relatively back-to-back, emphasizing $\tau\bar{\tau}$ final states. The $n_{\mathrm{trk}} = 0$ requirement will emphasize $\gamma\gamma \to \tau^+\tau^-$ production, though the majority of events in this CR will be DY $\tau\bar{\tau}$ due to its higher cross-section. We are also able to check the modelling of the $\tau\bar{\tau}$ backgrounds against $p_{\mathrm{T},\ell\ell}$.

2. "CR2": This CR has a $e^{\pm}\mu^{\mp}$ pair with $p_{\mathrm{T},\ell\ell} > 30$ GeV and $1 \leq n_{\mathrm{trk}} \leq 4$ within the ± 1 mm exclusivity window. Because the leptons are not back-to-back, this region will emphasize the dominant background for the signal, which is inclusive WW. We are also able to double check the n_{trk} modelling.

[3]This is a 2-vector sum: $p_{\mathrm{T},\ell\ell} = \sqrt{(p_{x,e} + p_{x,\mu})^2 + (p_{y,e} + p_{y,\mu})^2}$

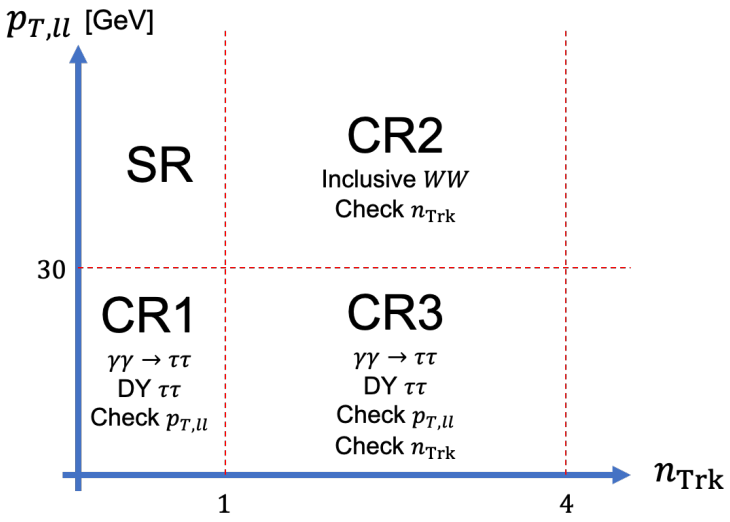


Figure 11.1: Visualization of the layout of the signal region and main control regions in $p_{\mathrm{T},\ell\ell}$ - n_{trk} space.

3. "CR3": This CR has a $e^{\pm}\mu^{\mp}$ pair with $p_{\mathrm{T},\ell\ell} < 30$ GeV and $1 \leq n_{\mathrm{trk}} \leq 4$ within the ± 1 mm exclusivity window. This is the CR that is the least like the SR, strongly emphasizing DY $\tau\bar{\tau}$, so we can check its $p_{\mathrm{T},\ell\ell}$ and n_{trk} shape modelling. In the highest $p_{\mathrm{T},\ell\ell}$ bins of this CR, inclusive WW will also make a sub-dominant, but non-negligible, contribution.

The cuts for both these CRs and the SR are given in Table 11.1, which includes the preselection requirements common to all four regions.

SR	**CR1**	**CR2**	**CR3**
$m_{\ell\ell} > 20$ GeV			
$p_{\mathrm{T},1} > 27$ GeV			
$p_{\mathrm{T},2} > 20$ GeV			
$e^{\pm}\mu^{\mp}$ or $\mu^{\pm}e^{\mp}$			
$p_{\mathrm{T},\ell\ell} < 30$ GeV	$p_{\mathrm{T},\ell\ell} < 30$ GeV	$p_{\mathrm{T},\ell\ell} > 30$ GeV	$p_{\mathrm{T},\ell\ell} < 30$ GeV
$n_{\mathrm{trk}} = 0$	$n_{\mathrm{trk}} = 0$	$1 \leq n_{\mathrm{trk}} \leq 4$	$1 \leq n_{\mathrm{trk}} \leq 4$

Table 11.1: Definitions of the signal region and control regions including preselection requirements, which are common to all four regions.

11.3 Control Plots

As described in the introduction to this chapter, rather than using the three CRs separately to retrieve scale factors for the major backgrounds, a fit will be performed that simultaneously uses the yields in the SR, the three CRs of the preceeding section, and the high-$m_{\ell\ell}$ region of Section 10.1[4]. In this fit, the normalization of the $\gamma\gamma \to W^+W^-$, $\gamma\gamma \to \ell^+\ell^-$, inclusive WW, and DY $\tau\bar{\tau}$ processes are allowed to float in order to maximize the Poisson probability of the yields[5].

However, before performing the fit, we can look at the data and simulation yields in the CRs. Plots of the $m_{\ell\ell}$, $p_{\mathrm{T},\ell\ell}$, and n_{trk} distributions for data and simulation in the the three CRs are given in Figures 11.2, 11.3, and 11.4. With these plots, we can visualize the discrepancy between simulation and data, both in terms of the overall yields and the shapes of the distributions, and validate some of our corrections. For example, the data vs. simulation discrepancy for the n_{trk} distribution for CR3 is flat against the number of tracks, indicating that the UE correction was effective.

In Figure 11.2, the Herwig 7 elastic $\gamma\gamma \to W^+W^-$ events have been scaled by S_{Excl} to account for SD and DD contributions. In Figures 11.3 and 11.4, S_{Excl} is not applied, with the SD and DD contributions included through the MG5+Pythia8 simulation, corrected as described in Section 10.2. The PU and UE corrections are applied as described in Chapters 7 and 8.

Events with fake leptons are included in these plots, represented by the gray areas. In simulation, a lepton truth-matching scheme has been applied to ensure that the estimate of fake leptons is not duplicated after applying the fake-factor method of Chapter 9. The truth-matching checks that a reconstructed electron (muon) corresponds to a truth-level electron (muon) from the simulated hard scatter. From Table 11.2, it can be seen that events with fakes are expected to comprise about 7% of the SR, 13% of CR1, 13% of CR2, and 12% of CR3.

A few observations can be drawn from these plots:

- CR1 is be well modelled, especially after accounting for events with fake leptons.
- The n_{trk} plot of Figure 11.3 shows a relatively flat discrepancy between of about 10% for n_{trk} bins 2–4, but a discrepancy of about 30% for n_{trk} bin 1. The $m_{\ell\ell}$ and $p_{\mathrm{T},\ell\ell}$ distributions appear relatively well-modelled. The normalization exhibits an acceptable level of agreement, as the uncertainty on the inclusive WW yield from simulation is a bit over 10% due to the application of the UE correction[6].

[4]Please see Chapter 13

[5]The CRs are relatively pure in the background they are designed to isolate, so this fit is almost equivalent to normalizing the backgrounds in their respective CRs. For example, we could extract a correction for the normalization of inclusive WW based on CR2 alone.

[6]Please see Chapter 12, particularly Section 12.2.6.

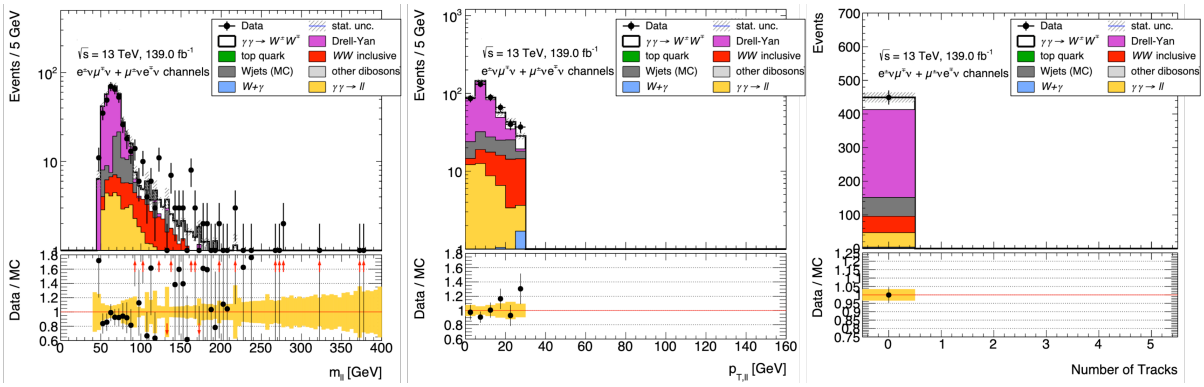


Figure 11.2: Distributions for the $m_{\ell\ell}$, $p_{\mathrm{T},\ell\ell}$, and n_{trk} values of simulated and data events that fall in "CR1", which has $p_{\mathrm{T},\ell\ell} < 30$ GeV and $n_{\mathrm{trk}} = 0$ requirements. Events with fake leptons are plotted in gray.

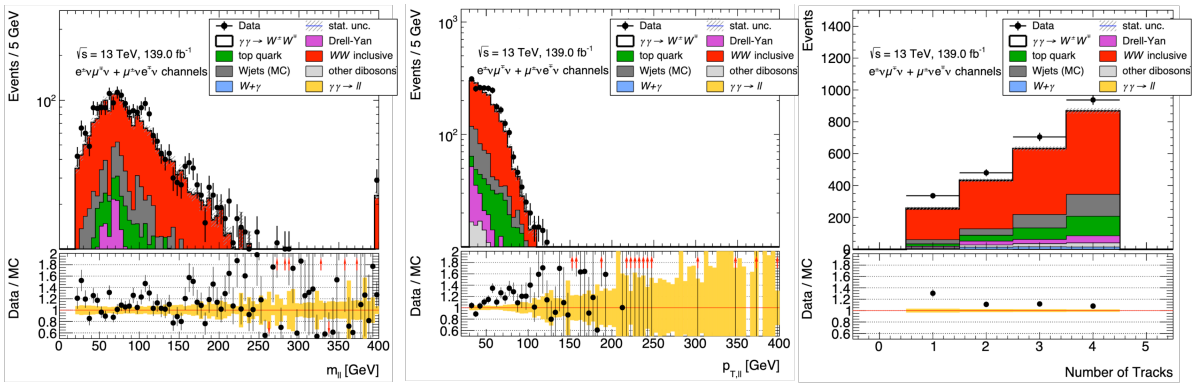


Figure 11.3: Distributions for the $m_{\ell\ell}$, $p_{\mathrm{T},\ell\ell}$, and n_{trk} values of simulated and data events that fall in "CR2", which has $p_{\mathrm{T},\ell\ell} > 30$ GeV and $1 \leq n_{\mathrm{trk}} \leq 4$ requirements. Events with fake leptons are plotted in gray. The predominance of inclusive WW events in this region is clear.

- CR3 is relatively well-modelled up to an overall normalization discrepancy of about 10%. There is good agreement in the shapes of the $p_{\mathrm{T},\ell\ell}$ and n_{trk} distributions comparing data to simulation.

In general, these plots do not suggest that additional specific corrections must be made before proceeding to the fit.

The pre-fit yields of the major processes in the CRs are given in Table 11.2 for both data and simulation. For the SR column of this table, only the simulated yields are presented. Uncertainties are statistical only. In the SR and CR1, the $\gamma\gamma \to W^+W^-$ estimate uses the

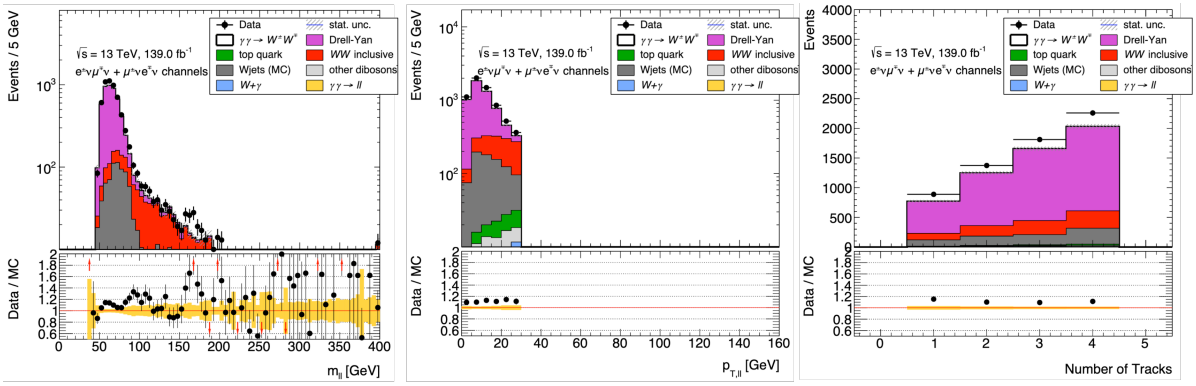


Figure 11.4: Distributions for the $m_{\ell\ell}$, $p_{\mathrm{T},\ell\ell}$, and n_{trk} values of simulated and data events that fall in "CR3", which has $p_{\mathrm{T},\ell\ell} < 30$ GeV and $1 \leq n_{\mathrm{trk}} \leq 4$ requirements. Events with fake leptons are plotted in gray.

S_{Excl} scale factor from Chapter 10, and for CR2 and CR3, the MG5+Pythia8 simulation is used, corrected as described in Section 10.2.

Process	**SR**	**CR1**	**CR2**	**CR3**
$\gamma\gamma \to W^+W^-$	133 ± 1.1	35.3 ± 0.6	97.8 ± 1.1	25.2 ± 0.6
$\gamma\gamma \to \ell^+\ell^-$	5.7 ± 0.2	41.5 ± 0.6	8.0 ± 1.2	42.0 ± 2.0
Inclusive WW	88.4 ± 1.6	48.9 ± 0.9	1422 ± 6.7	816 ± 4.8
DY $\ell^+\ell^-$	4.3 ± 1.3	261 ± 10.0	102 ± 7.4	4053 ± 49.2
$W + \gamma$	2.8 ± 1.6	2.0 ± 1.1	45.9 ± 6.4	36.0 ± 5.0
Fakes	17.6 ± 5.3	56.0 ± 11.1	289 ± 22.9	704 ± 42.1
Others	7.1 ± 0.8	2.0 ± 0.3	314 ± 6.2	82 ± 2.6
Total Sim.	259 ± 6.2	447 ± 15.0	2279 ± 26.6	5758 ± 65.2
Background Sim.	126 ± 6.2	412 ± 15.0	2181 ± 26.6	5732 ± 65.2
Data	–	449	2458	6332

Table 11.2: Data and simulation yields within the three main control regions, along with the simulation yields for the signal region. Uncertainties are statistical only. "Other" backgrounds include WZ and ZZ production and top-quark production.

Chapter 12

Systematics and Uncertainties

In order to determine the statistical significance of a potential observation of $\gamma\gamma \to W^+W^-$, we must understand the various sources of uncertainty in the analysis. This analysis suffers from both statistical and systematic uncertainties. The statistical uncertainties, discussed in Section 12.1, arise from the fact that have only a finite number of data and simulated events. The systematic uncertainties, described in Section 12.2, are caused by potential flaws in our simulation. Systematic uncertainties associated with several of the corrections derived throughout Chapters 7 – 10 will be discussed.

12.1 Statistical Uncertainties

An unavoidable component of uncertainty in any analysis is statistical. The yields of data or simulated events that fall into a region are typically assumed to follow the Poisson distribution, which is a probability distribution for the number of times a phenomenon will occur in a given interval if each occurrence is independent of the last and there is a constant mean rate. In the context of this analysis, the "interval" is interchangeable with integrated luminosity; every proton-proton collision at the LHC is independent, and there is a given probability for any process to occur in an individual proton-proton interaction.

This is to say, if we exactly duplicated Run 2 of the LHC infinitely many times, then the distribution of the number of observed events in our signal region should follow the Poisson distribution, which is governed by a single parameter, normally called λ. With infinitely many Run 2's, we could determine λ exactly; for the Poisson distribution, the mean observed value will actually be λ. Of course, there was only one Run 2, so its yield is our best estimate for the "true" mean value of yields[1]. The variance of yields for the Poisson distribution is also λ, and the standard deviation is $\sqrt{\lambda}$. Therefore, if we observe N events in a region, we typically assign that value a statistical uncertainty of $\sqrt{N}$[2]. As N

[1] And therefore λ

[2] In the case that λ is large, the Poisson distribution can be approximated by a Gaussian distribution with mean of λ and standard deviation of $\sqrt{\lambda}$. In this case, uncertainties will be symmetric. The unweighted

gets larger, the relative uncertainty decreases. This concept is why large simulated samples are preferable[3].

12.2 Systematic Uncertainties

Statistical uncertainties are unavoidable, but their relative importance is reduced with larger sample sizes. This is not the case for systematic uncertainties, which attempt to capture uncertainty due to potential flaws in simulation, such as poorly modelled resolution, biases in reconstruction efficiencies, or uncertainties on theoretical cross-section calculations.

12.2.1 Lepton, Trigger, and Tracking Uncertainties

Systematic uncertainties are assigned for the efficiencies of electron and muon reconstruction, for their energy and momentum resolution, and for the efficiencies of the identification and isolation variables [186, 130]. The values of the electron and muon uncertainties for the $\gamma\gamma \to W^+W^-$, $\gamma\gamma \to \ell^+\ell^-$, inclusive WW, and Drell-Yan (DY) $\ell^+\ell^-$ simulations in the signal region and control regions are given in Table 12.1. These are evaluated by varying the associated values in simulation to determine the impact on the eventual yield. Uncertainties in these regions for other samples are generally similar.

Similarly, systematic uncertainties due to the triggers are determined by varying the trigger efficiencies in simulation to observe the impact on the yields[4] [112, 113]. Generally, these uncertainties are below 0.25%.

Because of the importance of tracking in this analysis, its associated uncertainties are especially relevant. Uncertainties are assigned to account for:

1. Track parameter resolution (particularly z and d_0 resolution).
2. Potential systematic biases on tracks' z, d_0, and p_{T} parameters. This could occur due to mis-alignment of the inner detector when comparing simulation and data.
3. Reconstruction efficiency uncertainties.
4. Uncertainty in the track fake rate.

In order, they are assessed by:

simulated samples used in this analysis typically have large enough yields that this approximation is valid. The data yield in each region is above 300 events, so this approximation is valid for data as well.

[3]It is best to generate a significantly larger number of events than expected, and assign them a weight such that the sum of the weights matches the physical cross-section, as indicated in Equation 5.1.

[4]The variation is generally a few percent or less, depending on lepton p_{T}. Standard ATLAS prescriptions were used to set the variation, which are based on discrepancies in trigger efficiencies measured in data vs. simulation.

	SR	**CR1**	**CR2**	**CR3**
	$\gamma\gamma \to W^+W^-$			
Electron Syst.	0.39%	0.37%	0.39%	0.39%
Muon Syst.	0.43%	0.64%	0.43%	0.54%
	$\gamma\gamma \to \ell^+\ell^-$			
Electron Syst.	0.73%	0.64%	0.77%	1.56%
Muon Syst.	0.79%	0.52%	0.67%	2.35%
	Inclusive WW			
Electron Syst.	0.39%	0.44%	0.43%	0.43%
Muon Syst.	0.42%	0.43%	0.42%	0.59%
	Drell-Yan $\ell^+\ell^-$			
Electron Syst.	0.49%	1.15%	1.61%	0.55%
Muon Syst.	0.38%	0.76%	2.71%	0.61%

Table 12.1: Systematic uncertainties for lepton reconstruction efficiencies, energy and momentum resolution, and identification and isolation tagging efficiencies. Values are given for the major samples in the signal region and main control regions discussed in Chapter 11.

1. Smearing the z and d_0 values of tracks in simulation[5]. This is done in all simulated samples.

2. Biasing the tracks' z, d_0, and p_{T} in the simulation[6]. This is done in all simulated samples.

3. Removing good[7] tracks to replicate efficiency loss. This method is valid for simulated samples that have pileup-truth information[8]: $\gamma\gamma \to W^+W^-$, $\gamma\gamma \to \ell^+\ell^-$, $Z \to \tau\tau$, and inclusive WW. These are the most important simulated samples in the analysis. Track removal is also performed for samples without pileup-truth information, but some fake pileup tracks will be removed inadvertently. However, the samples that don't have pileup-truth information are those corresponding to subdominant backgrounds. Conservative theory uncertainties are assigned to these backgrounds (see Section 12.2.6), so tracking uncertainties will be subdominant.

4. Removing all fake tracks. Similar to the above discussion, this is only valid for samples with full pileup-truth; the impact of the fake rate will be underestimated for samples

[5]The variation is performed within track resolution, as discussed in Section 4.1

[6]Biased within track resolution, as discussed in Section 4.1

[7]A "good" track here is one that is truth-matched to a real particle [206]; in other words, it is not a fake track.

[8]Most simulated samples in ATLAS do not retain this information, as it is extraneous to most analyses.

without pileup-truth. As above, the conservative theory uncertainties will dominate uncertainties from the fake rate for these subdominant processes.

The uncertainty is assigned based on the change's impact on the yield in various regions. It is symmetrized, which is important for uncertainties associated with efficiency and fake rate, which are difficult to artificially increase. The uncertainties due to resolution and biases, efficiency, and fake rate are shown in Table 12.2.

	SR	**CR1**	**CR2**	**CR3**
	$\gamma\gamma \to W^+W^-$			
Resolution and Bias	0.78%	0.88%	1.34%	0.06%
Efficiency	0.22%	0.18%	0.19%	0.26%
Fake Rate	0.42%	0.56%	0.14%	0.13%
	$\gamma\gamma \to \ell^+\ell^-$			
Resolution and Bias	0.71%	0.86%	0.67%	2.20%
Efficiency	0.00%	0.24%	0.79%	0.04%
Fake Rate	0.47%	0.55%	0.18%	3.57%
	Inclusive WW			
Resolution and Bias	1.63%	0.20%	0.18%	0.42%
Efficiency	2.34%	2.04%	1.74%	1.77%
Fake Rate	4.90%	5.56%	3.54%	3.11%
	Drell-Yan $\ell^+\ell^-$			
Resolution and Bias	0.00%	0.75%	0.70%	0.51%
Efficiency	0.00%	1.97%	4.88%	1.37%
Fake Rate	0.00%	4.36%	6.88%	2.63%

Table 12.2: Systematic uncertainties for track parameter resolution and potential biases, track reconstruction efficiency, and fake rate. Values are given for the major samples in the signal region and main control regions discussed in Chapter 11.

12.2.2 Pileup Modelling Uncertainties

As described in Chapter 7, three main corrections are applied to simulation as pileup corrections:

- Correcting the distribution of the number of pileup interactions per bunch crossing in simulation to match that observed in data.
- Correcting the longitudinal size of the beamspot in simulation to match that in data.

- Correcting the number of pileup tracks ($n_{\text{trk}}^{\text{PU}}$) observed in the ± 1 mm window around the lepton vertex, which is derived by correcting the distribution in given random windows in the beamspot.

The uncertainty due to the first correction is evaluated by varying the scale factors used to correct the simulation by 4%, corresponding to the scale factors' uncertainty. The resulting uncertainties are less than 0.5% for most samples in all regions, but are about 2% for $\gamma\gamma \to \ell^+\ell^-$ events in CR3.

An uncertainty for the beamspot length and pileup track corrections is assigned based on information similar to that presented in Figure 7.2. If the simulated sample has a beamspot width of 35 mm to begin with, then a 1% uncertainty on the yield with $n_{\text{trk}}^{\text{PU}} = 0$ and a 2% uncertainty on the yield with $1 \leq n_{\text{trk}}^{\text{PU}} \leq 4$ are assigned. If the simulated sample has a beamspot width of 42 mm to begin with, then a 3% uncertainty on the yield with $n_{\text{trk}}^{\text{PU}} = 0$ and a 10% uncertainty on the yield with $1 \leq n_{\text{trk}}^{\text{PU}} \leq 4$ are assigned. For the most part, the photon induced samples were produced with a beamspot of 35 mm, and the QCD-induced processes were produced with a 42 mm beamspot. The values of these uncertainties are motivated by the fractional difference between simulation and data prior to the application of the track multiplicity correction, visualized as the short-dashed lines in Figure 7.2.

12.2.3 Underlying Event Correction Uncertainties

Several uncertainties result from the correction for the number of charged particles produced in QCD-induced backgrounds' underlying events described in Chapter 8. We should recall that the correction was derived in the Z-peak region, involving several steps before deriving an eventual reweighting: subtraction of photon-induced contributions, subtracting pileup tracks, and unfolding the number of reconstructed tracks to the truth-level number of charged particles. Uncertainties are derived by

- Studying the non-closure between the unfolded and truth-level simulation
- Varying the mass range of the Z-peak region and the $m_{\ell\ell}$ lower-bound of the region used for the photon-induced subtraction
- Using the uncertainties from the fit performed to subtract photon-induced events and the difference between a muon-only fit and an electron-only fit
- Using Sherpa in the place of Powheg+Pythia8 as the Drell-Yan template for the photon-induced event subtraction
- Using Sherpa in the place of Powheg+Pythia8 as the sample used to derive the $P(n_{\text{trk}}^{\text{UE}} = k \mid n_{\text{ch}} = n)$ PDF for the reconstructed-track to truth-particle unfolding
- Studying the difference between the predictions of the Powheg+Pythia8 and Sherpa $q\bar{q} \to W^+W^-$samples after the application of the correction.

The largest of these uncertainties are due to the non-closure between unfolded and truth-level simulation and the difference between POWHEG+PYTHIA8 and SHERPA in the reconstructed-track to truth-particle unfolding. These uncertainties both amount to about 2.5% for both Drell-Yan and inclusive WW modelling in most regions, though for Drell-Yan events in the signal region, the uncertainty due to the unfolding is about 5%.

12.2.4 Uncertainties for Fake Lepton Estimates

The estimation of the number of events with fake leptons described in Chapter 9 carries large and conservative uncertainties. In the method used, a fake factor is found as the ratio between "passing" fake leptons and "failing" fake leptons, where passing and failing refer to whether the leptons pass our isolation and identification criteria. The fake factor is derived in a same-charge lepton region. The impacts of three changes in the derivation are considered for evaluating uncertainties on the fake factor:

- The fake factors were rederived in regions that required $p_{\mathrm{T},\ell\ell} > 30$ GeV (no such cut is used in the baseline derivation).
- The fake factor is rederived for events with $n_{\mathrm{trk}} < 10$ ($1 \leq n_{\mathrm{trk}} \leq 4$ is used in the baseline derivation).
- To conservatively account for the systematic and statistical uncertainties associated with the subtraction of events with two genuine leptons in the regions used for the fake factor derivation, the number of events subtracted in each region is varied up and down by 50%.

The uncertainty associated with genuine lepton subtraction is the largest.

The fake factor is varied within its uncertainties to evaluate their effect on the expected yields of events with fake leptons in the SR and CRs. In the SR, this leads to a cumulative uncertainty of 140%, and uncertainties of about 130% for the CRs.

12.2.5 Uncertainty on Dissociative Production

Several sources of systematic uncertainty are assessed for the derivation of S_{Excl} discussed in Chapter 10. First, we should recall that we used high-$m_{\ell\ell}$ same-flavor lepton events in the scale factor's derivation. S_{Excl}itself was the fraction of all photon-induced events in the region to simulated elastic-only events. Background events were subtracted using a template method, where the template came from events with 5 tracks within ± 1 mm of the lepton vertex, being normalized in the Z-peak region. Essentially, this amounts to three major choices in the derivation:

- The lower bound of the mass for the high-$m_{\ell\ell}$ region,
- The choice of $n_{\mathrm{trk}} = 5$ for the template spectrum,

- The choice of 83.5 GeV $< m_{\ell\ell} <$ 98.5 GeV for the Z-peak region width.

Unsurprisingly, the uncertainty for the lower bound of the high-$m_{\ell\ell}$ region is studied by changing the region's lower bound. The bound is varied in steps of 10 GeV from 110 GeV to 400 GeV. An 11% uncertainty was assigned based on this study.

The uncertainties due to the second two choices are assessed by using a template based on the spectrum for events with $n_{\text{trk}} = 2$ and by using a variety of window sizes for the Z-peak. The final uncertainty evaluated for the Z-peak systematic comes from the window size 73.5 GeV $< m_{\ell\ell} <$ 108.5 GeV. Both the $n_{\text{trk}} = 2$ spectrum and the larger Z-peak window will have larger contamination from photon-induced events. Combined the uncertainties amount to 4.2%.

Cumulatively, there is a 15% uncertainty on S_{Excl}.

Uncertainty is also assigned for the estimation of photon-induced events with $1 \leq n_{\text{trk}} \leq 4$. These are evaluated by comparing the expected yields of $\gamma\gamma \to \ell^+\ell^-$ from LPAIR and MG5+ Pythia8, as these generators can produce elastic, single-dissociative, and double-dissociative events. The difference between the generators is 20% for elastic events and 30% for single-dissociative events. For double-dissociative events, a 100% uncertainty is assigned. Cumulatively, the uncertainty for photon-induced events in the $1 \leq n_{\text{trk}} \leq 4$ control regions comes to 20%.

While systematics on the predicted $\gamma\gamma \to W^+W^-$ yield in the SR do not affect the observation of the process, these systematics are important in a comparison of a measured cross-section for $\gamma\gamma \to W^+W^-$ to the Standard Model prediction.

12.2.6 Theory Uncertainties

The main theoretical uncertainties in this analysis come from the modelling of QCD-induced processes with few underlying event tracks. The Drell-Yan and $q\bar{q} \to W^+W^-$ processes comprise the bulk of expected backgrounds, and they were studied in detail to derive specific corrections. Both of these processes have $q\bar{q}$ initial states, but no similar correction was made for processes with initial-state gluons. Such processes form subleading contributions to the signal and control regions, and assigning conservatively large uncertainties does not lead to a large overall uncertainty on the background yield.

Backgrounds involving top quarks were studied in some detail, as outlined in Section 9.2. A 30% uncertainty was assessed for these backgrounds. A 30% uncertainty is also applied to vector-boson fusion induced WW production and other gluon-fusion induced WW production mechanisms, as these processes either have a similar initial state to the main top backgrounds or have forward jets, similar to those studied in the 2 forward-jet control region.

A 100% uncertainty was applied to the yield of $W + \gamma$ events, as these events are difficult to isolate and study directly.

Uncertainties on the Drell-Yan, $q\bar{q} \to W^+W^-$, $q\bar{q} \to WZ$, and $q\bar{q} \to ZZ$ yields are assessed by comparing the yields between the "main" and "alternate" generators in Table 5.2. For both $q\bar{q} \to W^+W^-$ and Drell-Yan, Powheg+Pythia8 is the main generator, with

POWHEG+HERWIG7 and SHERPA serving as alternates. The underlying event correction is applied to all of these samples[9], as they have similar initial states, but the uncertainties are assigned using different procedures.

The main and alternate samples exhibit greater agreement for the Drell-Yan process than for the $q\bar{q} \to W^+W^-$ process. Therefore, for Drell-Yan production, the POWHEG prediction is taken as the central value, with the SHERPA-to-POWHEG difference serving as both an upwards and downwards uncertainty. This method is also applied to WZ and ZZ production, as these samples also demonstrate adequate agreement between the generators.

The SHERPA-to-POWHEG difference is noticeably bigger for $q\bar{q} \to W^+W^-$ events with $n_{\text{trk}} = 0$. This discrepancy is shown in Figure 12.1. It can be seen that the two POWHEG-based predictions are mutually closer to each other than to the SHERPA yield. In this case, taking POWHEG's as the baseline prediction with the difference to SHERPA as a symmetric uncertainty, then the uncertainty would extend into an implausibly high regime. Two alternatives were considered:

1. Taking the POWHEG+PYTHIA8 prediction as the baseline prediction for $q\bar{q} \to W^+W^-$ with a one-sided uncertainty extending to the SHERPA.

2. Taking the average of the POWHEG+PYTHIA8 and SHERPA yields as the baseline $q\bar{q} \to W^+W^-$ yield, and using the difference between this central value and the generators' prediction[10] as a symmetric uncertainty.

The latter choice was used in this analysis. In the $n_{\text{trk}} = 0$ bin, the POWHEG+HERWIG7 yield is actually 2% higher than the POWHEG+PYTHIA8 yield, so in this bin, the final prediction is shifted upwards by 2%. With this method, the uncertainty is 7% for events with $n_{\text{trk}} = 0$, and less than 1% for events with $1 \leq n_{\text{trk}} \leq 4$.

An additional uncertainty is assigned for the $p_{\text{T},WW}$ reweighting procedure described in Section 8.7 and for the overall $q\bar{q} \to W^+W^-$ cross-section, coming to about 5%.

[9]The correction is derived for each generator+shower combination separately.

[10]Indicated by the grey band in Figure 12.1

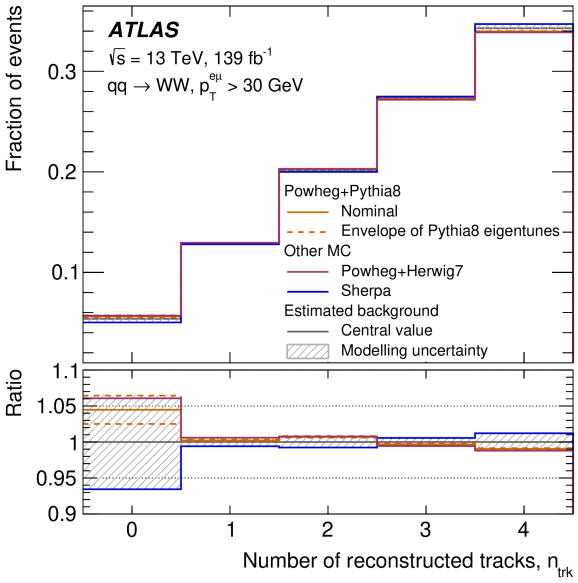


Figure 12.1: Normalized modelling differences between POWHEG+PYTHIA8, POWHEG+HERWIG7, and SHERPA for the $q\bar{q} \rightarrow W^+W^-$ process with fewer than 5 reconstructed tracks near the lepton vertex [187]. The prediction for the $q\bar{q} \rightarrow W^+W^-$ yield comes from taking the average value of POWHEG+PYTHIA8 and SHERPA, with uncertainties assigned based on the envelope defined by the generators' respective yields.

Chapter 13

Fit and Results

At this point, our simulation is fully corrected, and we understand its systematic uncertainties. We are ready to determine the statistical significance of the $\gamma\gamma \to W^+W^-$ process in the signal region and to measure its cross-section. Much of Chapters 7 – 10 was framed as an attempt to make an accurate prediction of $N_{\text{bkgd,SR}}$. What remains for this determination is to check the normalization of the major backgrounds using the control regions. If there is an overall discrepancy between the data and simulation[1], we will scale the normalization of the background simulations appropriately. One option for this would be to treat the CRs separately. For example, because CR2 is designed to emphasize inclusive WW, we could use that region alone to find a normalization correction for the simulation of inclusive WW.

However, we should note that *all* of our backgrounds will contribute events to each of the signal and control regions (see Table 11.2). For example, while DY events are expected to dominate in CR1, inclusive WW events and $\gamma\gamma \to \ell^+\ell^-$ events also make important contributions. Even $\gamma\gamma \to W^+W^-$ has a non-negligible expected yield in this region. Because of this cross-contamination, we do not want to take single control regions to derive normalization scale factors for single processes. We will instead perform a fit using the signal region, the three control regions defined in Chapter 11, and the high-$m_{\ell\ell}$ control region used for the derivation of S_{Excl}. In this fit, we will allow the overall normalization of the $\gamma\gamma \to W^+W^-$, $q\bar{q} \to W^+W^-$, DY $\ell^+\ell^-$, and $\gamma\gamma \to \ell^+\ell^-$ processes to float, seeking to maximize the Poisson likelihood of the data observed in the five regions. Because there are five regions and four free parameters, this is an overconstrained fit. Systematic uncertainties will be included in this fit as nuisance parameters.

13.1 Fit Overview

In Section 12.1, it was explained that the probability distribution governing the expected yields of different processes in the various analysis regions is the Poisson distribution. This

[1] Such as that seen in Figure 11.3

distribution tells us that the likelihood to observe a value k is

$$P(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}. \tag{13.1}$$

Here, λ is the defining parameter of the distribution. If observations were made many times, then the mean value of k would actually be λ. In Section 3.1.3, we discussed that the expected yield due to a process is the product of the integrated luminosity and the cross-section: $N = \sigma * \mathcal{L}$. This suggests that, for our purposes, λ is related to the cross-section of a process.

Given the definition of the Poisson distribution in Equation 13.1, the likelihood to observe N_{obs} events in a particular region will be:

$$L(N_{\text{obs}}) = \frac{(N_{\text{sig}} + N_{\text{bkgd}})^{N_{\text{obs}}} \times e^{-(N_{\text{sig}}+N_{\text{bkgd}})}}{(N_{\text{obs}})!}. \tag{13.2}$$

Here, λ, which is the expected yield, has been replaced by $N_{\text{sig}} + N_{\text{bkgd}}$, and k has been replaced by N_{obs}. The above equation neglects systematic uncertainties.

In Equation 13.2, N_{sig} is the expected yield of the $\gamma\gamma \to W^+W^-$ process based on simulation. For the fit, we will let $N_{\text{sig}} = \mu \times \sigma \times \mathcal{L} \times A \times \varepsilon_S$. Here, σ is the predicted cross-section, and μ is the normalization scaling factor that can float in the fit. The A and ε_S factors are the expected kinematic acceptance of the signal and its selection efficiency, respectively. Similarly, in Equation 13.2, N_{bkgd} is the expected yield of the background processes (if the region is the signal region, this is $N_{\text{bkgd,SR}}$). In the fit, this is simply set to $N_{\text{bkgd}} = \sum_b \beta_b N_b$, where the sum runs over the background processes in Table 11.2, with N_b being the yields in the table. The β_b values can modify the normalization of the background processes, and are allowed to deviate from 1 for the $\gamma\gamma \to \ell^+\ell^-$, $q\bar{q} \to W^+W^-$, and DY processes.

To add in the effect of systematic uncertainties, we must modify Equation 13.2. The systematic uncertainties could modify the expected signal yield by a factor of $(1+\sum_f x_f S_f)$, and they could modify the background yields by a factor of $(1 + \sum_f x_f B_f)$. In both cases, the sum over f runs over the relevant sources of systematic uncertainty, with S_f being the fractional uncertainty on the signal for the uncertainty corresponding to x_f and B_f being the corresponding fractional uncertainty on the background. The likelihoods for the x_f values follow normal distributions which have an average value of 0 and a standard deviation of 1. These are called "nuisance parameters" and cause a systematic shift in expected yields depending on their value. Each x_f is allowed to float in the fit. The cumulative likelihood for a given N_{obs} in a single region when accounting for systematics is

$$L(N_{\text{obs}}) = \frac{(N_{\text{sig}} + N_{\text{bkgd}})^{N_{\text{obs}}} \times e^{-(N_{\text{sig}}+N_{\text{bkgd}})}}{(N_{\text{obs}})!} \times \prod_f \frac{1}{\sqrt{2\pi}} e^{-\frac{x_f^2}{2}}. \tag{13.3}$$

Now, $N_{\text{sig}} = \mu \times \sigma \times \mathcal{L} \times A \times \varepsilon_S \times (1 + \sum_f x_f S_f)$, $N_{\text{bkgd}} = \sum_b (\beta_b N_b \times (1 + \sum_f x_f B_f^b))$, and the final product factor accounts for the probabilities of choosing particular x_f values.

To clarify, if the fit finds a positive value for an x_f value, that suggests that the N_b value associated with that systematic was overpredicted by the simulation.

The total likelihood across the five regions will be a product of five distributions like that in Equation 13.3. However, when attempting to find μ, $\beta_{\gamma\gamma\to\ell\ell}$, $\beta_{q\bar{q}\to WW}$, and β_{DY} values to maximize the five-region likelihood, it is customary to instead *minimize* the "negative log likelihood". When we take the natural log of the likelihood function, products become sums, and factors such as $1/\sqrt{2\pi}$ become irrelevant, making the math simpler. Therefore, the value that we are trying to minimize is:

$$-\ln L(N_{\{\mathrm{SR,CRs}\}}) = \sum_i -\ln\left(\frac{(N^i_{\mathrm{sig}}+N^i_{\mathrm{bkgd}})^{N^i_{\mathrm{obs}}}\times e^{-(N^i_{\mathrm{sig}}+N^i_{\mathrm{bkgd}})}}{(N^i_{\mathrm{obs}})!}\right)+\sum_f \frac{x_f^2}{2}. \tag{13.4}$$

The sum over i runs over the five regions. The μ, $\beta_{\gamma\gamma\to\ell\ell}$, $\beta_{q\bar{q}\to WW}$, and β_{DY} values are shared by each region, as are the nuisance parameters. The final sum over the nuisance parameters comes from taking the logarithm of the product of the normal distributions and dropping the irrelevant coefficient terms.

The exclusivity scale factor is actually derived in this fit rather than treating the high-$m_{\ell\ell}$, same-flavor region as truly independent[2]. The background in that region *is* estimated using the template method, but deriving S_{Excl} in the five region fit allows the $\gamma\gamma\to\ell^+\ell^-$ yields in the other regions, which are correlated, to influence the final value of S_{Excl}. In the CRs with $1\le n_{\mathrm{trk}}\le 4$, the $\gamma\gamma\to\ell^+\ell^-$ yields are treated separately, as those estimates do not rely on S_{Excl}.

The fit itself is performed using the HistFactory program described in [207]. The free parameters in the fit are the μ, $\beta_{\gamma\gamma\to\ell\ell}$, $\beta_{q\bar{q}\to WW}$, and β_{DY} values and the x_f nuisance parameters.

13.2 Fit Results

13.2.1 Blinded Fit Validation

Prior to the unblinding of the SR, two studies are performed to validate the fit and determine expected results:

- The fit is performed over the four control regions alone, without the SR included[3].
- The fit is performed with an Asimov dataset [208] that uses μ fixed at 1 for the expected signal yield and the $\beta_{\gamma\gamma\to\ell\ell}$, $\beta_{q\bar{q}\to WW}$, β_{DY}, and nuisance parameter values derived in the preceding bullet to be as realistic as possible.

[2] $\beta_{\gamma\gamma\to\ell\ell} = S_{\mathrm{Excl}}$.

[3] In this fit, the parts of Eq. 13.4 pertaining to the SR are not included.

When the fit is first performed to the CRs alone, the $\beta_{\gamma\gamma\to\ell\ell}$, $\beta_{q\bar{q}\to WW}$, and β_{DY} values are found to be:

$$\begin{aligned} \beta_{\gamma\gamma\to\ell\ell} = S_{\text{Excl}} &= 3.59^{+0.15}_{-0.15} \\ \beta_{q\bar{q}\to WW} &= 1.15^{+0.21}_{-0.23} \\ \beta_{DY} &= 1.14^{+0.12}_{-0.12}. \end{aligned} \tag{13.5}$$

The fitted values of the x_f parameters associated with systematic uncertainties are presented in Figure 13.1. The sources of uncertainties were discussed in Chapter 12. The "Electrons", "Muons", and "Tracks" sets are discussed in Section 12.2.1. The "Nchrew" set are the uncertainties associated with the UE correction, discussed in Section 12.2.3. The "Misidlep" set is associated with the estimation of events with fake leptons, discussed in Section 12.2.4. The "Pileup" set is associated with the pileup corrections, discussed in Section 12.2.2. The "DataDriven" set is associated with the estimation of the contribution of dissociative events in regions with $n_{\text{trk}} = 0$, discussed in Section 12.2.5. The "Theory" uncertainties were discussed in Section 12.2.6. Lastly, the "Luminosity" uncertainty is simply that due to the uncertainty on the delivered good Luminosity, which is 2.4 fb^{-1}, or 1.7%.

Next, an Asimov dataset is generated using the β values of Eq. 13.5, $\mu = 1$, and the x_f values that were more than 3% away from 0 in the CR-only fit. The μ value resulting from the fit to this dataset is $\mu = 0.99^{+0.13}_{-0.12}(\text{stat.})^{+0.17}_{-0.14}(\text{syst.})$, with an expected significance of 6.7σ.

The fit parameters with the greatest impact on the uncertainty of μ are shown in Figure 13.2. The most important values here are

1. "EXCL_ApplyToWW": The systematic uncertainty associated with changing the lower bound of the mass for the high-$m_{\ell\ell}$ region (see Section 12.2.5).

2. "γ (SR bin 0)": The statistical uncertainty of the background prediction in the SR. This is mostly due to the estimate of events with fakes

3. "THEO_WW": The systematic uncertainty associated with the inclusive WW yield after applying the UE correction.

4. "ExclDilepOverSM": The uncertainty associated with $\beta_{\gamma\gamma\to\ell\ell} = S_{\text{Excl}}$, which is used to scale the elastic-only prediction of $\gamma\gamma \to W^+W^-$.

5. "FAKE_NOBKG": The uncertainty associated with the subtraction of events with two genuine leptons in the the fake factor derivation (see Section 12.2.4).

13.2.2 Unblinded Fit

The yields in Table 11.2 and the systematics described throughout Chapter 12 are used to perform the fit over the SR and four CRs. Several key features of the fit result are worth highlighting.

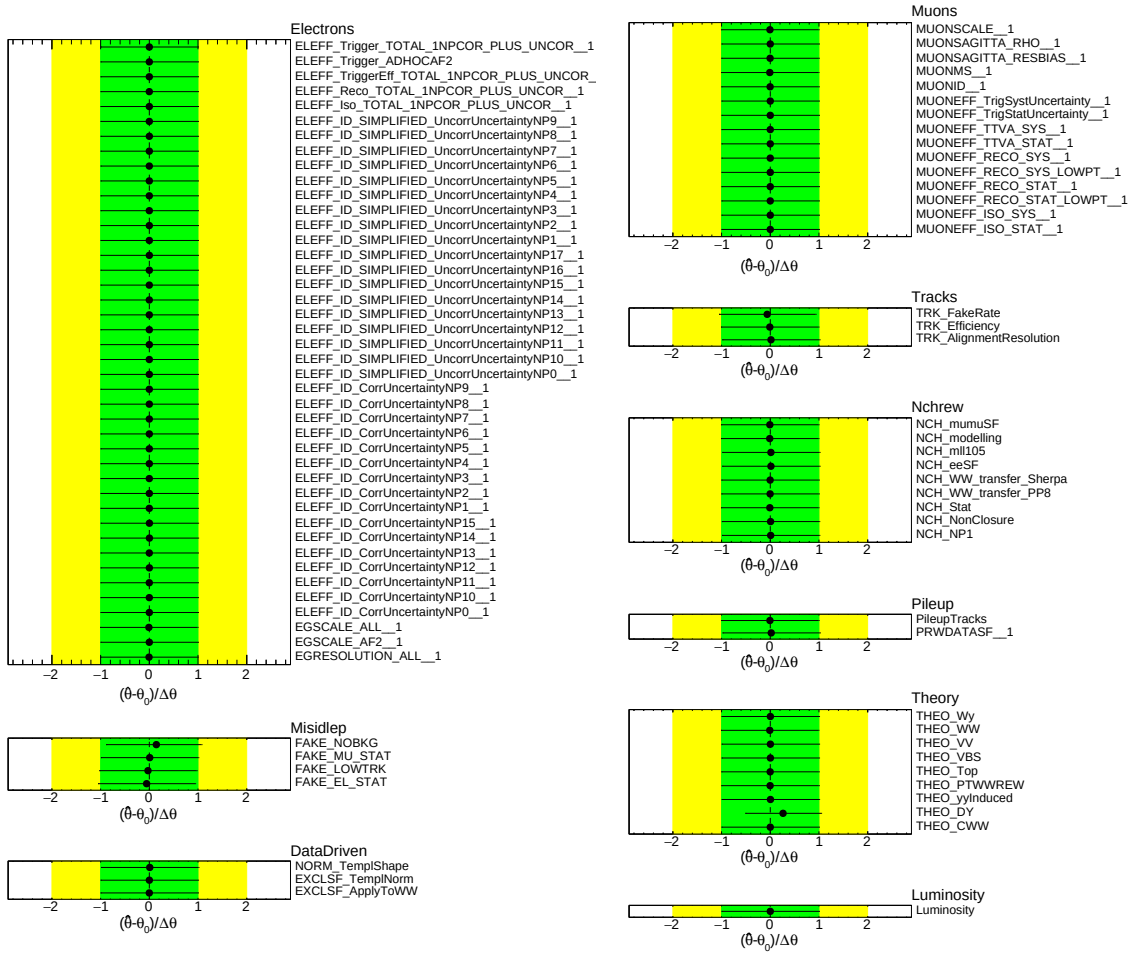


Figure 13.1: Fitted values and uncertainties of the x_f parameters in Eq. 13.4 for the Control Region-only fit. Each x_f value is associated with a systematic uncertainty, and the majority of the values are within ± 0.05 of their nominal value of 1. The uncertainties associated with the largest x_f values are "FAKE_NOBKG" and "THEO_DY". FAKE_NOBKG is the uncertainty associated with the subtraction of events with two genuine leptons in the the fake factor derivation. THEO_DY is the uncertainty associated with the Drell-Yan yield after the application of the UE correction.

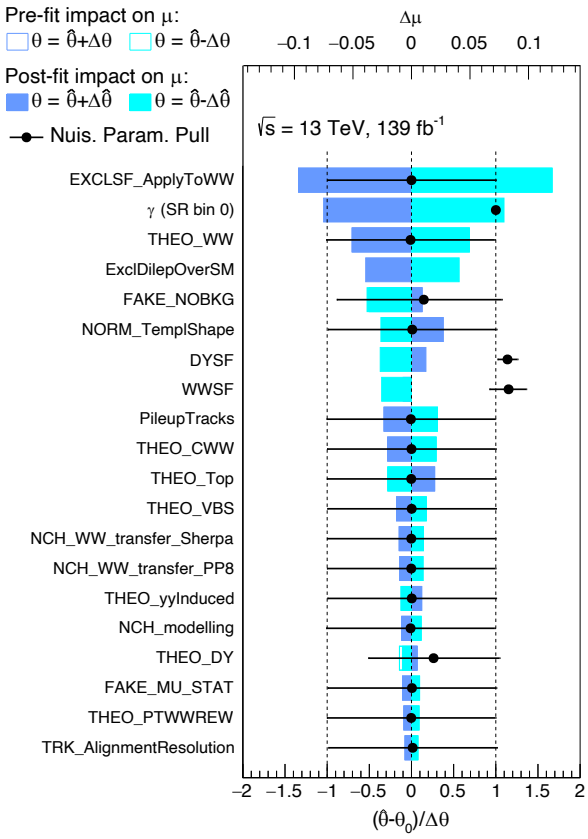


Figure 13.2: Parameters with the largest impact on the uncertainty of μ in the fit to the Asimov dataset. The blue boxes and top axis quantify the impact on the μ measurement, and the black points and bottom axis indicate the fitted values and their own uncertainties.

First, the normalization factors for the backgrounds, $\beta_{\gamma\gamma\to\ell\ell}$, $\beta_{q\bar{q}\to WW}$, and β_{DY} come out as:

$$\begin{aligned} \beta_{\gamma\gamma\to\ell\ell} = S_{\text{Excl}} &= 3.59^{+0.15}_{-0.15} \\ \beta_{q\bar{q}\to WW} &= 1.21^{+0.19}_{-0.23} \\ \beta_{DY} &= 1.16^{+0.10}_{-0.12}. \end{aligned} \tag{13.6}$$

The normalization factor for $\gamma\gamma \to \ell^+\ell^-$ is the S_{Excl} scale factor, which is multiplied by the elastic $\gamma\gamma \to W^+W^-$ and elastic $\gamma\gamma \to \ell^+\ell^-$ process yields in the signal region and control regions with $n_{\text{trk}} = 0$.

The majority of the x_f values in the final fit have a value less than ± 0.05, as shown in Figure 13.3. However, the x_f associated with FAKE_NOBKG[4] had a value of 0.35. Because this systematic uncertainty was so large, at around 100% in most regions, this mod-

[4]The systematic associated with the subtraction of events with two genuine leptons in the the fake factor derivation.

ifies the fakes estimate by around 35%. The single largest x_f parameter, which was fit at 0.61, was THEO_DY[5], though that systematic uncertainty was only about 2.0% for regions with $n_{\text{trk}} = 0$. Another relatively large x_f parameter was that associated with fake tracks (TRK_FakeRate). It was fit at -0.14, and while that systematic was sub-percent level for photon-induced processes, it was about 5% for QCD-induced processes in the signal and control regions.

Similar to Figure 13.2, Figure 13.4 shows the parameters with the greatest impact on the uncertainty of μ. The ranking is similar to that from the Asimov fit, though now the "NORM_TemplShape" uncertainty, which is the systematic uncertainty associated with the used of events with $n_{\text{trk}} = 5$ in the derivation of S_{Excl}, is in the top 5 most influential parameters.

The most important results from the fit were the μ parameter associated with the $\gamma\gamma \to W^+W^-$ signal strength and the statistical significance of the $\gamma\gamma \to W^+W^-$ yield. The μ parameter comes out to

$$\mu = 1.33^{+0.14}_{-0.14}(\text{stat.})^{+0.22}_{-0.17}(\text{syst.}). \tag{13.7}$$

That is to say that, within the fit, the cross section of the $\gamma\gamma \to W^+W^-$ process is 1.33 times higher than that predicted by the HERWIG 7 sample prediction after scaling by S_{Excl}. Likelihood scans of the μ parameter are given in Figure 13.5. The minimum of the solid black curve, which corresponds to the full final fit, is at a value of 1.33. It can be seen that this value is less than 2 standard deviations away from the SM prediction of $\mu = 1$. The statistical significance of the $\gamma\gamma \to W^+W^-$ observation in the fit is 8.4 σ, which is more than the 5 σ significance needed to claim an observation.

The post-fit yields of the different processes in the signal region and CR1, CR2, and CR3 are given in Table 13.1. This table can be compared with Table 11.2, though the errors here include systematic uncertainties. In the signal region and CR1, the yields of the photon-induced processes come from the elastic-only prediction scaled by S_{Excl}. In CR2 and CR3, the photon-induced processes come from the elastic and dissociative samples. We can see from the table that our estimation for $N_{\text{bkgd,SR}}$ is 132 total events.

The data yields and post-fit yields for the different processes can be nicely visualized in Figures 13.6, 13.7, 13.8, and 13.9. Figures 13.6 and 13.7 show the signal region and CR1, CR2, and CR3 as functions of $p_{\text{T},\ell\ell}$. Figure 13.8, shows the signal region and CR2 as a function of n_{trk} within the ± 1 mm exclusivity window. Figure 13.9 shows the signal region only, where events are plotted as a function of $m_{\ell\ell}$. Overall, the shapes of processes match the data well when plotted against all variables.

13.3 Cross-Section Measurement

Beyond observing the $\gamma\gamma \to W^+W^-$ process, we would like to measure its cross-section in this analysis. In particular, here we would like to measure the "fiducial" cross-section,

[5]The systematic associated with the DY yield after the application of the UE correction, as in Section 12.2.6

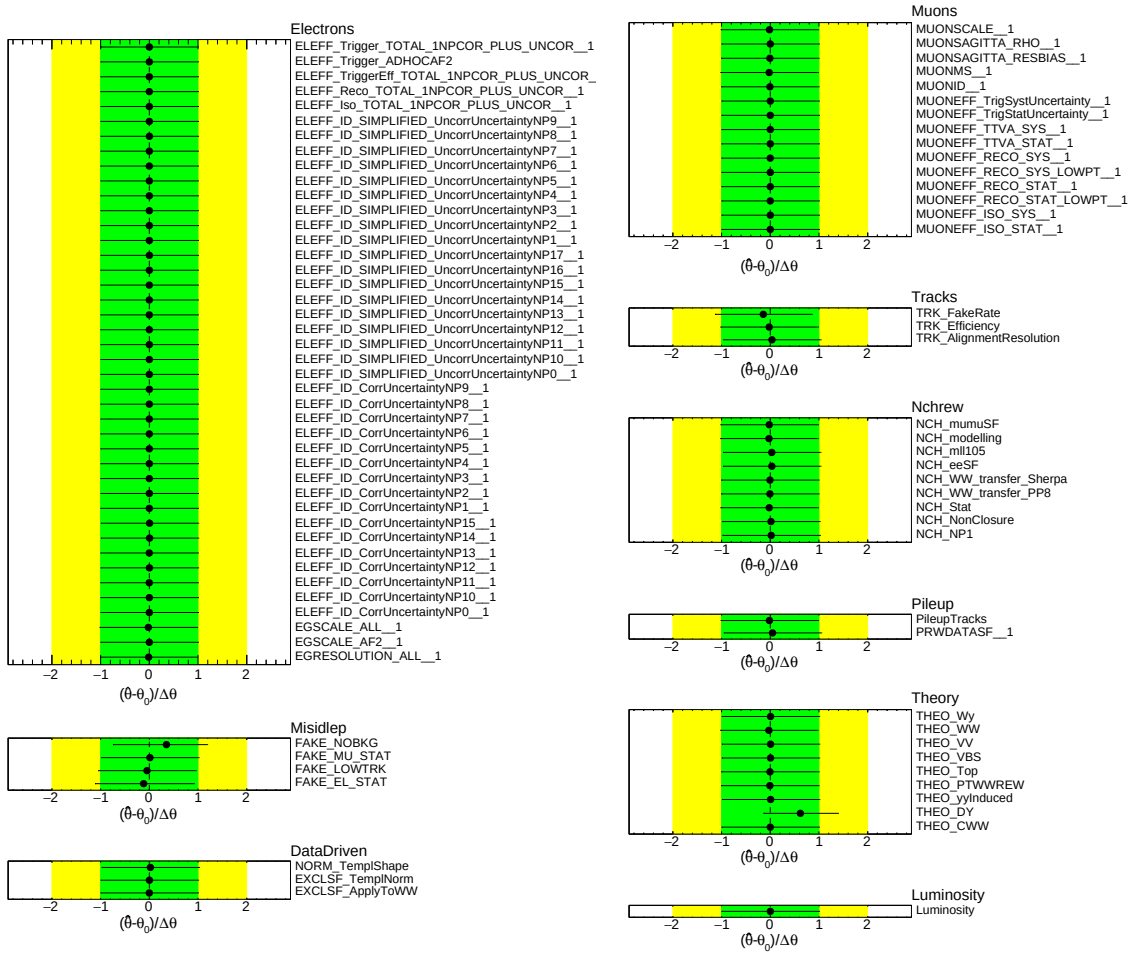


Figure 13.3: Fitted values and uncertainties of the x_f parameters in Eq. 13.4 for the final fit, which includes the SR and CRs. Each x_f value is associated with a systematic uncertainty, and the majority of the values are within ± 0.05 of their nominal value of 1. Similar to the case described in the caption of Figure 13.1, the uncertainties associated with the largest x_f values are FAKE_NOBKG and THEO_DY, though THEO_DY has increased a little. The x_f associated with "TRK_FakeRate", which is the systematic associated with the fake rate for tracks, has increased in magnitude as well.

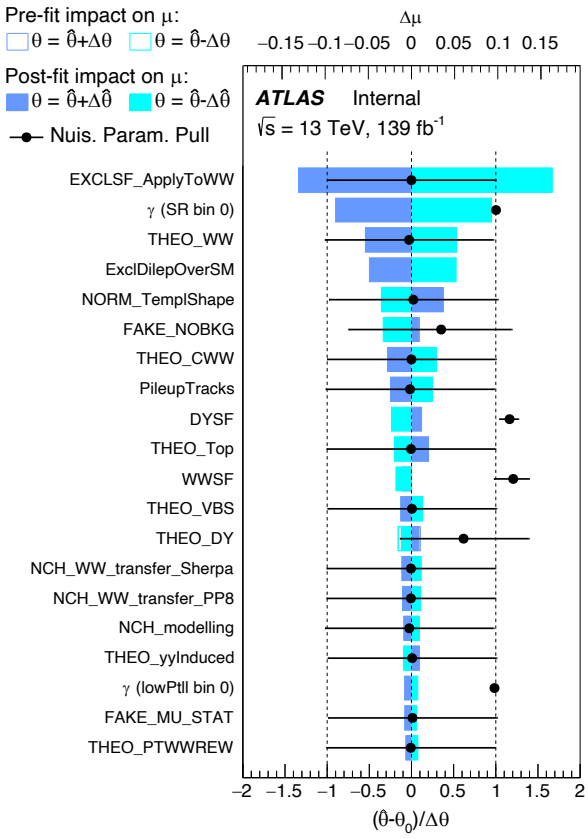


Figure 13.4: Parameters with the largest impact on the uncertainty of μ in the final fit, which included SR and CR data. The blue boxes and top axis quantify the impact on the μ measurement, and the black points and bottom axis indicate the fitted values and their own uncertainties.

which is the cross-section with constraints on the final state phase space. For example, we *do not* attempt to find $\gamma\gamma \to W^+W^-$ events with leading lepton p_{T} below 27 GeV. We are only able to directly measure the cross-section for events with final states that satisfy the conditions listed in Table 13.2. These cuts should be reminiscent of the lepton selections and preselections discussed in Chapter 6. Events where one or both of the W bosons decays to a τ lepton are also excluded. It is possible to extrapolate the fiducial cross-section to a total cross-section by considering the ratio of total simulated events to those that pass the fiducial cuts. However, this is not necessarily desirable, as we do not consider data outside of the fiducial region; it is possible that our simulation does not model the additional phase space well.

Ultimately, the fiducial cross-section prediction comes from the Equation 6.2, written

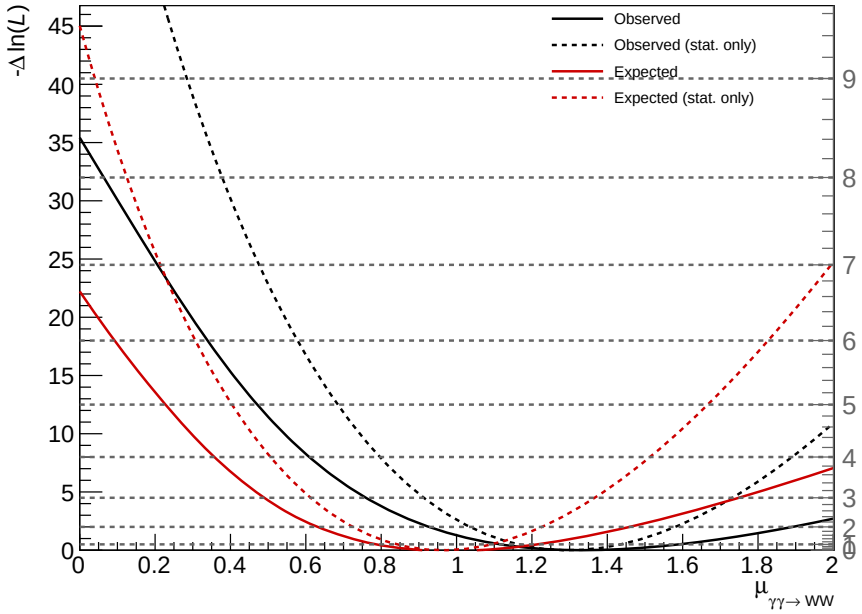


Figure 13.5: Likelihood curves for the fit parameter μ, which is used to scale the signal yield relative to the Standard Model prediction. The red curves are associated with the fit to the Asimov dataset, and the black curves are associated with the full final fit that uses unblinded SR data.

again here:

$$\sigma_{\text{fid.}} = \frac{N_{\text{data,SR}} - N_{\text{bkgd,SR}}}{\mathcal{L} \cdot \varepsilon_S}. \tag{13.8}$$

We have 307 data events in the signal region and 132 predicted background events. The luminosity was 139 fb^{-1}. The last unknown is ε_S, which is the expected signal selection efficiency; it can be broken down as:

$$\varepsilon_S = \frac{N_{\text{reco.}}}{N_{\text{truth}}} \cdot \frac{\epsilon_{\text{data}}}{\epsilon_{\text{sim.}}}. \tag{13.9}$$

In this equation, $N_{\text{reco.}}$ and N_{truth} are the number of simulated events that pass the fiducial cuts after reconstruction and the number of simulated events that pass the fiducial cuts at truth-level[6], respectively. The ϵ_{data} and $\epsilon_{\text{sim.}}$ values are exclusivity efficiencies in data and simulation respectively. While the $\frac{\epsilon_{\text{data}}}{\epsilon_{\text{sim.}}}$ fraction will be close to 1, it is still needed because the $\frac{N_{\text{reco.}}}{N_{\text{truth}}}$ fraction, which comes from simulation alone, implicitly relies on a factor of $\epsilon_{\text{sim.}}$.

[6]Prior to reconstruction

Process	**SR**	**CR1**	**CR2**	**CR3**
$\gamma\gamma \to W^+W^-$	174 ± 20.3	45.4 ± 6.1	95.0 ± 19.2	24.1 ± 4.9
$\gamma\gamma \to \ell^+\ell^-$	5.6 ± 0.3	40.1 ± 1.9	5.7 ± 1.2	32.3 ± 6.8
Inclusive WW	101 ± 16.9	54.8 ± 10.3	1719 ± 264	969 ± 152
DY $\ell^+\ell^-$	4.5 ± 0.9	281 ± 35.4	106 ± 19.1	4719 ± 396
$W + \gamma$	2.9 ± 2.9	1.9 ± 2.1	46.2 ± 46.1	35.9 ± 35.9
Fakes	10.9 ± 13.8	34.2 ± 34.2	177 ± 211	461 ± 418
Others	7.1 ± 1.7	1.9 ± 0.4	311 ± 76.3	81.4 ± 36.5
Total Sim.	305 ± 17.6	459 ± 19.1	2459 ± 65.0	6323 ± 131
Data	307	449	2458	6332

Table 13.1: Yields of the signal and background processes in the signal region and three control regions described in Chapter 11. These yields incorporate the effects of the fit, including the μ, $\beta_{\gamma\gamma\to\ell\ell}$, $\beta_{q\bar{q}\to WW}$, and β_{DY} values and the nuisance parameters. The errors include systematic uncertainties. The row for "Others" includes top background and other diboson backgrounds.

Feature	Selection
Leptons	Two opposite-charge leptons, 1 e + 1 μ
Leading Lepton	$p_{\mathrm{T}} > 27$ GeV; $\lvert\eta\rvert < 2.5$
Subeading Lepton	$p_{\mathrm{T}} > 20$ GeV; $\lvert\eta\rvert < 2.5$
Dilepton mass	$m_{\ell\ell} > 20$ GeV
Dilepton p_{T}	$p_{\mathrm{T},\ell\ell} > 30$ GeV
Non-lepton charged particles	0 charged particles with $p_{\mathrm{T}} > 500$ MeV and $\lvert\eta\rvert < 2.5$

Table 13.2: Definition of the fiducial region for the cross-section measurement.

Without the requirement of $n_{\mathrm{trk}} = 0$, the $\frac{N_{\mathrm{reco.}}}{N_{\mathrm{truth}}}$ fraction in the Herwig 7 elastic $\gamma\gamma \to W^+W^-$ simulation is 74.5%. Of those events, 52.6% pass the exclusivity selection in simulation[7], giving a final $\frac{N_{\mathrm{reco.}}}{N_{\mathrm{truth}}}$ value of 39.2%. A systematic uncertainty on this value is assessed by comparing the predictions for $\frac{N_{\mathrm{reco.}}}{N_{\mathrm{truth}}}$ found in Table 13.3. The largest difference is between the elastic-only baseline and the elastic Herwig 7 with SD and DD MG5+Pythia 8 setup. This difference leads to a symmetrized uncertainty of 1%.

The $\frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{sim.}}}$ fraction is found by applying the random window method of Section 7.2 to both data and simulated samples. The value is determined to be 1.009 ± 0.0003 (stat.).

[7]See Chapter 7

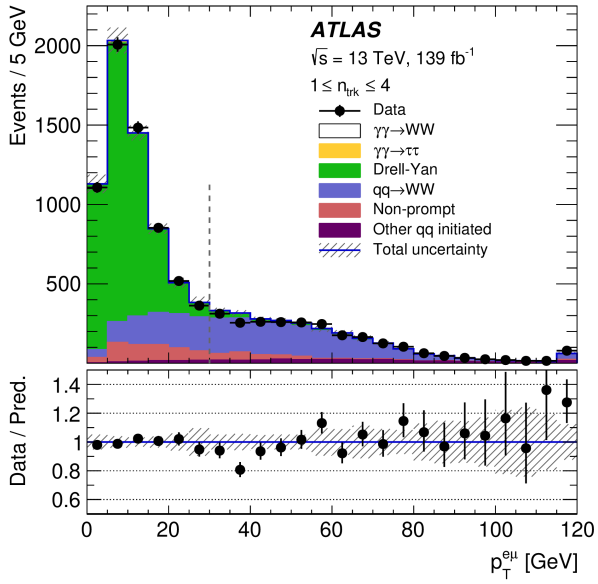


Figure 13.6: Data and post-fit yields in CR2 and CR3 [187]. CR2 is all events with $p_{\mathrm{T},\ell\ell} > 30$ GeV, and CR3 are those events with $p_{\mathrm{T},\ell\ell} < 30$ GeV. The dominance of Inclusive WW in CR2 and of Drell-Yan production in CR3 is evident.

$\gamma\gamma \to W^+W^-$ Generator	$\frac{N_{\mathrm{reco.}}}{N_{\mathrm{truth}}}$ (%)
Elastic Herwig 7 only	39.2 ± 0.3 (stat.)
Elastic Herwig 7 with SD and DD MG5+Pythia 8	39.6 ± 0.4 (stat.)
Elastic MG5 with SD and DD MG5+Pythia 8	38.9 ± 0.3 (stat.)

Table 13.3: Values of $\frac{N_{\mathrm{reco.}}}{N_{\mathrm{truth}}}$ fraction for different signal generator combinations.

Given the two fractions defined above, we have $\varepsilon_S = 0.396 \pm 0.004$. Additionally, 3% of SR events are expected to have had at least one W Boson that decayed into a τ, which then decayed into an e or μ. This contribution is subtracted.

After the τ-contribution's subtraction, we combine the data yield, the central value for the predicted $N_{\mathrm{bkgd,SR}}$, the luminosity, and ε_S to find a final fiducial cross section 3.13 fb. To account for systematic uncertainties, the background predictions are made with the various nuisance parameters shifted up and down by one standard deviation. Statistical uncertainty primarily comes from the yield of events in data. All uncertainties are symmetrized. In the end, we find

$$\sigma_{\mathrm{meas.}} = 3.13 \pm 0.31\ (\mathrm{stat.}) \pm 0.28\ (\mathrm{syst.})\ \mathrm{fb} \tag{13.10}$$

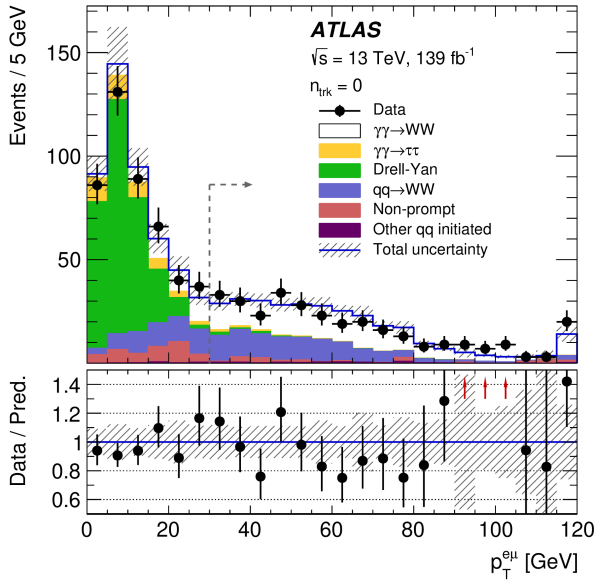


Figure 13.7: Data and post-fit yields in the signal region and CR1 [187]. The signal region is all events with $p_{\mathrm{T},\ell\ell} > 30$ GeV, and CR1 are those events with $p_{\mathrm{T},\ell\ell} < 30$ GeV. The dominance of $\gamma\gamma \to W^+W^-$ in the signal region is clear. In CR1, the most prominent contribution is from Drell-Yan production, though a contribution from $\gamma\gamma \to \ell^+\ell^-$ is also important.

as the measured fiducial cross-section for the process $pp(\gamma\gamma) \to p^{(*)}W^+W^-p^{(*)}$, where the W^+W^- system decays to $e^{\pm}\mu^{\mp}\nu\bar{\nu}$. There is a 8.9% uncertainty on this measurement from systematics. Figure 13.10 shows an explicit breakdown of the sources of systematic uncertainty on the cross section measurement.

This cross-section can be compared with the expectation from simulation. The expected fiducial cross-section of the elastic Herwig 7 simulation is 0.655 fb. When the S_{Excl} scale factor is multiplied onto this value, the total expected cross-section is

$$\sigma_{\mathrm{H7}} = 2.35 \pm 0.27 \text{ fb}, \tag{13.11}$$

which accounts for the uncertainty on S_{Excl} and the uncertainty on the applicability of S_{Excl} to the $\gamma\gamma \to W^+W^-$ process. As expected, there is a factor of 1.33 difference between the measured fiducial cross-section and the expected one from Herwig. This is in agreement with the measured value of μ discussed in the previous section.

The measured value can also be compared to a prediction from MG5. In this case, the predicted cross-section does not rely on S_{Excl}, rather on the exclusive, SD, and DD MG5[8] samples of Table 5.1. The prediction for the fiducial cross-section in this case comes to 4.3 ± 1.0 (scale) $\pm$ 0.1 (PDF) fb. The scale uncertainty comes from varying the choice of

[8] +Pythia 8 in the SD and DD cases

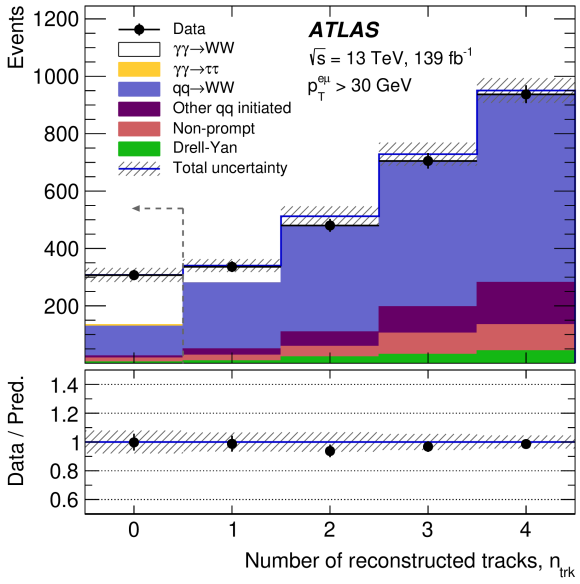


Figure 13.8: Data and post-fit yields in the signal region and CR2 [187]. All events in this plot have $p_{\mathrm{T},\ell\ell} > 30$ GeV, so the signal region is events with $n_{\mathrm{trk}} = 0$, and CR1 are those events with $1 \leq n_{\mathrm{trk}} \leq 4$. The signal region has about a 1.3:1 signal to background ratio.

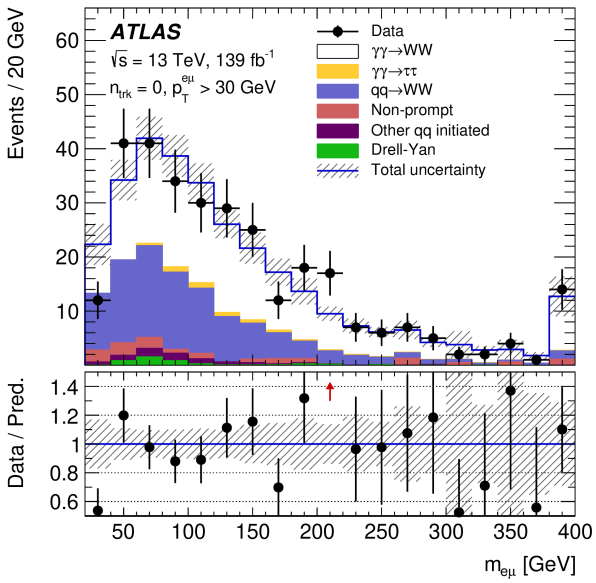


Figure 13.9: Data and post-fit yields in the signal region as a function of the dilepton mass, $m_{\ell\ell}$ [187]. The right-most bin includes overflows.

Source of uncertainty	Impact [% of the fitted cross section]
Experimental	
Track reconstruction	1.1
Electron energy scale and resolution, and efficiency	0.4
Muon momentum scale and resolution, and efficiency	0.5
Misidentified leptons, systematic	1.5
Misidentified leptons, statistical	5.9
Other background, statistical	3.2
Modelling	
Pile-up modelling	1.1
Underlying-event modelling	1.4
Signal modelling	2.1
WW modelling	4.0
Other background modelling	1.7
Luminosity	1.7
Total	8.9

Figure 13.10: The impact of the most important sources of systematic uncertainty on the cross-section measurement[187]. These impacts are determined by varying the source of the systematic up and down by one standard deviation; correlations are not taken into account. The effect of the variation is symmetrized here.

factorization scale for the dissociation. The MG5 samples do not rely on S_{Excl}, and they do not include the survival effects discussed in Section 2.2.4. As discussed in that section, two different predictions for the S^2 factor in the $\gamma\gamma \to W^+W^-$ case are 0.65 and 0.82. If these factors are applied to the MG5 prediction, then the final predictions are 2.8 ± 0.7 fb and 3.5 ± 0.9 fb. Both of these are consistent with the measurement of 3.13 fb within uncertainties.

Chapter 14

Part II Conclusions and Outlook

Part II of this thesis has laid out a complete analysis for the observation and measurement of the $\gamma\gamma \to W^+W^-$ process in $\sqrt{s} = 13$ TeV proton-proton collisions the LHC. This analysis probes a unique region of phase space at the LHC, where the exclusivity of an event is determined by a lack of reconstructed tracks. This led to the application of various corrections to the simulation, as desribed in Chapters 7 – 10. Having performed these corrections, analysis regions were defined and systematic uncertainties assessed. A fit was performed to maximize the Poisson likelihood of the observed data yields in the signal region and control regions, which simultaneously accounted for mismodelling of the normalization of the major backgrounds and extracted the signal strength relative to the Standard Model expectation.

In the 139 fb^{-1} of data collected in Run 2 of the LHC, the $pp(\gamma\gamma) \to p^{(*)}W^+W^-p^{(*)}$: $W^+W^- \to e^\pm\mu^\mp\nu\bar{\nu}$ process was observed with a statistical significance of 8.4σ, which is its first ever observation. The cross-section was measured to be

$$\sigma_{\text{meas.}} = 3.13 \pm 0.31 \text{ (stat.)} \pm 0.28 \text{ (syst.) fb}, \tag{14.1}$$

which constitutes a signal strength of

$$\mu = 1.33^{+0.14}_{-0.14}(\text{stat.})^{+0.22}_{-0.17}(\text{stat.}) \tag{14.2}$$

relative to the Standard Model prediction from Herwig 7. An event display for an event in the signal region is shown in Figure 14.1. The presence of an electron-muon pair with high $p_{\text{T},\ell\ell}$ and a lack of tracks near their vertex can be easily seen in this figure.

However, this is not the end of the story. While this result suggests a cross-section that agrees with the Standard Model within uncertainties, we can use that information to constrain Beyond the Standard Model (BSM) physics that would affect the processes contributing to our measurement. Such constraints are best placed by extracting differential cross-section measurements, which is to say, the cross-section as a function of kinematic variables, such as $p_{\text{T},\ell\ell}$ or $m_{\ell\ell}$. With an observation of the $\gamma\gamma \to W^+W^-$ process in hand, what we now want is to place the most stringent possible constraints on BSM physics.

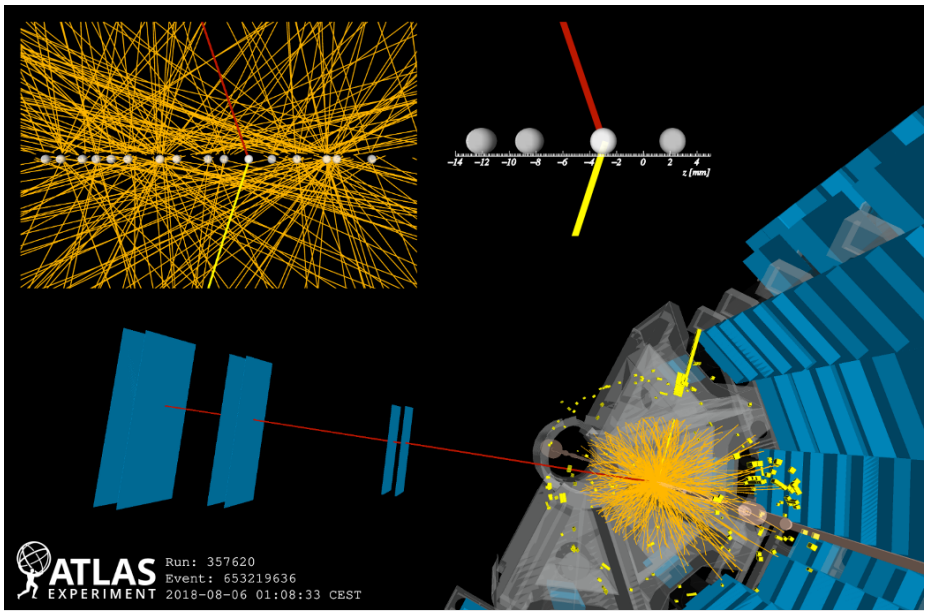


Figure 14.1: Sample event display for a single event in the SR for the $\gamma\gamma \to W^+W^-$ analysis [187]. The muon and electron with high-$p_{\mathrm{T},\ell\ell}$ are shown in red and yellow, respectively. The orange lines represent tracks. The insert shows track density in the z-direction; no tracks point near the muon-electron vertex.

While an 8.4σ observation is significant, we can do better. We should not forget the fact that the most important signal-vs.-background discriminant in our arsenal is the presence of reconstructed tracks, nor should we forget that there is an artificially placed lower-bound on our reconstructed track p_{T}. Lowering this threshold in ATLAS is not trivial, but it will lead to a large improvement in background rejection, improving our measurement. This pursuit is the subject of Part III.

Part III

Exploring $\gamma\gamma \to W^+W^-$ with low-p_{T} tracks

Chapter 15

Motivation for Low-p_{T} Tracking

The reconstruction of charged particles ("tracking") using the ATLAS detector was discussed in detail in Section 4.1. That section covered the algorithms used for tracking, the quality cuts made to select tracks, and the resolution of different track parameters. These tracks played a vital role in the observation of the $\gamma\gamma \to W^+W^-$: tracks are the key feature used to distinguish photon-induced processes from QCD-induced processes at a proton-proton collider. In Part II, our signal region was defined such that there were no tight[1] tracks with $p_{\mathrm{T}} > 500$ MeV and $d_0 < 1$ mm within a longitudinal distance of ± 1 mm from the lepton vertex[2]. Aggregated over all of LHC Run 2, about 50% of signal events in the fiducial region passed this cut, with the majority of failures caused by a pileup track falling within the exclusivity window. For the dominant background, $q\bar{q} \to W^+W^-$, only about 0.1% of events in the fiducial region passed this selection. Despite the large differences in the probability for $\gamma\gamma \to W^+W^-$ and $q\bar{q} \to W^+W^-$ events to pass the exclusivity selection, the post-fit signal-to-background ratio in Section 13.2.2 was 1.32:1. While the yields discussed in Chapter 13 were more than enough to establish a significant observation of $\gamma\gamma \to W^+W^-$, it is possible to perform an even more powerful version of the analysis.

15.1 Low-p_{T} Tracking for the $\gamma\gamma \to W^+W^-$ Analysis

In order to improve the analysis, a few different options can be considered:

- Increase the yields of both signal and background. The statistical significance would improve if the yields were increased, even if the 1.32:1 signal-to-background yield was maintained.
- Improve the signal acceptance efficiency.
- Improve the background rejection rate.

[1] see Section 4.1
[2] see Section 6.2.3

The first of these options is possible to some extent. For example, the fiducial region can be changed to include lower-p_{T} leptons, which would increase the yield of both signal and background[3]. The other way to effectively implement the first option is to increase the amount of data. Unfortunately, this is not possible at the moment. LHC data-taking will not resume until this thesis is already published, but this analysis *can* be revisited after Run 3 is completed.

The second point could be implemented with a relatively simple change: if a smaller exclusivity window is used, fewer signal events will be rejected due to a pileup track in the window. Unfortunately, this would have the side effect of increasing the background acceptance. As described in Appendix A, the ± 1 mm window size optimizes the statistical significance of the analysis. Almost any change that increases the signal acceptance will result in a more serious increase in background rates, making the second bullet an ineffective way to improve the analysis.

At face value, the last point seems to run into the same issues as the second. We could increase the background rejection by *widening* our exclusivity window, but that would lead to lower signal acceptance. Indeed, given the information available to the analysis detailed in Part II, there are not many realistic possible improvements.

However, we should remember that the lower-bound on track p_{T} is not set due to the physical limits of the ATLAS detector, rather due to reconstruction time and memory storage constraints[4]. Tracks are rejected if they have p_{T} below 500 MeV, but if we could lower this threshold, we could reject a higher rate of background. Charged particles created in an Underlying Event (UE) are actually most likely to have a p_{T} below 500 MeV, as shown in Figure 15.1[5] [87]. If charged particle tracking were performed down to about 100 MeV, the number of tracks would effectively double, which could dramatically reduce the 0.1% background acceptance rate. At the same time, more pileup tracks *will* be reconstructed, slightly lowering the signal acceptance rate, but the signal's lack of UE tracks will lead to a less steep loss in efficiency.

15.2 Limits of Tracking in ATLAS

Before addressing any algorithmic attempt at low-p_{T} track reconstruction, it is worth quickly noting the physical limits of the ATLAS when it comes to tracking. The relationship between a charged particle's p_{T}, its bending radius, and the magnetic field is approximately

$$p_{\mathrm{T}} = 0.3 * B * R, \tag{15.1}$$

[3]In the analysis in Part II, the leading lepton p_{T} is set based on the single-lepton trigger thresholds, and the subleading lepton p_{T} was set to reduce contributions from fake leptons.

[4]Low-p_{T} tracking also suffers from combinatorial issues discussed in Section 15.2, which exacerbate the reconstruction time problems relative to high-p_{T} tracking.

[5]The data in this figure was reconstructed with the low-pileup tracking algorithm described in Section 15.3.

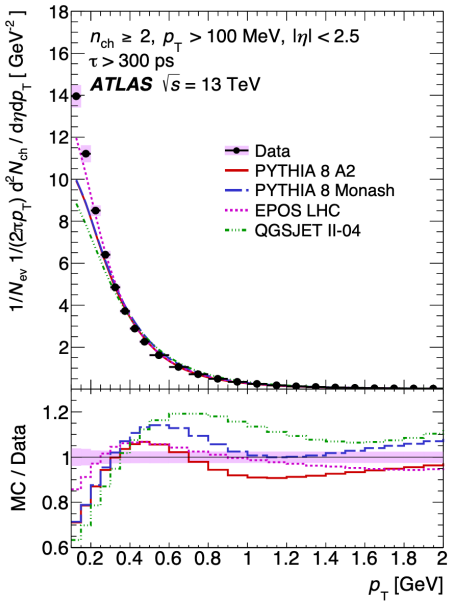


Figure 15.1: Charged particle multiplicity vs. p_{T} for data taken at very low-pileup [87]. Tracks in these events are most likely to have p_{T} below 500 MeV, which is the standard lower limit for tracking.

where the p_{T} is in units of GeV, the magnetic field, B, is in units of T, and the bending radius, R, is in units of m. The ATLAS Inner Detector is immersed in a field of 2T, with the outermost silicon layer having a radius of 0.5m. If a particle created at the interaction point has a bending radius of 0.25m, it will just reach that outermost layer[6]. Therefore, the minimum p_{T} particle that will reach the outermost silicon layer has a p_{T} of 150 MeV. To reach the outermost pixel layer, at a radius of 0.12m, the particle would have to have a p_{T} of at least 40 MeV. Such particles could be reconstructed with a pixel-layer-only tracking algorithm, but with only 4 hits expected per track, the efficiency would be relatively low and the fake rate relatively high.

As mentioned in the previous section, the main reasons for the current tracking p_{T} limit of 500 MeV are to reduce storage space requirements and to decrease per-event reconstruction time. Storage space issues are unavoidable unless a reduced set of information is desired for lower-p_{T} tracks. However, we want to retain as much tracking information as possible, so we cannot reduce storage space on a per-event level[7].

The processing time problem primarily occurs at the seeding stage of the tracking algorithm (see Section 4.1). As illustrated in Figure 15.2, inner detector hits associated with

[6]The diameter of the track's helix is 0.5m in this case.

[7]A strategy to decrease the overall storage required for events with reconstructed low-p_{T} tracks is given in Section 16.1

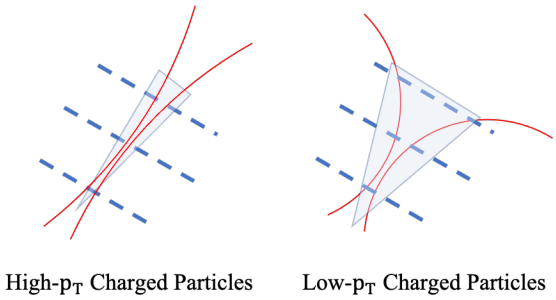


Figure 15.2: Lower-p_{T} charged particles curve more in ATLAS's magnetic field than high-p_{T} particles. Because of this, hits on different layers associated with a single particle can be further apart. To make a complete set of seeds, a much larger number of hit combinations must be considered.

lower-p_{T} particles will be spatially further apart than hits associated with high-p_{T} particles. When forming seeds for high-p_{T} tracks, a relatively small set of connections with hits on neighboring layers is possible. At low-p_{T} this set of connections is much larger. We want to make a complete set of seeds, so we must form as many connections between hits as physically reasonable. This problem is worsened by the effect of "multiple scattering" [2]. This effect is illustrated in Figure 15.3. As charged particles pass through matter, their trajectory can be deflected by Coulomb interactions with atomic nuclei, such that the initial incidence angle on the material is different from the angle at which it departs the material. This effect is more important for lower-p_{T} charged particles, resulting in the need for an even wider cone to be considered than shown in Figure 15.2. Because of multiple scattering, seeds for low-p_{T} tracks are not expected to be as high-quality as those for high-p_{T} tracks. The number of seeds produced increases faster than linearly as the minimum tracking p_{T} is lowered.

All of these problems become even more relevant as pileup increases. The number of hits in the Inner Detector increases approximately linearly with pileup, leading to a significant combinatorics problem.

To improve the processing slow-down associated with track seeding, an algorithmic speed up can be achieved by both reducing the number of hit connections that must be considered and the number of seeds that will be formed in total. The development of such an algorithm is detailed in Chapter 16.

15.3 Existing Low-p_{T} Tracking Algorithms

It is worth noting here that tracking below a p_{T} of 500 MeV *has* been performed in ATLAS before. The most trivial example of this is track reconstruction in data prior to 2017. In Run 1 of the LHC and for the first years of Run 2, the average pileup was lower than that in 2017 and 2018. Lower pileup leads to fewer charged particles, meaning smaller per-event readout

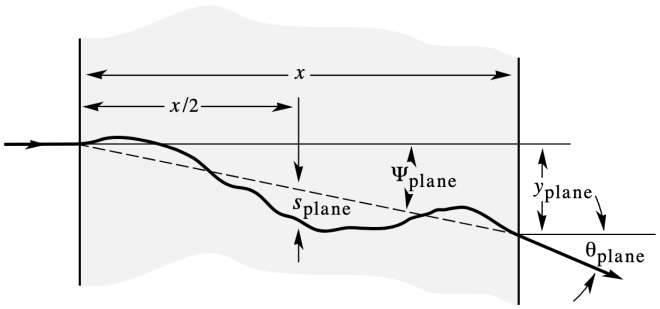


Figure 15.3: Illustration of multiple scattering of a charged particle as it passes through matter [2]. This effect is relatively stronger for lower-p_{T} charged particles.

sizes and shorter processing time. Because of this, the lower-bound on nominal tracking in the past was actually 400 MeV. However, the algorithm was exactly the same.

Another example is a tracking setup that was designed specifically for very low-pileup conditions [87]. In 2015, the LHC delivered 151 $\mu\mathrm{b}^{-1}$ of proton-proton data with an average of 0.005 interactions per bunch crossing. In this dataset, there is essentially a maximum of one proton-proton interaction per event. These events are exceptionally sparse, and the dataset itself is small, so processing time and storage space are not a major obstacles. A specialty tracking algorithm was applied to this dataset seeking to reconstruct charged particles down to a p_{T} of 100 MeV. This algorithm proceeds with essentially the same steps outlines in Section 4.1. However, the specific seeding algorithm was changed slightly to optimize for low-p_{T} seed finding, and the track fitting scheme was also slightly modified. Similarly, the minimum number of hits needed to constitute a track was reduced to account for the fact that the lowest-p_{T} tracks would not traverse all layers. The tracking efficiency of this algorithm vs. p_{T} is shown in Figure 15.4. The efficiency is relatively stable and above 70% for particles with p_{T} over about 250 MeV, with a steep dropoff at lower-p_{T}.

This algorithm was primarily designed to study the charged particle distributions of unbiased inelastic proton-proton collisions[8], and to help improve the modelling of such interactions[9].

However, we do not want to apply this low-pileup algorithm directly for the $\gamma\gamma \to W^+W^-$ analysis. For one, it would greatly increase the reconstruction time and storage space required for our data. It has also been algorithmically tuned for low-pileup environments. But perhaps more importantly, the tracking performance for high-p_{T} tracks, particularly those near 500 MeV will be slightly different than the nominal algorithm used in the analysis discussed in Part II. For the sake of consistency, we want to exactly reproduce the tracks used in that version of the analysis[10].

[8] Which comprise the majority of pileup interactions in nominal running.

[9] As seen in Figure 15.1, the simulations used in Ref. [87] did not match the data well.

[10] Changing tracking algorithms can also affect lepton reconstruction, as both electron and muons rely on

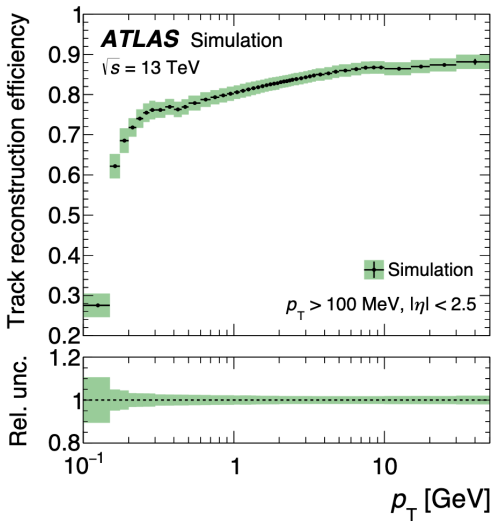


Figure 15.4: Tracking efficiency vs. p_{T} for charged particles when using the tracking algorithm implemented in very low-pileup environments [87].

15.3.1 Quantified Impact of Low-p_{T} Tracks

That being said, we *can* use this algorithm to inform our expectations of performance in a version of the $\gamma\gamma \to W^+W^-$ analysis with low-p_{T} tracking. A private processing of the hit-level simulation for $q\bar{q} \to W^+W^-$ events and elastic $\gamma\gamma \to W^+W^-$ events was performed with fully simulated pileup. Then the yields of these samples were considered for different track p_{T} cutoffs. That is to say, while tracks were reconstructed down to 100 MeV, we might only consider tracks with $p_{\mathrm{T}} > 300$ MeV. In the case that we only consider tracks with $p_{\mathrm{T}} > 500$ MeV, the performance should be similar to the first round of analysis. In this study, the same event selections detailed in Section 6.2.3 were used, and a ± 1 mm window was used for tracks. The results of this study are shown in Figure 15.5. The $q\bar{q} \to W^+W^-$ yields are shown both with and without the UE correction derived in Chapter 8. This correction is applied based on truth-level particles with $p_{\mathrm{T}} > 500$ MeV, so it is not dependent on the track reconstruction. However, it is possible that an additional correction will be needed to account for mismodelling of lower-momentum UE particles. Similarly, the elastic-only $\gamma\gamma \to W^+W^-$ yield is shown both with and with the S_{Excl} scale factor that accounts for dissociative events. It is not exactly clear that the same scale factor is applicable in the presence of low-p_{T} tracks; dissociating protons may actually create tracks in this case, which would reduce the scale factor by some unknown amount. It can also be seen in this image that the expected yields in the minimum $p_{\mathrm{T}} = 500$ MeV bin do not match those in Table 13.1. This

Inner Detector tracks. We would not be able to use the lepton and trigger scale factors recommended by ATLAS in this case.

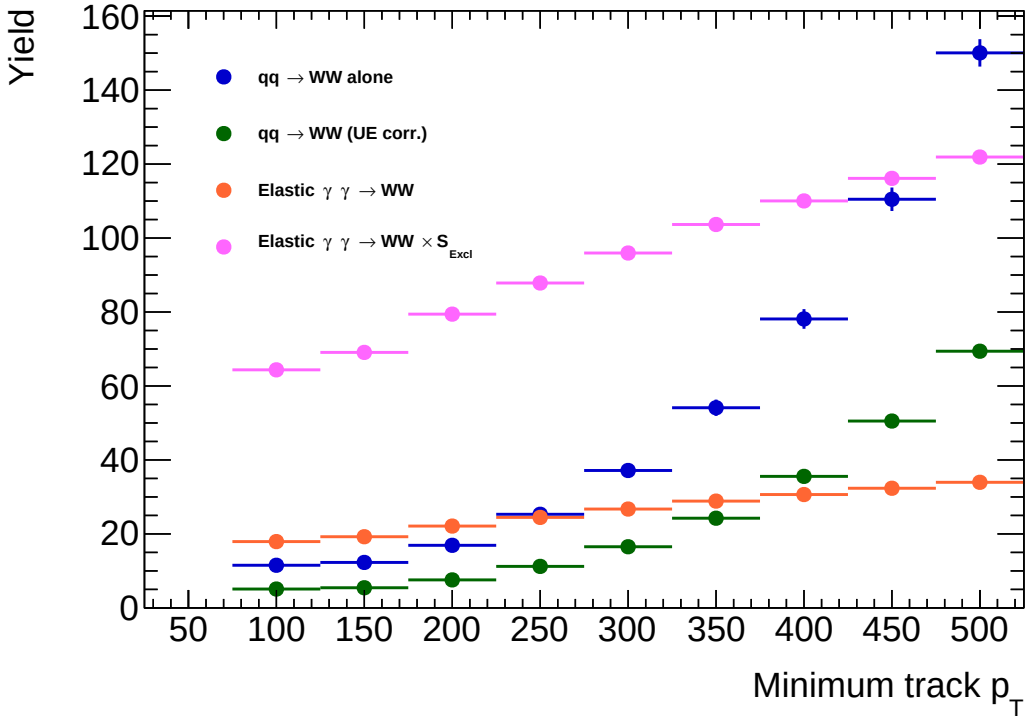


Figure 15.5: The yields of the $q\bar{q} \to W^+W^-$ and elastic $\gamma\gamma \to W^+W^-$ samples when different minimum-p_{T} cutoffs are applied in event selection. The samples were reconstructed using the low-pileup tracking algorithm. The events are selected using the selections detailed in Section 6.2.3, together with a ± 1 mm exclusivity window. The blue dots represent the $q\bar{q} \to W^+W^-$ yield without applying the underlying event correction detailed in Chapter 8, while the green dots do have that correction applied. Similarly, the orange dots represent the elastic-only $\gamma\gamma \to W^+W^-$ yield, whereas the magenta dots have the $S_{\mathrm{Excl}} = 3.59$ scale factor applied to account for dissociative events.

is mostly due to two facts. First, the tracking algorithm has been changed here, even above a p_{T} of 500 MeV, so the rates of signal and background passing the exclusivity selection will change. Second, the $q\bar{q} \to W^+W^-$ sample does not include gluon- or VBS-induced WW production, and the 1.21 $\beta_{q\bar{q}\to WW}$ factor found in the fit is not applied. The 1.33 signal strength factor for the signal is also not applied.

Regardless, this figure illustrates the key expected improvement. The background experiences an exponential-shaped fall-off as the minimum considered track p_{T} is lowered, whereas the signal experiences a relatively gentle fall. The signal's drop is due to newly reconstructed pileup tracks, whereas the driving factor behind the background drop-off is newly reconstructed UE tracks. The $q\bar{q} \to W^+W^-$ yield with the UE correction drops by

93% as the minimum p_{T} is lowered from 500 MeV, where the expected yield is 70 events, to 100 MeV, where the expected yield is 5 events. For that same change, the signal yield drops by 47%, decreasing from 122 expected events (with S_{Excl} applied) to 65 events. These values will change as the exclusivity selection is re-evaluated[11] and as the various simulation corrections are updated.

The expected impact of adding low-p_{T} tracks to the $\gamma\gamma \to W^+W^-$ analysis is illustrated in Figure 15.6. Here, the expected signal-to-background ratio and statistical significance of the analysis are plotted against the minimum track p_{T} considered for the exclusivity selection. A custom significance metric was used, with the assumed integrated luminosity of 139 fb^{-1}:

$$\sigma = \frac{3.59 * s}{\sqrt{3.59 * s + 1.3 * b + (0.2 * 1.3 * b)^2}}. \tag{15.2}$$

In this equation, s is the expected yield of the elastic-only $\gamma\gamma \to W^+W^-$ samples, and b is the expected yield of the $q\bar{q} \to W^+W^-$ sample with the UE correction applied. The factor of 3.59 is the same S_{Excl} scale factor which accounts for dissociative signal events. The factor of 1.3 applied to the $q\bar{q} \to W^+W^-$ yield is designed to account for additional background processes. In Table 13.1, it can be seen that there is about a 1:1.3 ratio of inclusive WW to other backgrounds. The last term in the denominator is designed to account for uncertainties on the background, here being conservatively assessed at 20%. The overall form of the significance metric takes the form of the signal yield divided by the Poisson uncertainty of the signal region yield.

The signal-to-background ratio increases as the minimum-p_{T} is lowered down to about 150 MeV before plateauing. When considering the significance, there is a balancing act which simultaneously rewards a higher signal yield and a higher signal-to-background ratio. Because the signal yield decreases with lower-p_{T} while the signal-to-background ratio increases, the significance peaks for a tracking cutoff of 250 MeV.

It should be noted that the metric defined in Equation 15.2 is relatively conservative. The expected significance is 5.3σ when a minimum p_{T} of 500 MeV is considered. In the version of the analysis *without* low-p_{T} tracks, the expected significance was 6.7σ based on the Standard Model alone. Because the observed signal strength was 1.33 times the Standard Model, the observed significance was 8.4σ. In light of this, the expected improvement should be compared to the 5.3σ in Figure 15.6. Based on this plot, we could expect about a 3.5σ increase in significance when applying low-p_{T} tracks.

Because we plan on developing a new tracking algorithm and re-evaluating our exclusivity selection, the numbers found here are expected to change when the analysis is performed in real data. However, as an exploratory study, these results are very promising and strongly motivate the use of low-p_{T} tracking for photon-induced analyses.

[11] See Chapters 16 and 17.

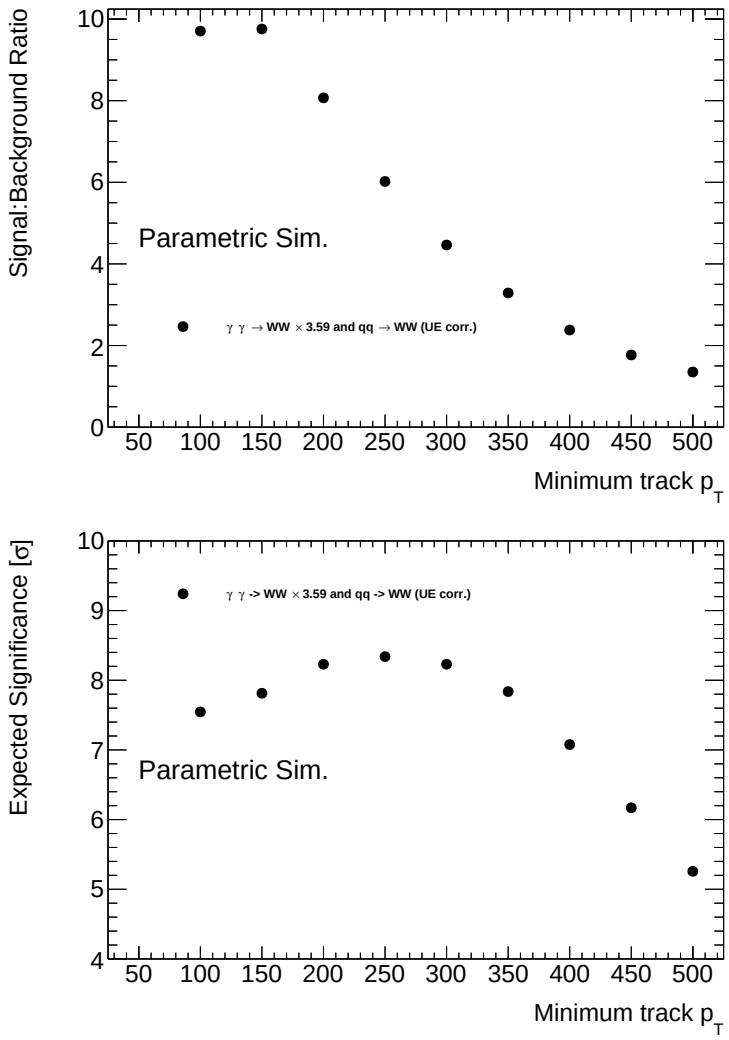


Figure 15.6: The expected signal-to-background ratio (top) and significance (bottom) of the $\gamma\gamma \to W^+W^-$ analysis when using low-p_{T} tracks. The significance metric, which is described in the text, attempts to account for missing samples and uncertainties. The yields of the $q\bar{q} \to W^+W^-$ and elastic $\gamma\gamma \to W^+W^-$ samples used here are the same as those in Figure 15.5, where the $q\bar{q} \to W^+W^-$ sample has the underlying event correction applied, and the $\gamma\gamma \to W^+W^-$ sample is scaled to account for dissociative events. The samples were reconstructed using the low-pileup tracking algorithm.

Chapter 16

Implementation of Low-p_{T} Tracking

In Chapter 15, it was explained that low-p_{T} charged particle reconstruction (tracking) represents an important possibility for improving the $\gamma\gamma \to W^+W^-$ analysis. In a relatively simple exploratory study using a pre-existing version of low-p_{T} tracking, the point of maximum expected statistical significance for the $\gamma\gamma \to W^+W^-$ process showed an 85% reduction in background compared to the nominal cutoff, while the signal yield was reduced by only 25%. This chapter illustrates a strategy for actually carrying out an improved $\gamma\gamma \to W^+W^-$ measurement with low-p_{T} tracks.

It was also explained that low-p_{T} tracking is not simple, especially in high-pileup environments. For lower-p_{T} tracks, increasingly large sets of hits must be combined to form a complete set of seeds, slowing down the tracking algorithm. A low-p_{T} tracking implementation is detailed in this chapter which seeks to alleviate these issues.

16.1 Re-Analysis Strategy

In order to use low-p_{T} tracks for a real analysis, data events must be reprocessed. Typically, analyses in ATLAS only refer to higher-level objects, such as muons, electron, and tracks, without reference to the individual detector hits that were used to reconstruct the object. However, information about the detector hits *is* stored, so it is possible to run new reconstruction algorithms on events.

It has been mentioned several times that low-p_{T} tracking will increase processing time and storage space, but it is possible to minimize the consequences of this fact by processing only a small subset of the Run 2 data. The $\gamma\gamma \to W^+W^-$ analysis team, together with the developers of ATLAS Eventindex [209], developed and tested a novel method which allows teams to reprocess small subsets of Run 2 data with specialty algorithms. Teams can specify the desired subset with a simple list of events. In particular, the $\gamma\gamma \to W^+W^-$ analysis team *knows* what events are most relevant for another round of the analysis: we want to reprocess events in the old signal region and control regions[1]. This comes to fewer than 10,000 events

[1]With low-p_{T} tracking, signal events in the signal region should stay in the signal region, and background

in data (not counting the Z-peak region or high-$m_{\ell\ell}$ region), which is significantly smaller than the billions of events in the full Run 2 dataset. By performing event picking, we simply analyze less data, which is a valid way to avoid time and storage constraints. Selecting these events for reprocessing is a major reason for wanting to preserve all the high-p_{T} tracks from Part II exactly. We want the event-picked dataset to have exactly the same information as before reprocessing with only the addition of new, low-p_{T} tracks.

Therefore, the re-analysis of $\gamma\gamma \to W^+W^-$ with low-p_{T} tracking must proceed in stages:

1. Events are analysed using only generic ATLAS objects in order to determine a small subset that is worth reprocessing with low-p_{T} tracks. No reprocessing is required for this step. This step is practically equivalent to the analysis presented in Part II, as the signal and control region events will be most relevant after reprocessing too. However, some event selections could be changed relative to Part II; for example, the subleading lepton p_{T} requirement could be decrease below 20 GeV.

2. The small subset of data picked in the preceding step is reprocessed with the low-p_{T} tracking algorithm described in Section 16.2.

No similar two-stage process is needed for simulated events. The simulated samples of Chapter 5 can be fully reprocessed as they are much smaller than the full Run 2 dataset.

Additional developments are needed for the analysis that will rely on both simulation and data:

- Selection criteria must be developed for the new low-p_{T} tracks, similar to the "tight" selection applied to generic tracks (see Section 4.1). These are explored in Chapter 17.

- The exclusivity selection must be re-optimized in the presence of low-p_{T} tracks. The old ± 1 mm window selection will no longer provide the best performance. This optimization is also presented in Chapter 17.

- The corrections of Chapters 7 – 10 will be rederived to account for potential changes due to low-p_{T} tracks. This will rely on both data and simulation, with similar methods employed as in Part II.

- The signal and control regions must be re-defined.

After these changes, the statistical analysis will resemble that of Part II.

events that were in the signal region should be rejected. Events should not migrate *into* the signal region, with one caveat: if the exclusivity selection is changed, then signal and background events *could* migrate back into the signal region. For example, if the exclusivity window's width was reduced, a signal event with a pileup interaction at the edge of the old window could appear in the newly defined signal or control regions. While this is not optimal, it should be a small effect. Furthermore, this effect can be accounted for in simulation by including a cut on whether the event would have fallen in an old region, or data events with a few additional tracks beyond the maximum of 4 in CR2 could be included in the reprocessing.

16.2 New Low-p_{T} Tracking Algorithm

Because will not use the low-pileup tracking algorithm of Section 15.3, we must develop a new tracking algorithm and explore its performance. In this algorithm, we will take two steps specifically to reduce the processing time. First, we will remove some hits from consideration when reconstructing low-p_{T} tracks. This will alleviate the combinatorics problem mentioned at the beginning of this chapter. Second, we have the option to only reconstruct tracks that point back to within a certain longitudinal distance of the lepton vertex. This will reduce the amount of time spent in the ambiguity-solving component of the tracking algorithm and significantly reduce the number of final tracks produced, leading to smaller file sizes.

A goal of this new setup is to achieve high tracking efficiency. Constraining fake rate is a smaller concern, as that aspect can be deferred to analysis after track reconstruction. To make things exactly comparable to a setup without low-p_{T} tracking, this implementation should exactly reconstruct all tracks that would be constructed in generic reconstruction.

The algorithm, which is schematically illustrated in Figure 16.1, proceeds as follows:

1. Generic tracking is performed as detailed in Section 4.1 with a minimum track p_{T} of 500 MeV. These tracks will be unchanged from the first version of the analysis.

2. The inner detector hits that were used for these tracks are removed from the space point collection.

3. After backtracking (see Section 4.1), a low-p_{T} specific round of tracking is performed. Here, the seed making algorithm is the same as that used in the generic tracking pass, though it searches for seeds within a larger cone of combinations[2]. The track fitting scheme is adjusted, with reduced hit and quality requirements. This tracking pass does *not* have an upper limit on track p_{T}. It is possible that a track with p_{T} *above* 500 MeV will be formed in this pass. All tracks formed in this stage of the algorithm are stored separately from those created in the initial, nominal tracking pass.

Low-p_{T} track information is not used in vertexing.

This algorithm differs from the low-pileup one; that algorithm sought to reconstruct tracks in a single pass, rather than a two-pass scheme, and it used a specialized seeding algorithm.

Another important aspect of this low-p_{T} tracking setup is the possibility to use a Region of Interest (RoI). In this context, an RoI is a region in z such that only seeds pointing into the RoI are extended into candidates. In other words, low-p_{T} tracks will only have z values in the RoI (and potentially slightly outside of the RoI if the track parameters are adjusted after the seeding stage). The RoI *only* applies for low-p_{T} seeds, not for the first tracking pass, so the nominal tracks are *not* affected. A visualization of an RoI is shown in Figure 16.2. This Low-p_{T} tracking can be run with or without an RoI.

[2] As in Figure 15.2

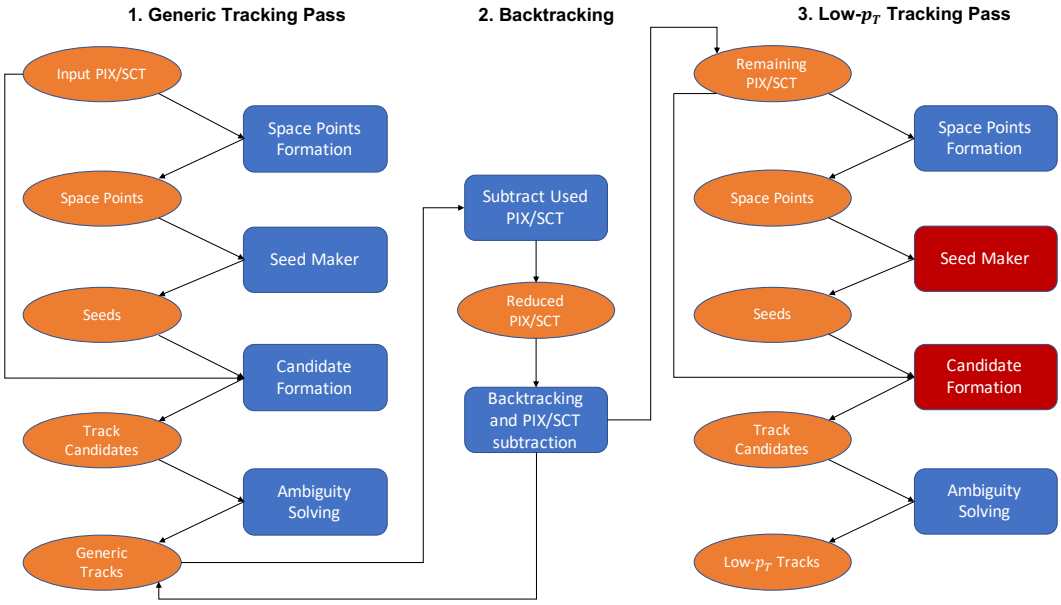


Figure 16.1: An illustrative schematic of the high-pileup low-p_{T} tracking algorithm. Objects and containers are illustrated with orange circles, and algorithms are shown in squares. Red squares indicate algorithms that were modified relative to that applied in generic tracking.

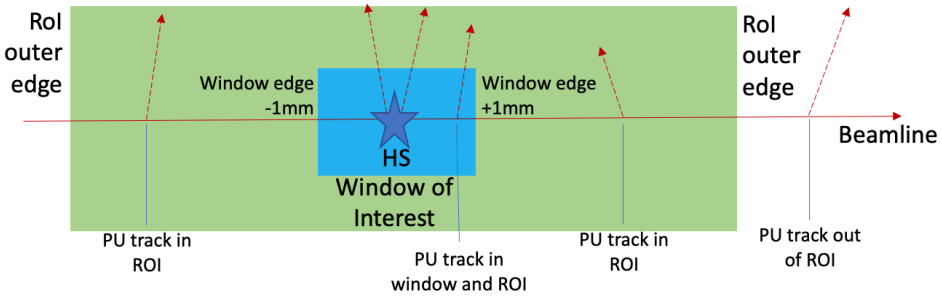


Figure 16.2: An illustration of what a Region of Interest for low-p_{T} tracking looks like. If a collision of interest occurs where the star labelled HS (for hard scatter) is located. We may only care about tracks that are "close", say within ± 1 mm, which is indicated by the blue box. The RoI, which is shown in green, is where low-p_{T} track seeds are allowed to point. It does not have to coincide with the window of *interest*, but can be wider or narrower as desired for performance purposes. Tracks are indicated by red dashed arrows. If a seed points within the RoI, it can be extended into a candidate and eventually be accepted as a track. Any low-p_{T} seed that points outside of the RoI would be rejected at the seeding stage.

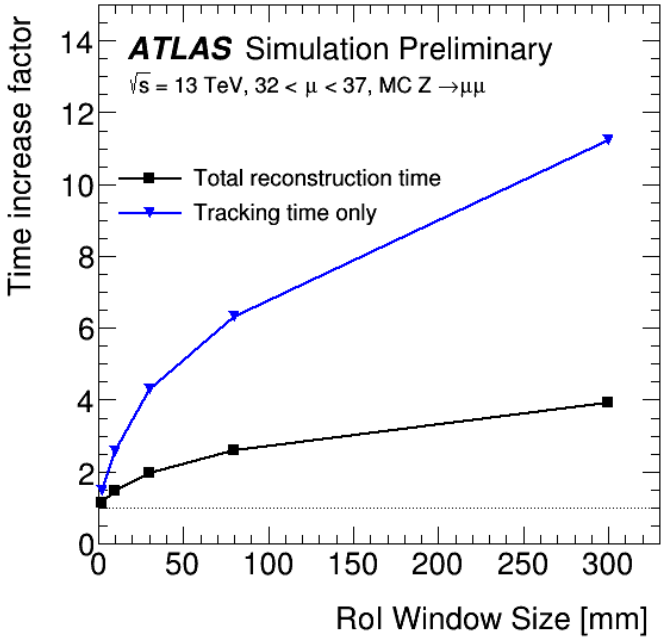


Figure 16.3: Impact of Region of Interest (RoI) size choices on the reconstruction time. The black squares show the fractional increase in the total reconstruction time over the default setup. The blue triangles show the fractional increase in the tracking time alone, where the tracking time is defined as the time spent in both the default tracking pass and in the low-p_{T} tracking pass [210].

The impact of the low-p_{T} tracking pass on reconstruction time as a function of RoI size is shown in Figure 2.10. The simulated sample used in this study was a Powheg+Pythia 8 $Z \to \mu\mu$ sample with pileup of 32–37 events. The RoI was centered around the truth-level location of the Z boson-producing interaction. The sample had a charged particle filter applied, such that there were fewer than 12 truth particles with $p_{\mathrm{T}} > 500$ MeV and $\eta < 2.5$ originating from the Z boson-producing interaction. It is clear that the inclusion of the RoI achieves the goal of reducing reconstruction time impact. The figure shows the impact on the total event reconstruction time and on the tracking component of reconstruction alone. Using an RoI smaller than ±5 mm has minimal impact on reconstruction time, but using a very large RoI, such as ±300 mm, can quadruple the per-event reconstruction time and increase the time spent on tracking by an order of magnitude.

In data, the RoI is centered around the location of the highest energy lepton in the event by default. It can also be centered around a pre-determined location for each event, as read in from an external list. This can be useful in the case that the events were previously processed without low-p_{T} tracking.

16.3 Low-p_{T} Tracking Performance

Having implemented low-p_{T} tracking, it is next necessary to understand the tracking efficiency and fake rate. Efficiency is defined as

$$\text{Efficiency} = \frac{\text{Number of truth charged particles with at least one matched track}}{\text{Number of charged particles}}, \quad (16.1)$$

and fake rate is defined as

$$\text{Fake rate} = \frac{\text{Number of fake tracks}}{\text{Number of tracks}}. \quad (16.2)$$

A fake track is one with a truth-matching probability of less than 50%. Truth-matching probability is defined as [119, 206]

$$\text{Truth-match probability} = \frac{10 * N_{\text{pix,1}} + 5 * N_{\text{SCT,1}} + 1 * N_{\text{TRT,1}}}{10 * N_{\text{pix,tot.}} + 5 * N_{\text{SCT,tot.}} + 1 * N_{\text{TRT,tot.}}}. \quad (16.3)$$

It is a weighted fraction of hits on a track that come from a single particle in different Inner Detector components, $N_{\text{pix,1}}$, $N_{\text{SCT,1}}$, and $N_{\text{TRT,1}}$, where the one particle is that which left the most hits. The denominator is a weighted sum of all of the hits on the track in the Inner Detector. Heuristically, truth-matching is a question of which truth particle contributed the most energy to the hits comprising the track.

16.3.1 Determining an RoI Size

First, we can once again ask about these performance metrics as a function of RoI size, using the same $Z \to \mu\mu$ sample as described in Section 16.2. Figure 16.4 illustrates the above performance metrics in the context of determining an RoI size. The top half of Figure 16.4 shows the efficiency as a function of the RoI size for hard-scatter tracks in a simulated sample, and the bottom half of that plot shows the fake rate for tracks near the hard-scatter interaction location. By considering Figures 16.3 and 16.4, the following conclusions are reached:

- A small RoI (< 10 mm) is fast but has lower efficiency and high fake rate.
- A large RoI (> 100 mm) has low fake rate, has generally slightly lower efficiency, and is time intensive.
- A medium RoI size (around 30 mm) has close to maximal efficiency, at the cost of medium fake rate and a moderate impact on the reconstruction time.

However, for the $\gamma\gamma \to W^+W^-$ analysis, the best choice is running *without* an RoI. While this leads to a slight decrease in efficiency and a large increase in reconstruction time,

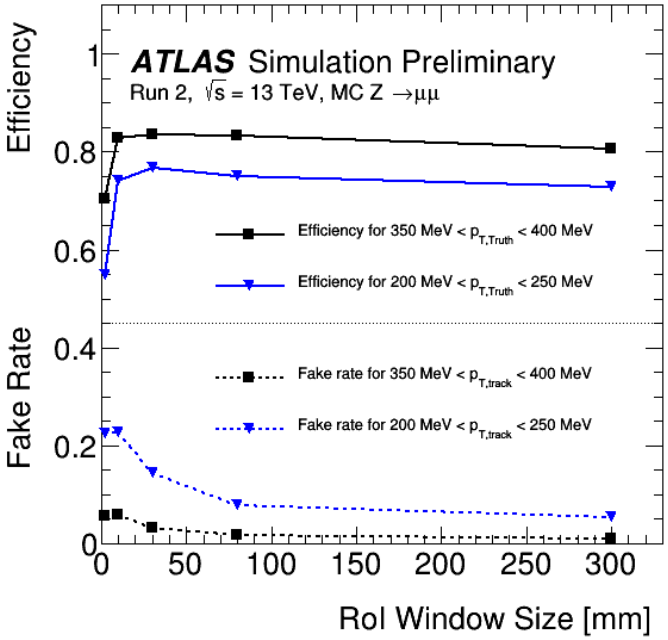


Figure 16.4: Tracking efficiency and fake rate as a function of the Region of Interest (RoI) size for two different truth particle or track p_{T} ranges. Only truth particles originating from the hard-scatter interaction are considered for the tracking efficiency, and only tracks pointing to the hard-scatter interaction location within 1 mm longitudinally along the beamline and with a transverse displacement from the beamline of less than 1 mm are considered for the fake rate [210].

it makes tracking systematics, efficiency, and fake rate studies easier. These studies are performed by looking at locations in the beamspot away from the center of the RoI, similar to the study performed in Chapter 7. Indeed it will be necessary to re-derive the pileup correction, which will also requires tracking without an RoI. When running with an RoI, tracking performance metrics depend not only on RoI size, but also on position within the RoI, so these studies will be distorted by using an RoI. The reconstruction time increase is acceptable due to only running on a small subset of the data (see Section 16.1).

16.3.2 Efficiency and Fake Rate vs p_{T}

Figure 16.5 shows the efficiency and fake rate as a function of $p_{\mathrm{T,truth}}$ and $p_{\mathrm{T,track}}$, respectively, for the no-RoI reconstruction case. The efficiency for default-tracking alone is also shown in the efficiency plot. It can be seen that the low-p_{T} tracking algorithm gives a smooth and relatively stable efficiency above 70% as $p_{\mathrm{T,truth}}$ decreases down to about 200 MeV. In the

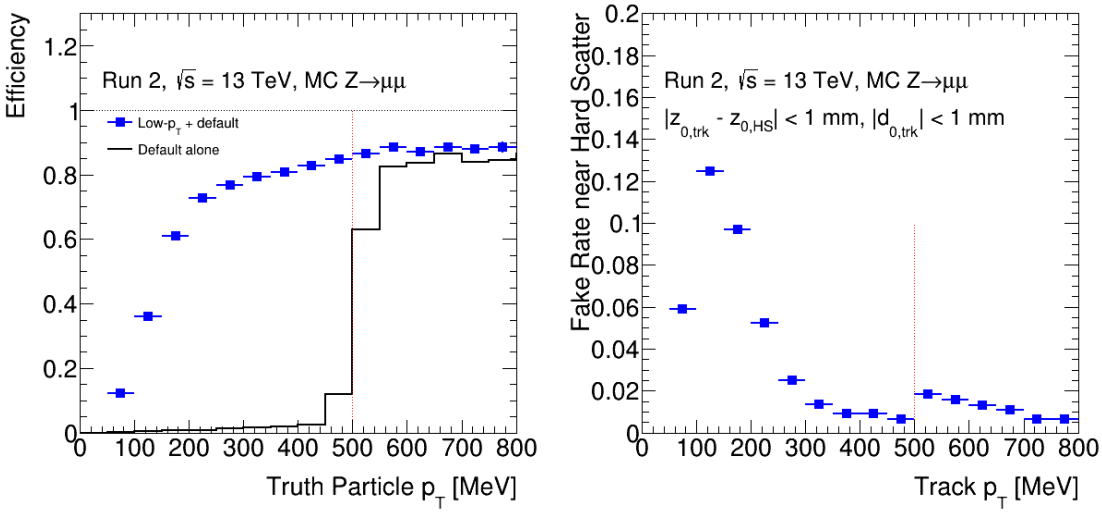


Figure 16.5: (Left) Tracking efficiency as a function of the truth particle p_{T}. The black line indicates the efficiency when default tracking alone is run, and the blue points indicate the efficiency when both default and low-p_{T} tracking are run as in the two-pass scheme. The truth particles considered here originate from the hard-scatter interaction. (Right) Fake rate as a function of the reconstructed track p_{T}, where the tracks are restricted to point near the hard-scatter interaction location. The reconstructed tracks considered here have to point to within 1 mm of the hard-scatter interaction longitudinally along the beamline and be displaced from the beamline by less than 1 mm in the transverse direction. Both default and low-p_{T} tracks are included when the considering fake rate.

very low-$p_{\mathrm{T,truth}}$ regime below 200 MeV, the efficiency falls off quickly. The fake rate stays below 10% for all bins, except between 100 and 150 MeV.

Figure 16.6 shows the seed efficiency, candidate efficiency, and the previously shown track efficiency as a function of $p_{\mathrm{T,truth}}$. The definitions for seed efficiency and candidate efficiency are the same as that in Eq. (16.1), but with "matched seed" and "matched candidate" respectively instead of "matched track". It can be seen that consistently across almost all $p_{\mathrm{T,truth}}$ values, over 90% of truth particles have at least one matched seed, and that above about 200 MeV, typically less than half of the seed-to-track efficiency loss occurs in the seed-to-candidate stage. In the very low-$p_{\mathrm{T,truth}}$ regime, the drop in seed-to-candidate efficiency is more dramatic.

Exploring Seed Quality

Given that a high fraction of truth particles have at least one matched seed, it might be possible to increase the tracking efficiency by accepting a more seeds. It is also worth investigating why the efficiency drop between seed and candidate efficiency is so dramatic

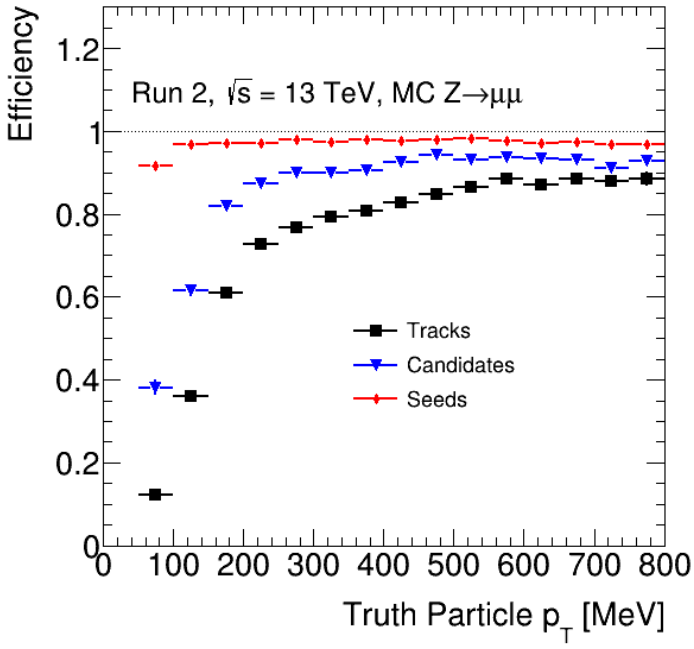


Figure 16.6: Efficiency for the three steps of ATLAS tracking as a function of the truth particle p_{T}. Seed efficiency (red diamonds) is defined here as the likelihood for a stable, charged truth particle within tracking acceptance to have at least one matched seed. Candidate efficiency (blue triangles) and track efficiency (black squares) are defined analogously for candidates and tracks respectively. The truth particles considered here originate from the hard-scatter interaction.

at very low-$p_{\mathrm{T,truth}}$. Addressing both of these concerns requires an investigation of the seed quality.

The primary metric for understanding the seed quality is again the truth-match probability. Seeds are formed from 3 pixel hits or 3 SCT hits, though occasionally extra confirmation hits are used, typically resulting in a seed with 4 or 5 SCT hits. If all 3 hits on a seed are from the same truth particle, the match probability will be 3/3 or 1. If all three hits on the seed are caused by different truth particles, the match probability will be 1/3 or 0.333. This latter case is considered a fake seed as it is comprised of a random collection of hits. A seed is considered "fake" if the match probability is less than 0.5.

Figure 16.7 shows the proportion of good and fake seeds as a function of the predicted $p_{\mathrm{T,seed}}$. In the very low-$p_{\mathrm{T,seed}}$ regime, almost 40% of the seeds are fake. This calls into question the idea of simply "accepting more seeds" to improve the efficiency. If the seed is fake, one cannot expect it to be extended through additional hits to make a candidate, which could explain the large seed-to-candidate efficiency drop at low-p_{T}. Given the high fake rate

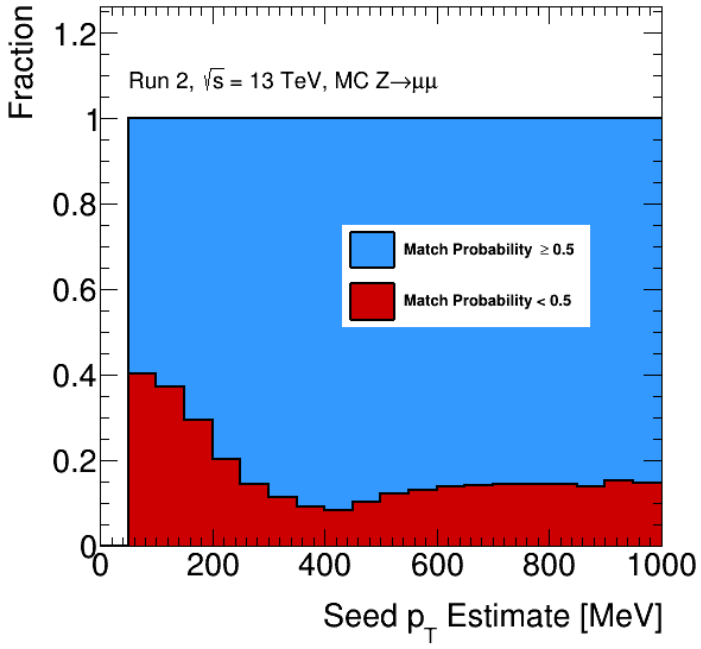


Figure 16.7: Seed composition as a function of the seed p_{T}, where the seed p_{T} is an estimate from the 3 (or more) hits that comprise the seed. The seeds considered for this plot are required to point to within 5 mm of the hard-scatter interaction longitudinally along the beamline and to have a transverse displacement of less than 2 mm. The blue region shows the fraction of seeds where the truth matching probability is greater than 0.5 ("good"), and the red area shows the fraction of seeds where the truth matching probability is less than 0.5 ("fake").

for seeds, loosening selection criteria would an increase in fake rate for tracks rather than a gain in efficiency.

Increasing the efficiency for very low-p_{T} tracks will be difficult; there is a large combinatorial background such that almost half of the seeds are comprised of hits of random origin. Any increase in efficiency will come at a high cost in fake rate.

Conversely, reducing the fake rate without a drastic loss of efficiency might be possible by tightening seed or track selection criteria. One option is to do this at reconstruction level, but in order to not discard information, this could also be done offline, once a complete set of tracks has been finalized. For example, there could be a set of track selections based on p_{T}, η, number of pixel hits, number of SCT hits, or number of holes. Any reduction in fake rate will necessarily come with a loss of efficiency. However, the appropriate balance of efficiency and fake rate is an analysis-specific question, which will be investigated in the next chapter.

Chapter 17

Selecting Tracks and Events

A strategy for the $\gamma\gamma \to W^+W^-$ re-analysis with low-p_{T} tracks was laid out in Section 16.1. In that section, it was noted that track selections and event selections used in the analysis would have to be re-evaluated with the addition of new tracks.

In the first round of the $\gamma\gamma \to W^+W^-$ analysis, tracks were required to pass a pre-existing "tight-primary" selection[1]. This is a cut-based boolean selection designed to decrease fake rate for nominal tracks. It is standard across ATLAS analyses, with well-understood performance. For low-p_{T} tracks, there is no such pre-existing selection. While the expected fake rate, shown in Figure 16.5, is not extremely high above a p_{T} of approximately 300 MeV, it grows rapidly at lower p_{T}. Because fake tracks could result in loss of signal events, we would like to develop a set of selections to ensure that these new tracks are high-quality.

Similarly, in the first round of the analysis, the exclusivity selection for events was a simple cut on whether a track was found within ± 1 mm of the lepton vertex in both the z and d_0 variable. While this choice optimized the expected significance of the analysis in Part II, low-p_{T} tracking will roughly double the number of tracks per event. Because of this, the exclusivity selection must be updated.

17.1 Low-p_{T} Track Selection

Two different schemes for track selection will be explored in this section. First, it is possible to make a cut-based selection similar to the tight selection used for high-p_{T} tracks. Second, it is possible to use a machine learning (ML) based approach, which should better take advantage between correlations in the variables used to distinguish true and fake tracks. While the performance of the ML based approach will likely be better than the cut-based approach, it does have the drawback of requiring more extensive validation.

[1] see Section 4.1

17.1.1 Cut-based Selection Approach for Low-p_{T} Tracks

The pre-existing tight selection for high-p_{T} tracks involves the following cuts:

- $|\eta| < 2.5$
- Minimum of 9 pixel and SCT clusters if $|\eta| \leq 1.65$
- Minimum of 11 pixel and SCT clusters if $|\eta| > 1.65$
- Track can share a maximum of either one pixel hit or two SCT clusters on the same layer with another track
- Track can have a maximum of 2 holes[2] in the SCT layers and no holes in the pixel layers
- At least 1 hit in the IBL or B-Layer.
- $|d_0| < 2.0$ mm and $|z_0 \sin\theta| < 3.0$ mm.

When the tight selection is applied to nominal tracks in the HERWIG 7 $\gamma\gamma \to W^+W^-$ sample, 90.6% of true tracks are accepted and 20.5% of fake tracks are accepted[3]. Here a true track is one with truth-match probability [206] above 0.5, and a "fake" track is one with truth-match probability below 0.5. The p_{T} dependence of this selection is shown in Figure 17.1. Both the true and fake acceptance rates are fairly stable, with the true rate decreasing slightly with lower p_{T}, and the fake rate increasing slightly.

To put this in context, in the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample, 99% of nominal tracks *without* the tight selection are considered true. There are an average of 0.05 fake tracks per event in a ± 1 mm window around the lepton vertex. After the tight selection, this drops to 0.01 fake tracks per event[4]. In the $q\bar{q} \to W^+W^-$ sample, there is more activity around the leptons, and a slightly higher fake rate, resulting in 0.06 fake tracks per event in the exclusivity window prior to the tight selection. Figure 17.2 shows the average number of fake tracks in the exclusivity window before and after the application of the tight selection, binned in p_{T}.

To contextualize the fake rates for low-p_{T} tracks, a similar exercise can be performed. For tracks produced in the second pass of the new tracking algorithm (called "low-p_{T}" tracks here, though they may have $p_{\mathrm{T}} > 500$ MeV), 96.4% of tracks are considered true. However, the ± 1 mm window around a lepton vertex contains 0.25 fake tracks on average. Because of this, reducing the fake rate will be very important. Figure 17.3 shows the p_{T} dependence the fraction of tracks that are fake and the average number of fake tracks in the exclusivity window, binned in p_{T}. This can be compared directly to Figure 17.2.

[2]From Section 4.1: a "hole" occurs when a track's trajectory is extrapolated through an active silicon layer, but the expected hit there is not found.

[3]These values are slightly sample-dependent, but the dependence is at the level of 1%.

[4]The number of true tracks per event is highly process-dependent, but about 90% of of true tracks are retained.

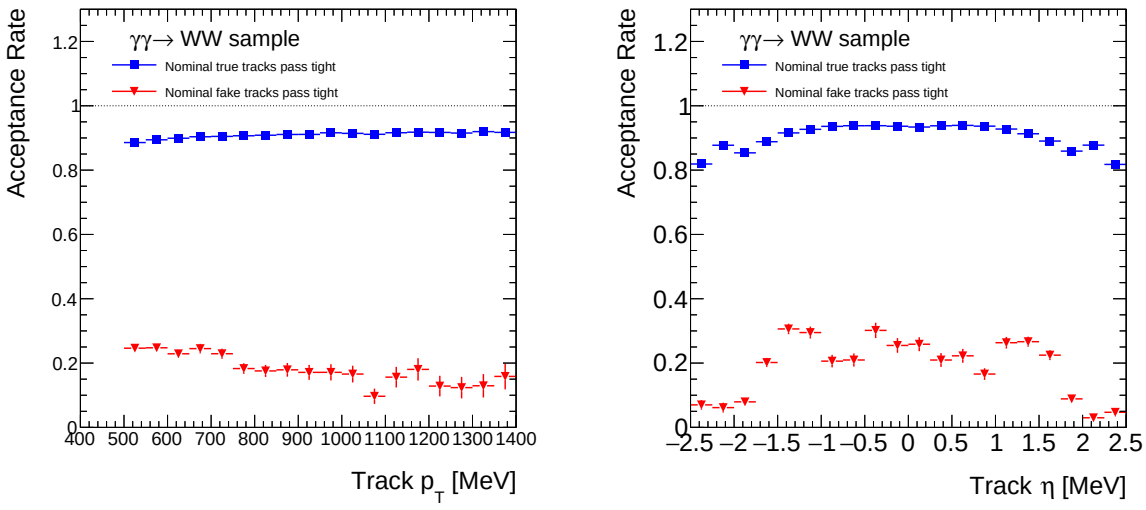


Figure 17.1: The p_{T} dependence (left) and η dependence (right) of the acceptance rate of true and fake nominal tracks after the application of the tight selection. The tracks must have $|d_0| < 1.0$ mm. Both of these plots were made using tracks from the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks.

To develop a cut-based track selection similar to the tight selection, we must consider possible discriminating variables, or variables that we expect to differ on average between true and fake tracks. The distributions of the following potential variables are shown in Figure 17.4:

1. Track p_{T}
2. Track η
3. Track $|d_0|$
4. Uncertainty of the track z parameter resulting from the track fit
5. Uncertainty of the track d_0 parameter resulting from the track fit
6. Number of holes in the SCT Layers
7. Number of hits in the SCT layers plus the number of dead SCT modules the track is projected to pass through
8. Number of holes in the pixel Layers
9. Number of hits in the pixel layers plus the number of dead pixel modules the track is projected to pass through

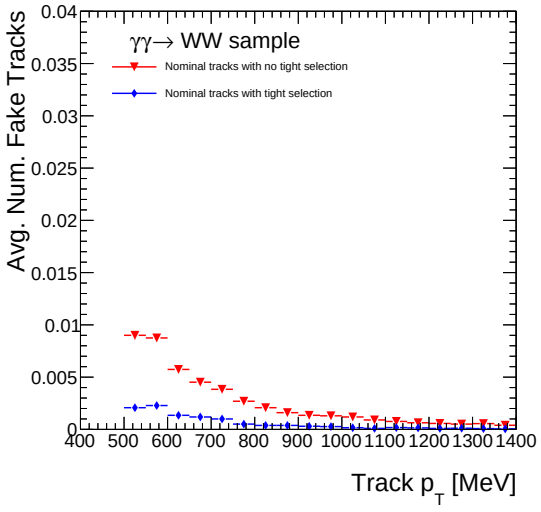


Figure 17.2: The average contribution of fake nominal tracks with and without tight selection that fall into a ±1 mm window around the lepton vertex, as a function of track p_{T}. When there is a fake track in such a window, it is more likely to have lower-p_{T}. This plot was made using tracks from the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks. The tracks are required to have $|d_0| < 1.0$ mm. On average 99% of nominal tracks are considered true.

10. Boolean for whether the track has a hit on the IBL or B Layer (if no IBL hit is expected)

11. The number of degrees of freedom of the track fit

12. The dE/dx of the track as it passed through the pixel layers. This is the rate of energy loss per unit length in the pixels' silicon [211], which depends on a particle's velocity.

The true and fake distributions exhibit some degree of overlap for each of these variables, so it will be impossible to make a perfect cut-based classifier.

To optimize the cut-based track selection, a grid scan was performed over a number of track variables: d_0, σ_z, σ_{d_0}, number of SCT holes, number of SCT hits and dead modules, number of pixel holes, number of pixel hits and dead modules, whether or not there is a hit in the IBL or B-Layer (if no IBL hit is expected), and the number of degrees of freedom in the track fit. The goal of the grid scan was to optimize the significance of the $\gamma\gamma \to W^+W^-$ analysis, based on the metric from Equation 15.2, using the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample and the POWHEG+PYTHIA 8 $q\bar{q} \to W^+W^-$ sample. The potential cuts were subdivided into a p_{T} and η grid, depending whether the track p_{T} is greater than or equal to 250 MeV and whether the $|\eta|$ is greater than or equal to 1.0.

Ultimately the set of variables used in the cleaning cuts were:

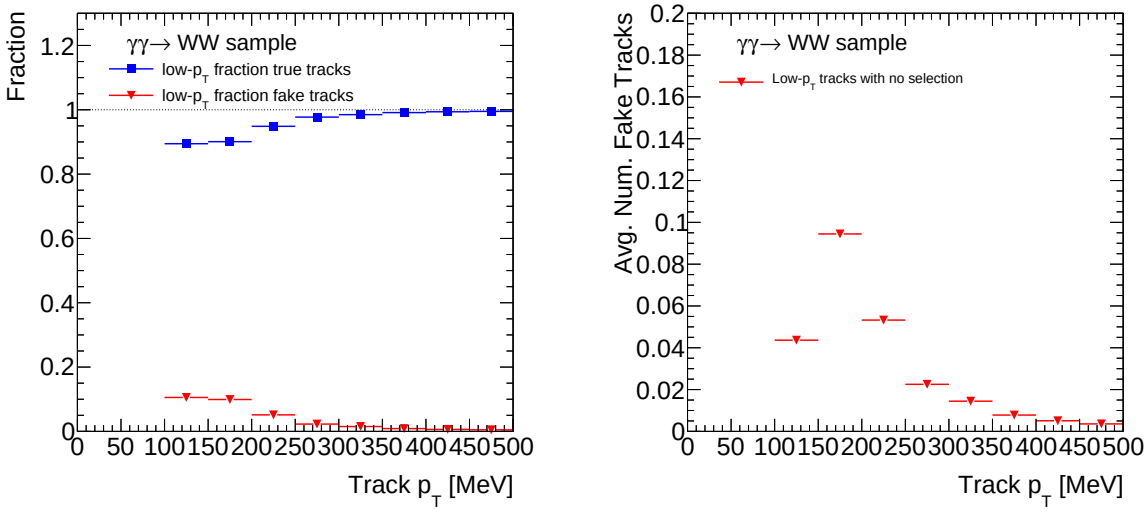


Figure 17.3: (Left) The fraction of low-p_{T} tracks that are true and fake as a function of track p_{T}, *before* any selection is applied. On average 96.4% of these tracks are considered true. (Right) The average contribution of fake low-p_{T} tracks that fall into a ± 1 mm window around the lepton vertex, as a function of track p_{T}. No selection is applied to these tracks. Both of these plots were made using tracks from the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks. The tracks for both must have $|d_0| < 1.0$ mm and $p_{\mathrm{T}} > 100$ MeV.

1. track $p_{\mathrm{T}} > 100$ MeV
2. Number of SCT holes ≤ 2
3. Number of pixel holes $= 0$
4. Number of pixel hits + dead modules ≥ 4
5. $\sigma_{d_0} < 3.0$ for tracks with $p_{\mathrm{T}} > 250$ MeV
6. There must be a hit in either the IBL or the B-Layer if no hit in the IBL is expected.

The other variables were did not provide significant additional discrimination.

When this selection is applied, 82.9% of true tracks and 23.2% of fake tracks pass. The p_{T} dependence of the cut survival is shown in Figure 17.5. This performance is similar to, but slightly worse than the application of the tight cut to high-p_{T} tracks; however, the low-p_{T} tracks are generally lower quality, so this is not unexpected. Prior to the application of the cuts, there was an average of 0.25 fake tracks per event in a ± 1 mm window around the leptons, but with the cuts, this is reduced to 0.06 fakes per event. This drop can be also be seen in Figure 17.5. In the $q\bar{q} \to W^+W^-$ sample, the performance is similar, with

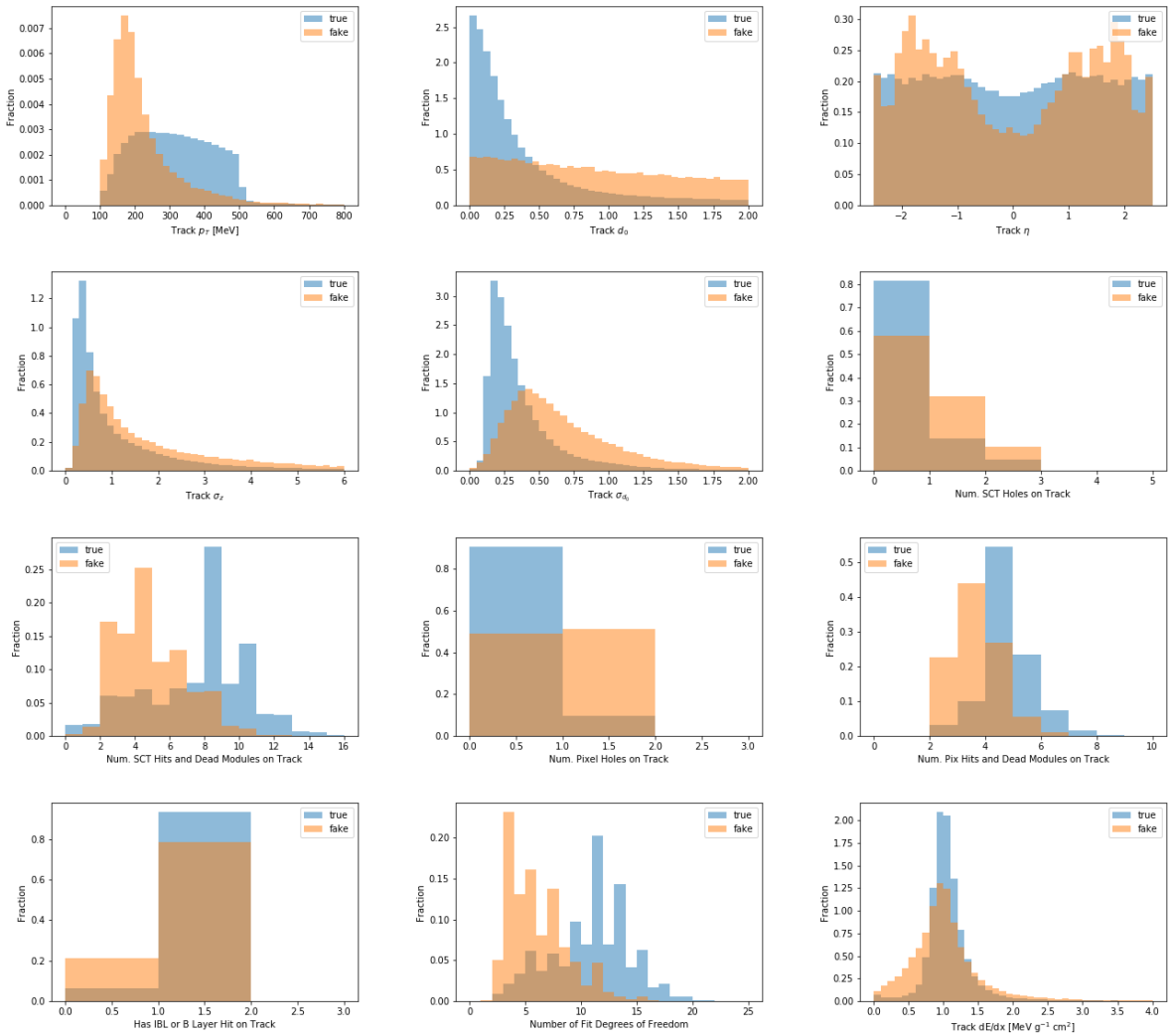


Figure 17.4: Distributions of discriminating variables for true and fake tracks. The true distributions are shown in blue and the fake distributions are in orange. Shown are track p_{T}, η, d_0, the resolutions of the z and d_0 parameters, the number of holes in the SCT and pixel layers, the number of hits and dead modules in the SCT and pixel layers, a boolean for whether there is a hit in the IBL or B Layer, the number of degrees of freedom in the track fit, and the track dE/dx. The distributions are normalized to have an area of 1.0. In this sample, 96.4% of tracks are true, so the fake distributions are much smaller when normalized to the actual yields.

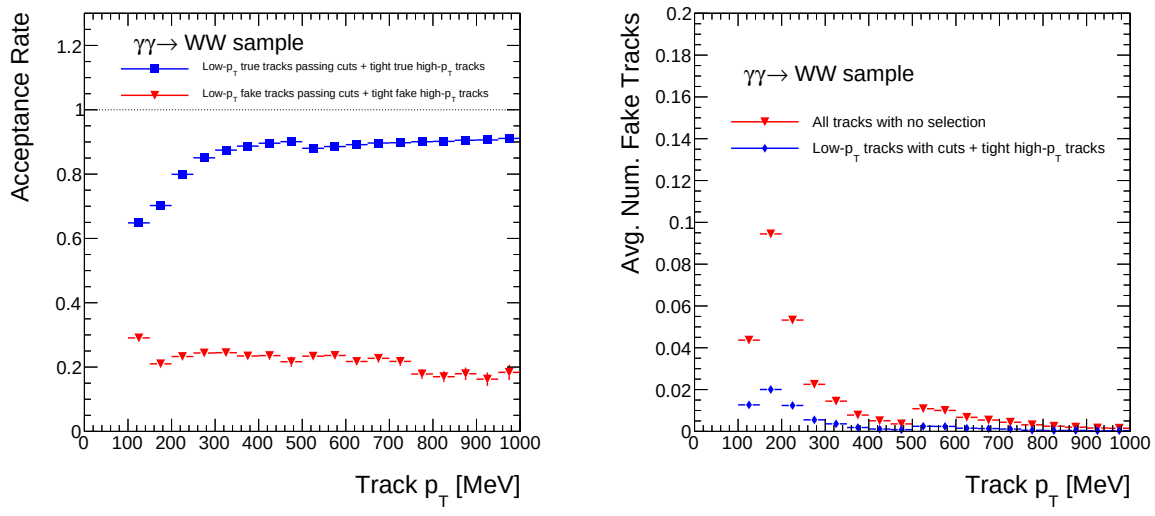


Figure 17.5: (Left) The p_{T} dependence of the acceptance rate of true and fake low-p_{T} tracks upon the application of the recommended cleaning cuts. (Right) The average contribution of fake low-p_{T} tracks that fall into a ± 1 mm window around the lepton vertex, as a function of track p_{T}. The contribution is shown before and after the recommended cut-based selection is applied. The tracks in this plot must have $|d_0| < 1.0$ mm and $p_{\mathrm{T}} > 100$ MeV. Both of these plots were made using tracks from the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks.

84.6% of true tracks and 23.9% of fake tracks passing the cuts. Similarly, 0.07 fake tracks are expected per event, where 0.29 are expected before the selection. This independent set of simulated tracks provides validation of the cuts and their expected performance.

Other working points achieved true and fake track acceptance rates of:

- 71.3% true and 17.6% fake, by allowing no holes in any pixel layer but placing no hit and dead modules requirements
- 80.5% true and 23.7% fake, by allowing 1 SCT hole, 0 pixel holes, and requiring at least 4 pixel hits plus dead modules for tracks with $|\eta| < 1.0$ and 3 pixel hits otherwise
- 87.4% true and 59.0% fake, by allowing 1 SCT hole, 1 pixel hole, and requiring at least 3 pixel hits plus dead modules for all low-p_{T} tracks except those with $p_{\mathrm{T}} > 250$ MeV and $|\eta| < 1.0$, where 4 hits are required.

Because hits and holes are are integer values, the performance can leap dramatically when changing the values, as can be seen in Figure 17.4.

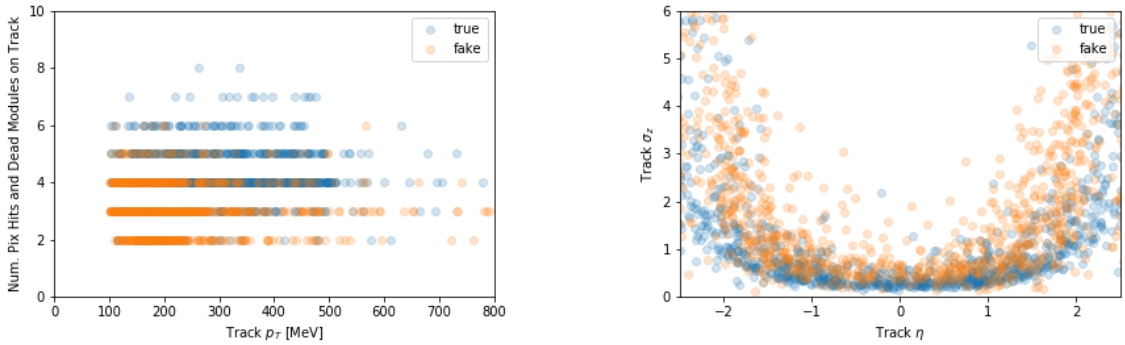


Figure 17.6: (Left) Scatter plot of p_{T} and number of pixel hits plus dead modules for 1000 true and 1000 fake tracks. (Right) Scatter plot of η and uncertainty on the z parameter in the track fit for 1000 true and 1000 fake tracks. The tracks in these plots must have $|d_0| < 1.0$ mm and $p_{\mathrm{T}} > 100$ MeV. Both of these plots were made using tracks from the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks.

17.1.2 Machine Learning Approach for Low-p_{T} Tracks

While the cut-based approach to track selection exhibits good performance, it is not optimal. If we represent the possible feature space of a track as existing in some high-dimensional space, simple cuts represent "horizontal" or "vertical" decision planes, where tracks with a feature on one side of the plane can be accepted and tracks on the other side must be rejected. Admittedly, some granularity was included by allowing for different cuts at different p_{T} and η values, but the optimal working point actually did not take advantage of this.

In reality, the optimal decision boundary in the high-dimensional space will be some complicated function of all of the track features. Figure 17.6 shows that even in two dimensions the decision boundary can be non-linear. For example, we could imagine making a circular boundary in the p_{T} and pixel hit plot or a hyperbolic boundary in the η and z uncertainty plot. If all four variables were used, the boundary would be more complicated and essentially impossible to visualize.

To estimate such a multidimensional function, we can turn to the Machine Learning (ML) technique known as a Neural Network (NN) [212]. The natural language of higher dimensional spaces is linear algebra, and NNs are essentially an algorithm for numerical function approximation. The approximation is performed by minimizing a value called the "loss", which is a metric to express how good or bad a set of classifications are. In the context of this thesis for example, we have a collection of true and fake tracks. The more accurate our classification, the lower the loss. A simple schematic of a NN is shown in Figure 17.7. Features of the tracks are fed into the first layer, and then linear combinations of these features are formed in "hidden" layers. There may be multiple hidden layers, where linear combinations of linear combinations are created, but eventually these values are reduced into

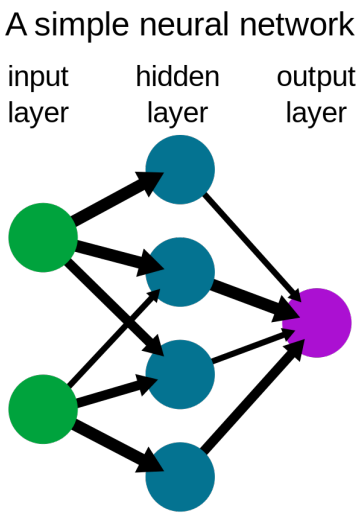


Figure 17.7: Diagram of a simple neural network [213]. In the track selection case, features of the track are fed into the first layer. Linear combinations of the features are formed and fed into hidden layers, before these values are reduced into a single value which attempts to predict whether the track is a true track or a fake track. The weights that form the linear combinations algorithmically tuned to make the guess as accurate as possible.

a single value. This value attempts to predict whether the track is a true track or a fake track.

It is possible to control the number of hidden layers, the number of nodes in these layers, the definition of the loss function, and many other features of the NN. Practically speaking though, what happens is that we feed a single track, described by some number of features, into the NN, and a single value gets output. If the value is close to 1, the NN thinks that the track is more likely to be true, and if the value is closer to 0, then the track is more likely to be fake.

To build our true vs. fake track discriminating NN, we use the NN package provided by Keras [214, 215]. We represent each track by 15 features:

1. Track p_{T}.
2. Absolute value of track η
3. Absolute value of track d_0
4. Uncertainty of track z parameter from track fit
5. Uncertainty of track d_0 parameter from track fit
6. Number of SCT holes

7. Number of SCT hits and dead modules
8. Number of pixel holes
9. Number of pixel hits and dead modules
10. Boolean for whether there is a hit in either the IBL or the B-Layer if no hit in the IBL is expected
11. Track fit χ^2
12. Number of degrees of freedom in the track fit
13. Absolute value of track q/p, where q is charge and p is momentum
14. Covariance of track q/p resulting from the track fit
15. Whether the track was made in the generic tracking pass or in the low-p_{T} tracking pass.

The dE/dx of tracks was also explored as a potential discriminant, though the impact of its inclusion was negligible.

The NN was set up using two hidden layers of 34 nodes each with rectified linear unit activation functions for each hidden layer and a final layer with 1 node and a sigmoid activation function to make the output of the NN between 0 and 1. A binary cross-entropy loss function was used with the Keras "adam" optimizer. For the training itself, 22 epochs were used to train on tracks from the $q\bar{q} \to W^+W^-$ sample. Both low-p_{T} and high-p_{T} tracks were used for training. The results were validated using the $\gamma\gamma \to W^+W^-$ sample. There was no evidence of overtraining.

Receiver-Operator Characteristic (ROC) curves for the performance of the NN are shown in Figure 17.8 for low-p_{T} tracks and high-p_{T} tracks. These figures also display the working points associated with cut-based selection schemes discussed in Section 17.1.1. Having a higher true acceptance rate and a lower fake acceptance rate is better, and it can be seen that the NN outperforms the cut-based approaches for both low-p_{T} and high-p_{T} tracks.

Figure 17.9 shows 1D plots of the NN scores for true and fake tracks, split by whether they are low-p_{T} or high-p_{T} tracks. As expected, true tracks have scores that peak near 1. A scan of NN cut values was performed to find the optimal working point for the $\gamma\gamma \to W^+W^-$ analysis. This was found to be 0.92 for low-p_{T} tracks, resulting in a 87.9% true acceptance rate and a 13.1% fake acceptance rate. For high-p_{T} tracks, the optimal cut value was found to be 0.97, resulting in a 94.9% true acceptance rate and a 6.5% fake acceptance rate. The performance of these cuts against track p_{T} is shown in Figure 17.10. The performance is similar in the $q\bar{q} \to W^+W^-$ sample, where 89.4% of true low-p_{T} tracks are accepted with 13.0% of fake low-p_{T} tracks, and 95.0% of true high-p_{T} tracks are accepted with 6.7% of fake high-p_{T} tracks for the prescribed NN score cuts. This serves as important validation of the NN.

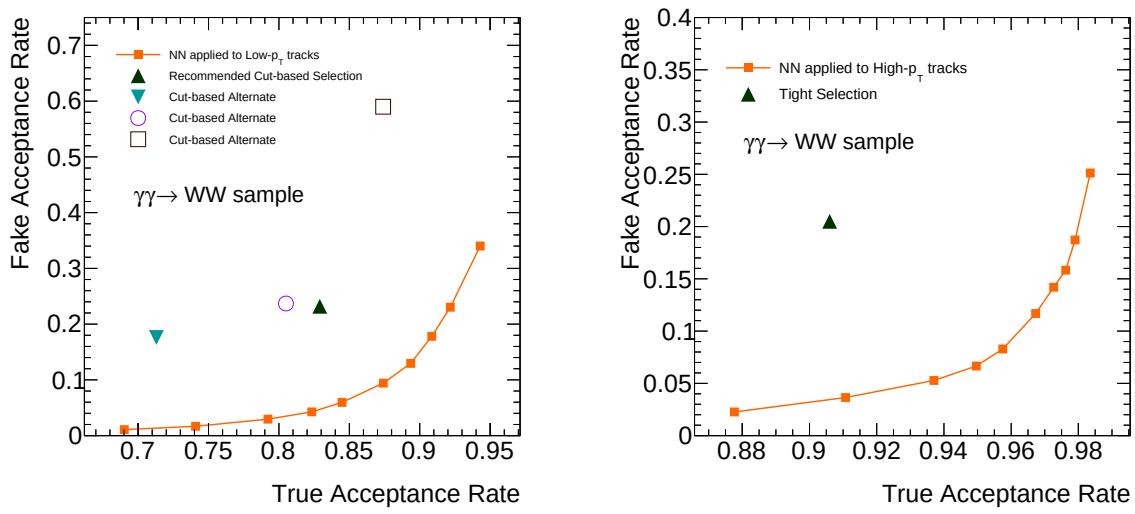


Figure 17.8: (Left) Receiver-Operator Characteristic (ROC) curve for the application of the true-fake track discrimination neural network when applied to low-p_{T} tracks. The x-axis shows the true acceptance rate, and the y-axis shows the fake acceptance rate for a given cut choice. Working point associated with the cut-based selections discussed in Section 17.1.1 are shown for comparison. (Right) ROC curve of the NN when applied to high-p_{T} tracks. The "tight-primary" working point is also shown. Both of these plots were made using tracks from the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks. Down and to the right also represents better performance in both plots.

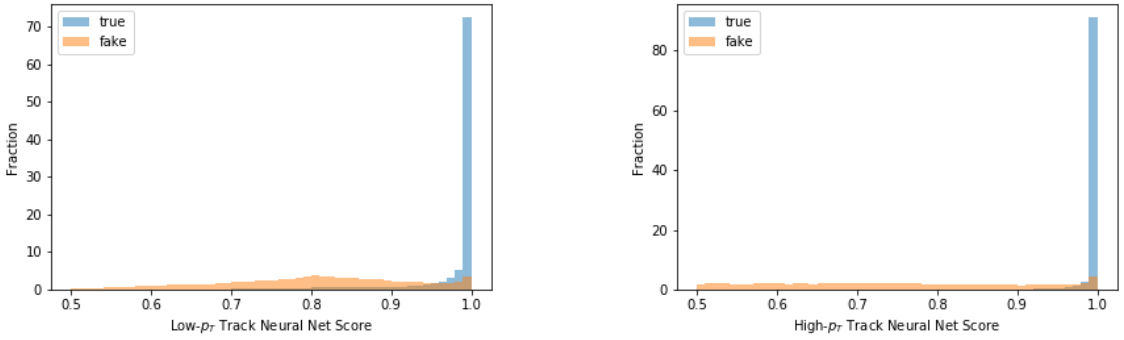


Figure 17.9: (Left) Neural network scores for true (blue) and fake (orange) low-p_{T} tracks. (Right) Neural network scores for true (blue) and fake (orange) high-p_{T} tracks. Both of these plots were made using tracks from the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks.

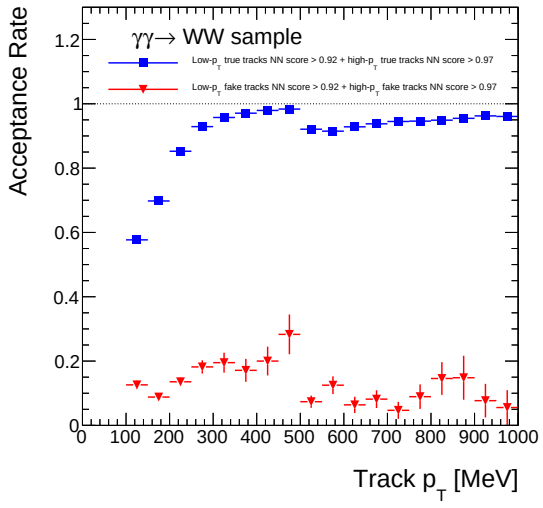


Figure 17.10: Acceptance rates for true and fake low-p_{T} and high-p_{T} tracks after the application of a neural net score cut. A cut of 0.92 is applied to low-p_{T} tracks, and a cut of 0.97 is applied to high-p_{T} tracks. This plot was made using tracks from the HERWIG 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks.

With the NN selection, an expected contribution of 0.03 fake low-p_{T} tracks is expected within a ± 1 mm window of the lepton vertex per event. The expected contribution of fakes in bins of p_{T} is given in Figure 17.11. Similarly, about 0.007 fake high-p_{T} tracks are expected with the application of the NN selection. In the $q\bar{q} \to W^+W^-$ sample, there is an expected average contribution of 0.05 fake low-p_{T} tracks and 0.01 fake high-p_{T} tracks.

As a validation that the NN creates non-linear decision boundaries, Figure 17.12 shows 2D scatter plots analogous to those in Figure 17.6, but where "true" and "fake" tracks have been replace with accepted and rejected tracks, based on the suggested NN cut for low-p_{T} tracks. The overlapping distributions would not be possible with cut-based selections.

While the NN performs well when compared to the cut-based methods, it is important to keep in mind that the NN was trained and tested on simulated tracks. Simulating tracks is difficult, so the performance of the NN in data may be different from expected. Because of this, validation of the NN performance in data is needed, and systematics uncertainties must be assessed.

17.2 New Exclusivity Selections

At this point, we have a new set of low-p_{T} tracks and algorithms to reject poor quality tracks from consideration. We can now re-evaluate our exclusivity selection. If a study similar to that in Chapter 15 is performed, with the same exclusivity selection of Part II, a

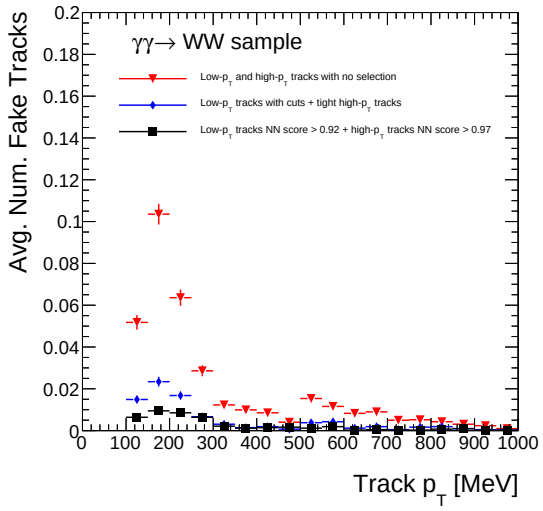


Figure 17.11: Average contribution of fake low-p_{T} tracks that fall into a ± 1 mm window around the lepton vertex, as a function of track p_{T}. The contribution is shown for the application of the track-selection neural net and before and after the recommended cut-based selection for comparison. The tracks in this plot must have $|d_0| < 1.0$ mm and $p_{\mathrm{T}} > 100$ MeV. Both of these plots were made using tracks from the Herwig 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks.

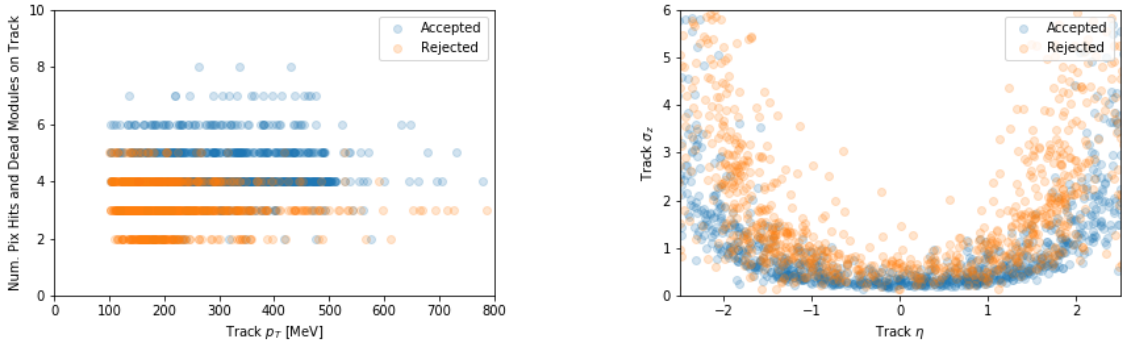


Figure 17.12: (Left) Scatter plot of p_{T} and number of pixel hits plus dead modules for 1000 accepted and 1000 rejected low-p_{T} tracks after applying the suggested neural network cut. (Right) Scatter plot of η and uncertainty on the z parameter in the track fit for 1000 accepted and 1000 rejected low-p_{T} tracks. The tracks in these plots must have $|d_0| < 1.0$ mm and $p_{\mathrm{T}} > 100$ MeV. Both of these plots were made using tracks from the Herwig 7 elastic $\gamma\gamma \to W^+W^-$ sample, which are either pileup or fake tracks.

peak expected significance of 8.7σ is found, which uses a minimum track p_{T} of 300 MeV. If the minimum considered track p_{T} is 500 MeV, there is an expected significance of 6.1σ.

However, the ± 1 mm window size is no longer the best choice when low-p_{T} tracks are introduced. Similarly, we can restrict the $|\eta|$ of considered tracks. When a grid scan is performed over the minimum track p_{T}, the maximum low-p_{T} track $|\eta|$, and the window size, the following values combine to give maximal expected significance:

- Window size of ± 0.65 mm around the lepton vertex
- Minimum considered track p_{T} of 120 MeV
- Maximum track $|\eta|$ of 1.82 for low-p_{T} tracks. High-p_{T} track are still considered for any $|\eta| < 2.5$.

These cuts actually maximize both the significance when tracks are selected using the NN and the significance when tracks are selected with the cut-based method.

Figure 17.13 shows the expected significance as a function of the cuts. Each plot shows a scan over one variable, with the other variables held constant at the value in the list above.

The expected significance at the optimized working point is 9.65σ. When considering the expected impact of low-p_{T} tracks on the $\gamma\gamma \to W^+W^-$ analysis, this significance estimate should be compared to the 6.1σ expected with a non-optimized event selection referenced above, rather than the 8.4σ observed in data. The uncertainties are accounted for conservatively in Equation 15.2, and the signal yield prediction does not include the factor of 1.33 observed in data.

17.2.1 Additional Signal Region Changes

Adding low-p_{T} tracks to the $\gamma\gamma \to W^+W^-$ analysis will inevitably result in some signal loss, at least for a given window size. Of course, the major reason for adding these tracks is the dramatic suppression of background yields, which consequently gives us some flexibility in selecting events based on lepton kinematics.

For example, Figure 13.7 shows that the $p_{\mathrm{T},\ell\ell}$ cut of 30 GeV for the signal region is justified. Below this cut, backgrounds dominate, particularly Drell-Yan (DY) $\tau^+\tau^-$ production. However, the yield of this background, as well as the sub-dominant $q\bar{q} \to W^+W^-$ background, will decrease with the addition of low-p_{T} tracks, potentially allowing us to shift the $p_{\mathrm{T},\ell\ell}$ cut down. If the exclusivity selection suggested in Section 17.2 is used with the track-selection NN, we can scan over possible $p_{\mathrm{T},\ell\ell}$ requirements for the signal region. The results are visualized in Figure 17.14.

The expected significance metric here is similar to Equation 15.2, with a slight modification:

$$\sigma = \frac{3.59 * s}{\sqrt{3.59 * s + 1.2 * b + DY + (0.2 * 1.2 * b + 0.2 * DY)^2}}. \tag{17.1}$$

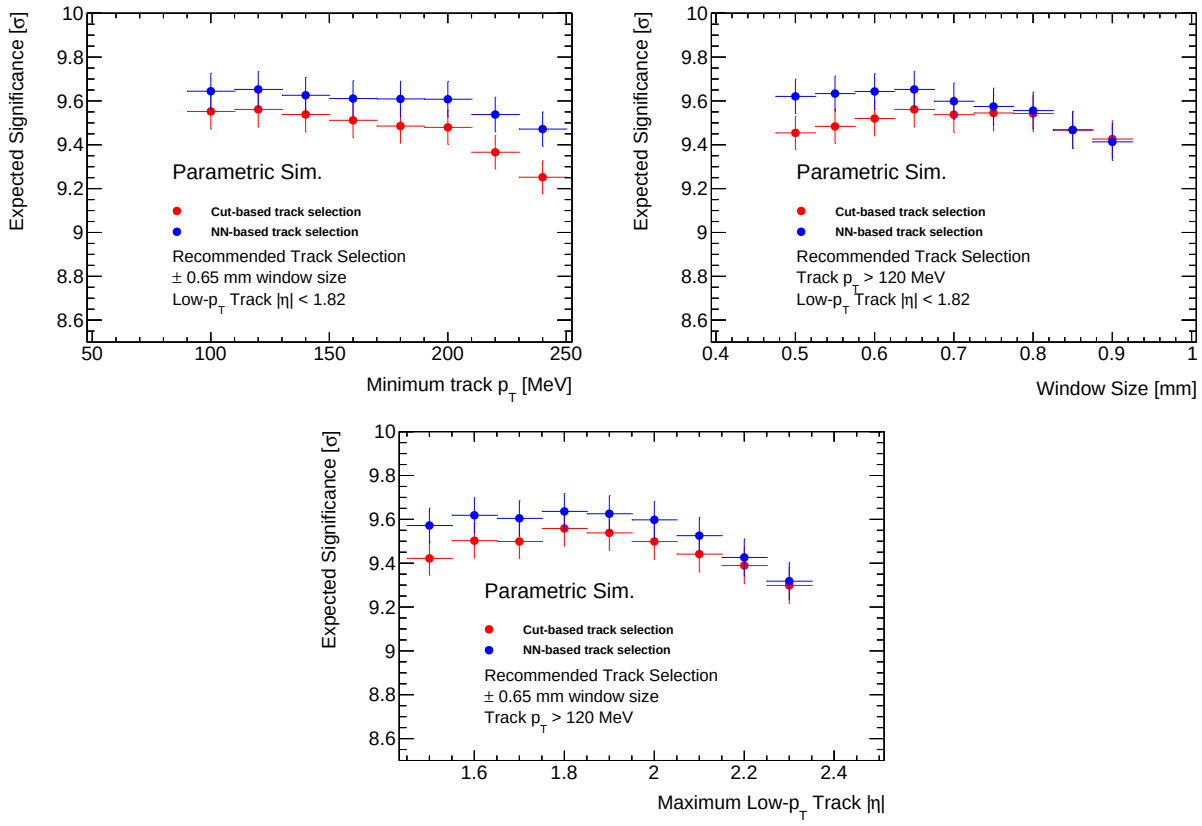


Figure 17.13: The expected significance as a function of the minimum considered track p_{T} (top left), the exclusivity window size (top right), and the maximum considered η of low-p_{T} tracks (bottom). For the window size plot, the values on the x-axis represent 1/2 of the window size; for example, the observed peak occurs for a window that is ± 0.65 mm around the lepton vertex.

Here, s is the expected yield of the elastic-only $\gamma\gamma \to W^+W^-$ samples, and b is the expected yield of the $q\bar{q} \to W^+W^-$ sample with the UE correction applied. DY is new in this equation, representing the expected yield of Drell-Yan events. The factor applied to the $q\bar{q} \to W^+W^-$ yield has been changed to 1.2, because a DY sample is being used to estimate that yield independently. This sample must be used here, as the expected $p_{\mathrm{T},\ell\ell}$ distribution is different for the DY and $q\bar{q} \to W^+W^-$ processes. Based on this study, the significance is maximized for a signal region $p_{\mathrm{T},\ell\ell}$ cut of 20 GeV, though the expected significance at 15 GeV and 25 GeV are similar. These estimates neglect any $p_{\mathrm{T},\ell\ell}$ dependence of other backgrounds, such as $\gamma\gamma \to \tau^+\tau^-$ and contributions from fake leptons. The expected yields of the three samples are given in Table 17.1, as are the expected significances[5]. Because of the

[5]The significance prediction for the $p_{\mathrm{T},\ell\ell} > 30$ GeV requirement is slightly different from the 9.65σ

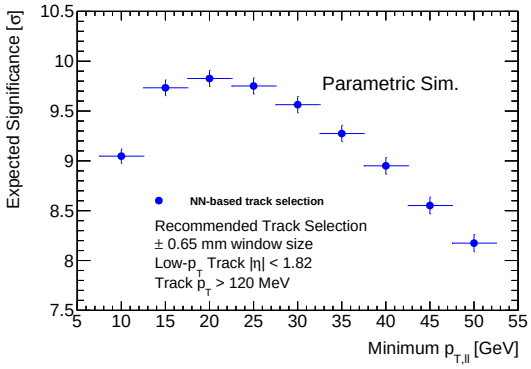


Figure 17.14: The expected significance as a function of the minimum $p_{\mathrm{T},\ell\ell}$ cut for a new signal region. In the first round of the analysis, this was placed at 30 GeV.

suppression with low-p_{T} tracks, even when the $p_{\mathrm{T},\ell\ell}$ cut is lowered to 20 GeV, the expected yield of DY is lower than in the analysis without low-p_{T} tracks, given in Table 13.1. We also see that if the $p_{\mathrm{T},\ell\ell}$ cut is lowered to 20 GeV from 30 GeV, the expected signal yield increases by 13%.

Process	$p_{\mathrm{T},\ell\ell} > 15$ GeV	$p_{\mathrm{T},\ell\ell} > 20$ GeV	$p_{\mathrm{T},\ell\ell} > 25$ GeV	$p_{\mathrm{T},\ell\ell} > 30$ GeV
$\gamma\gamma \to W^+W^-$	128 (35.7)	122 (34.0)	116 (32.2)	108 (30.2)
$q\bar{q} \to W^+W^-$	17.4 (14.5)	16.3 (13.6)	14.4 (12.0)	12.7 (10.6)
DY $\ell^+\ell^-$	5.9	2.2	1.0	0.37
Significance	9.74σ	9.83σ	9.75σ	9.57

Table 17.1: Expected yields of the $\gamma\gamma \to W^+W^-$, $q\bar{q} \to W^+W^-$, and Drell-Yan processes in the signal region for various $p_{\mathrm{T},\ell\ell}$ cuts. The $\gamma\gamma \to W^+W^-$ yield is given with and without scaling by $S_{\mathrm{Excl}} = 3.59$, and the $q\bar{q} \to W^+W^-$ yield is given with and without a factor of 1.2, which seeks to account for additional backgrounds. The events must pass the exclusivity selection described in Section 17.2. The expected significances based on Eq. 17.1 are also presented for each $p_{\mathrm{T},\ell\ell}$ cut.

Similarly, we can change the p_{T} of the single leptons required for events. In the first round of the analysis, the leading lepton was required to have $p_{\mathrm{T}} > 27$ GeV and the subleading lepton was required to have $p_{\mathrm{T}} > 20$ GeV. The requirement on the leading lepton p_{T} is primarily motivated by the single lepton trigger (see Section 3.3.1). The requirement on the

quoted for the "optimized" significance after the track selection scan because Eq. 17.1 includes the Drell-Yan contribution directly, whereas Eq. 15.2 does not.

subleading lepton is mainly to suppress backgrounds and maintain a high efficiency relative to lepton identification and isolation requirements (see Sections 4.2 and 4.3). Assuming there is some lepton or other object with high enough energy to trigger event readout, ATLAS *does* reconstruct leptons below 20 GeV, so while we cannot easily lower the leading lepton p_{T} while using the single-lepton triggers, we can imagine lowering the subleading lepton p_{T}. It would be possible to lower the leading lepton p_{T} by using di-lepton triggers, however [216]. In simulated WW events, the leading lepton p_{T} peaks at 40 GeV, and the subleading lepton p_{T} peaks at 25 GeV. There are actually very few events where the leading lepton p_{T} is below 27 GeV, so it is not immediately obvious that lowering the p_{T} requirements will increase the significance. There are more events where the subleading lepton p_{T} is below 20 GeV, but this may be offset by an increase in the yield from fake leptons, though the fake lepton yield should be suppressed by the low-p_{T} tracks. A scan was performed in the two variables, and it was determined that the optimal working point was at $p_{\mathrm{T}} > 24$ GeV for the leading lepton and $p_{\mathrm{T}} > 16$ GeV for the subleading lepton if one- and two-lepton triggers are both used, but this has a relatively large uncertainty due to the possibility of fake leptons. In particular, the two-lepton triggers have much lower p_{T} requirements on the leading lepton, so more events will be triggered due to fake leptons. Rather than processing data and performing event picking with new triggers, it is easier to use the same single lepton triggers used in Part II, retaining the $p_{\mathrm{T}} > 27$ GeV cut for the leading lepton and reducing the subleading lepton requirement. Here, the significance is maximized when the subleading lepton requirement is reduced to 15 GeV. Both Herwig 7 and MadGraph 5 predict a 16% increase in the WW yield with this change. This brings about a 1σ increase in statistical significance relative to the yield with the old p_{T} and $p_{\mathrm{T},\ell\ell}$ cuts, though again, this carries a large uncertainty due to a potential increase in fakes.

Chapter 18

Expected Results

It was mentioned in Section 16.1 that the signal and control regions for the $\gamma\gamma \to W^+W^-$ re-analysis would be re-evaluated. For example, in Part II, CR2 in Part II was designed to isolate inclusive WW. It required that events have $1 \leq n_{\text{trk}} \leq 4$, where n_{trk} strictly counts high-p_{T} tracks. This definition could be changed to count both low-p_{T} and high-p_{T} tracks, or the two sets of tracks could be used separately. While a complete set of regions has not yet been finalized for the re-analysis, the following control regions are explored using the data and simulated samples described in Section 18.1:

- Z-Peak Region: $\mu^+\mu^-$ events with 70 GeV $< m_{\ell\ell} <$ 105 GeV and $0 \leq n_{\text{trk}} \leq 2$ (where n_{trk} refers to the number of high-p_{T} tracks). This region has little to do with our $\gamma\gamma \to W^+W^-$ signal, with the bulk of events coming from the DY $\mu^+\mu^-$ process, which is relatively well modelled in simulation. However, this control region will importantly allow us to compare the low-p_{T} tracks reconstructed in simulation to those in data for the first time. We can additionally look at the rate of events that pass and fail the event selections with the track-selection NN detailed in Section 17.2.

- High-$m_{\ell\ell}$ Region: $\mu^+\mu^-$ events with $m_{\ell\ell} >$ 160 GeV and $0 \leq n_{\text{trk}} \leq 2$ (where n_{trk} refers to the number of high-p_{T} tracks). While there should be a non-negligible contribution from DY $\mu^+\mu^-$, this region will be dominated by $\gamma\gamma \to \mu^+\mu^-$. Here, we can check the modelling of $\gamma\gamma \to \ell^+\ell^-$, and in particular, we will examine the modelling in dissociative LPAIR samples. We would like to investigate whether dissociative events can be distinguished from elastic events, which may be possible if proton dissociation generates low-p_{T} tracks. Similar to the Z-peak region, we can look at the rate of events that pass and fail the event selections detailed in Section 17.2.

- Old SR Background Region: $e^\pm\mu^\mp$ events that fell into the SR of Part II[1], which now fail the event selections detailed in Section 17.2. This region should be dominated by inclusive WW, and will provide a dramatic illustration of the impact of low-p_{T} tracks

[1] $p_{\text{T},\ell\ell} >$ 30 GeV, $n_{\text{trk}} = 0$

to the $\gamma\gamma \to W^+W^-$ analysis. Low-p_{T} tracks were introduced to strengthen our background rejection, and we will find that the majority of the background events in the SR of Part II fall into this new region.

All of these regions use the same general pre-selections given in Section 6.2.3.

While this analysis is still blinded, we can also look at the simulation-only yields in a new signal region that includes $e^{\pm}\mu^{\mp}$ events that pass the event selections detailed in Section 17.2[2] with $p_{\mathrm{T},\ell\ell} > 20$ GeV.

18.1 Available Data and Simulation with Low-p_{T} Tracks

About 50,000 data events have been processed based on the results of the analysis in Part II. All events that fell in the old signal region, CR2, and CR3 were reprocessed. As additional control regions, all $\mu^+\mu^-$ events with $m_{\ell\ell} > 115$ GeV and $0 \leq n_{\mathrm{trk}} \leq 2$, and all e^+e^- events with $m_{\ell\ell} > 160$ GeV and $0 \leq n_{\mathrm{trk}} \leq 2$ were reprocessed. Here, n_{trk} refers to the number of nominal, first-pass tracks within ± 1 mm of the lepton vertex. Finally, a Z-peak region corresponding to $\mu^+\mu^-$ events with 70 GeV $< m_{\ell\ell} <$ 105 GeV and $0 \leq n_{\mathrm{trk}} \leq 2$ was selected with a prescale factor of 50, meaning that 1 out of every 50 events would be processed. These selections are summarized in Table 18.1.

Region	Leptons	Definition	Scale Factor
Signal	$e^{\pm}\mu^{\mp}$	Presel., $p_{\mathrm{T},\ell\ell} > 30$ GeV, $n_{\mathrm{trk}} = 0$	1
CR2	$e^{\pm}\mu^{\mp}$	Presel., $p_{\mathrm{T},\ell\ell} > 30$ GeV, $1 \leq n_{\mathrm{trk}} \leq 4$	1
CR3	$e^{\pm}\mu^{\mp}$	Presel., $p_{\mathrm{T},\ell\ell} < 30$ GeV, $1 \leq n_{\mathrm{trk}} \leq 4$	1
Z-peak	$\mu^{\pm}\mu^{\mp}$	Presel., 70 GeV $< m_{\ell\ell} <$ 105 GeV, $0 \leq n_{\mathrm{trk}} \leq 2$	50
High-$m_{\ell\ell}$ $\mu\mu$	$\mu^{\pm}\mu^{\mp}$	Presel., $m_{\ell\ell} > 115$ GeV, $0 \leq n_{\mathrm{trk}} \leq 2$	1
High-$m_{\ell\ell}$ ee	$e^{\pm}e^{\mp}$	Presel., $m_{\ell\ell} > 160$ GeV, $0 \leq n_{\mathrm{trk}} \leq 2$	1

Table 18.1: Selections used for the currently available event-picked data. There are slightly fewer than 50,000 events in this samples total, with 8,800 $e\mu$ events and 41,000 $\ell\ell$ events. The n_{trk} selection here refers to the number of tight high-p_{T} tracks within ± 1 mm of the lepton vertex, or the standard selection from the previous round of the analysis. "Presel." here refers to the standard preselection discussed in Section 6.2.3, without any changes to the subleading lepton p_{T} requirements.

[2]They must have 0 low-p_{T} or high-p_{T} tracks that pass the track-selection NN cuts.

Similarly, a small set of simulated samples have been processed with low-p_{T} track reconstruction, which are given in Table 18.2. The samples include the bulk of the yield expected in the event-picked dataset[3].

Process	Generator
Elastic $\gamma\gamma \to W^+W^-$	Herwig 7
$q\bar{q} \to W^+W^-$	Powheg+Pythia 8
DY $\mu^+\mu^-$ with $m_{\ell\ell} > 60$ GeV	Powheg+Pythia 8
DY $\tau^+\tau^-$ (for all $m_{\ell\ell}$)	Powheg+Pythia 8
Double Diss. $\gamma\gamma \to \mu^+\mu^-$ with $m_{\ell\ell} > 60$ GeV	Pythia 8
Double Diss. $\gamma\gamma \to \tau^+\tau^-$ with $m_{\ell\ell} > 60$ GeV	Pythia 8
Single Diss. $\gamma\gamma \to \mu^+\mu^-$ with $m_{\ell\ell} > 60$ GeV	LPAIR
Elastic $\gamma\gamma \to \mu^+\mu^-$with $m_{\ell\ell} > 100$ GeV	Herwig 7

Table 18.2: Simulated samples that have been processed with low-p_{T} tracking. Most of these are the same samples presented in Tables 5.1 and 5.2. These samples simulate the bulk of the yield expected in the $\mu\mu$ and $e\mu$ events that make up the event-picked dataset.

18.2 Studies of Data with Low-p_{T} Tracks

18.2.1 Z-Peak Investigations

The first data vs. simulation comparisons that we make are for $\mu^+\mu^-$ events in the Z-peak region. This region has little to do with modelling of our signal and backgrounds, but allows us to investigate the modelling of low-p_{T} tracks. It is an important region for track modelling in this analysis, as it is where the underlying event (UE) correction is derived. The definition of this region is given in the chapter introduction and in Table 18.1.

The data and simulation agreement for lepton kinematic variables are shown in Figure 18.1. The vast majority of events in this region are accounted for with the DY $\mu^+\mu^-$ simulation, as should be expected. The yields, which are given in Table 18.3, agree to the 5% level between simulation and data. The shapes also appear to be well-modelled. All of the corrections discussed in Part II are applied to the simulation presented here, except that the S_{Excl} scale factor is not applied to the elastic $\gamma\gamma \to W^+W^-$ yield. The UE correction discussed in Chapter 8 *is* applied here. This correction is based on the number of truth-level charged particles with $p_{\mathrm{T}} > 500$ MeV and $|\eta| < 2.5$, which is a factor that is not affected by the reconstruction of low-p_{T} tracks or lack thereof. However, this correction does not account for low-p_{T} truth particles, which is why it will be rederived for the second round of the analysis.

[3]A full set of simulated samples will be processed eventually; see Section 18.4

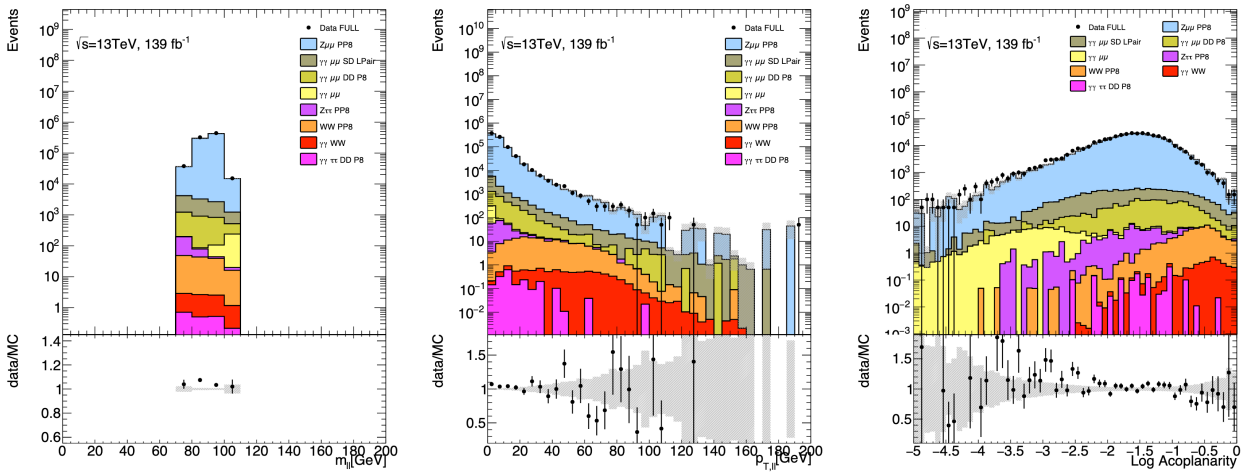


Figure 18.1: The data and simulation agreement for di-lepton $m_{\ell\ell}$ (left), $p_{\mathrm{T},\ell\ell}$ (middle), and acoplanarity (right) for $\mu\mu$ events in the Z-peak region. The natural log of the acoplanarity is plotted, such that the most back-to-back leptons are on the left-hand side of the plot. The data comes from the event picked data selected with the requirements of Table 18.1, and the simulation is that laid out in Table 18.2. The data in this region has a scale factor of 50 applied.

Sample	Yield
Elastic $\gamma\gamma \to W^+W^-$	7.27 ± 0.13
$q\bar{q} \to W^+W^-$	138.05 ± 2.80
DY $\mu^+\mu^-$	767840.40 ± 4210.77
DY $\tau^+\tau^-$	191.20 ± 9.70
DD $\gamma\gamma \to \mu^+\mu^-$	2852.86 ± 78.10
DD $\gamma\gamma \to \tau^+\tau^-$	1.94 ± 0.45
SD $\gamma\gamma \to \mu^+\mu^-$	7872.75 ± 72.63
Elastic $\gamma\gamma \to \mu^+\mu^-$	291.86 ± 3.14
Total	779196.33 ± 4212.13
Data	817900.00 ± 6394.92

Table 18.3: Yields of the event picked data (with scale factor of 50 applied) and simulation in the Z-peak region. The ratio of the simulation to data yield is 0.95. Uncertainties are statistical only.

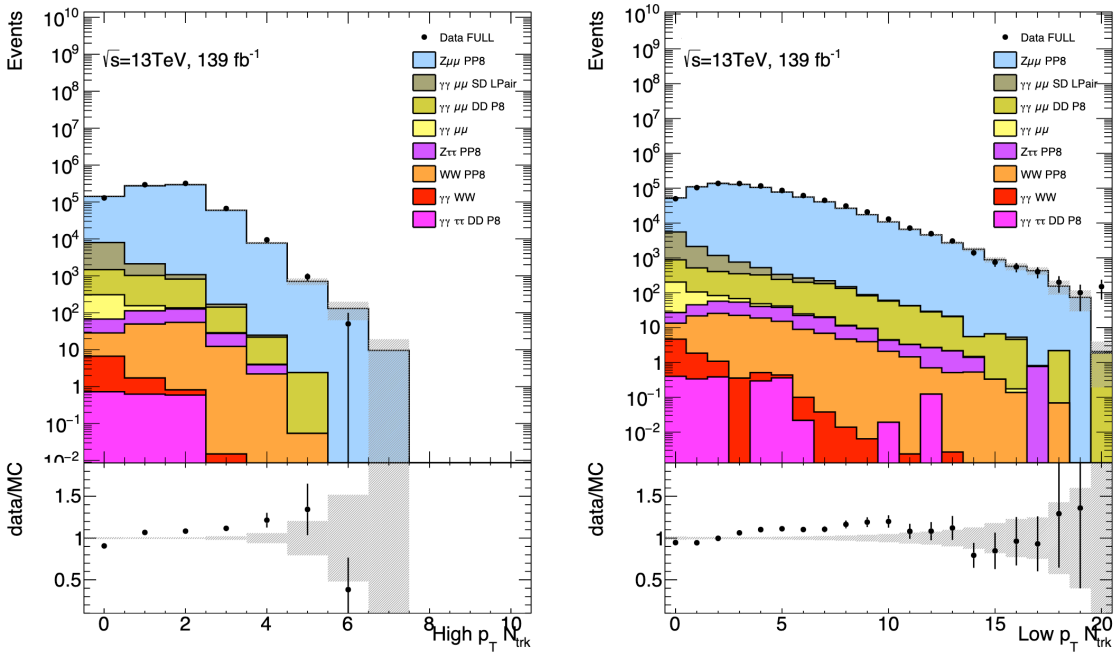


Figure 18.2: The data and simulation agreement for the number of high-p_{T} tracks (left) and low-p_{T} tracks (right) within ± 0.65 mm of the lepton vertex, selected with the neural network-based selection discussed in Section 4.1. While the Z-peak events were selected with $0 \leq n_{\mathrm{trk}} \leq 2$, this selection was based on the number of high-p_{T} tracks that pass the "tight" requirement. The neural network based selection is different than the tight one, which is how there are data events with more than 2 high-p_{T} tracks.

While the modelling of the lepton kinematics is expected to be correct, the modelling of the tracks, particularly of the low-p_{T} tracks, is interesting here. This represents the first-ever look at ATLAS data with low-p_{T} tracks reconstructed in nominal-pileup environments. Figure 18.2 shows the spectrum of the expected number high-p_{T} and low-p_{T} tracks within ± 0.65 mm of the lepton vertex that pass the lepton selections discussed in Section 4.1. It is promising to see here relatively good agreement between the data and simulation for most bins in both the high- and low-p_{T} cases. The UE correction of the first round is fairly effective, even for low-p_{T} tracks, suggesting that the modelling of low-p_{T} charged particles is correlated with the modelling of high-p_{T} charged particles.

Lastly, we can investigate the new event selections suggested in Section 17.2. The yields of events that pass and fail the version of the selection with the neural network-based track selection are shown in Figure 18.3. The majority of events should fail the selection, and the data agrees well with the simulation for such events. However, the simulation overpredicts the number of events that pass the selection by almost 20%, but a contributing factor here

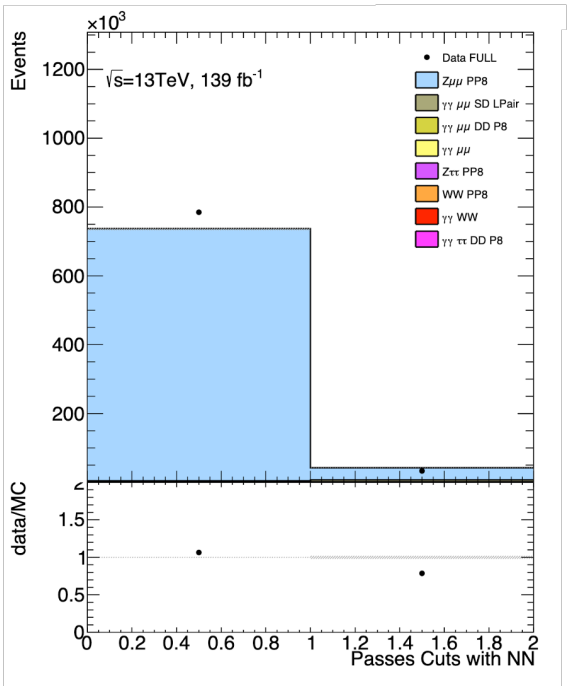


Figure 18.3: The data and simulation agreement for the number of events that pass and fail the event selection proposed in Section 17.2 using the neural network-based track selection. Because these events are dominated by the DY $\mu\mu$ process, the majority should fail the selection. The simulation overpredicts the number of events that pass the selection by almost 20%, but a contributing factor here is mismodelling of the $\gamma\gamma \to \mu^+\mu^-$ process, which will be discussed in Section 18.2.2.

is mismodelling of the $\gamma\gamma \to \mu^+\mu^-$ process, which will be discussed in Section 18.2.2.

18.2.2 High-$m_{\ell\ell}$ Investigations

The preceding subsection showed that the DY process dominates the Z-peak region as expected and that the overall modelling of this process was relatively good after the application of the corrections derived in Part II. We can now look at the high-$m_{\ell\ell}$ region to inspect the modelling of the $\gamma\gamma \to \ell^+\ell^-$ process. Here, we will look at $\mu\mu$ events with $m_{\ell\ell} > 160$ GeV, which is a region reminiscent of that used to derived the S_{Excl} scale factor. In the first round of the analysis, the yield of the $\gamma\gamma \to \ell^+\ell^-$ and $\gamma\gamma \to W^+W^-$ processes in regions with $n_{\text{trk}} = 0$ was estimated using the elastic-only simulation, with the S_{Excl} scale factor used to account for dissociative components. In the study presented in this subsection, the dissociative components will be estimated with the dedicated generators from Table 18.2.

Figure 18.4 shows the same lepton kinematics plots as Figure 18.1 did for the Z-peak region. We see generally good agreement here too, suggesting that the overall cross-section

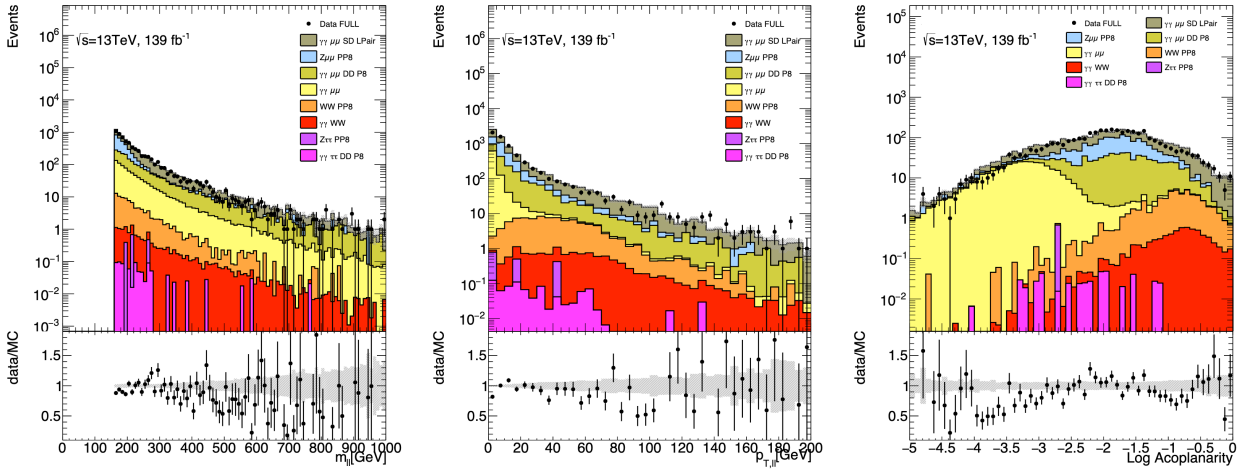


Figure 18.4: The data and simulation agreement for di-lepton $m_{\ell\ell}$ (left), $p_{\mathrm{T},\ell\ell}$ (middle), and acoplanarity (right) for $\mu\mu$ events with $m_{\ell\ell} > 160$ GeV. The natural log of the acoplanarity is plotted, such that the most back-to-back leptons are on the left-hand side of the plot. The data comes from the event picked data selected with the requirements of Table 18.1, and the simulation is that laid out in Table 18.2.

expectation from the simulation is accurate to within 10%. The shape estimates for the kinematic spectra are also generally accurate, though there is some mismodelling in the acoplanarity spectrum. The yields in this region are detailed in Table 18.4.

Figure 18.5 shows the spectra of the numbers of high- and low-p_{T} tracks within ± 0.65 mm of the lepton vertex, when using the neural network-based track selection. It is clear here that these spectra are not well modelled, with the simulation dramatically overestimating the yield with 0 tracks. This has particularly dramatic consequences when considering the exclusivity selection, which is based on track counting. The yields of events that pass the exclusivity selection are given in Table 18.4 and are visualized in Figure 18.6.

The simulation overpredicts the yield of events passing the exclusivity selection by 60%, which seems to be mostly due to the LPAIR single-dissociative $\gamma\gamma \to \ell^+\ell^-$ generator. If a scale factor is used to account for dissociative events, as in the first round of analysis, this mismodelling is not extremely consequential. However, the use of the scale factor contributed large uncertainties to the analysis, which we would like to avoid. The mismodelling in LPAIR is a contributing factor for why we would like to investigate the SuperChic generator [19] for $\gamma\gamma \to \ell^+\ell^-$ events.

18.2.3 Blinded Investigation of $e\mu$ Events

The investigation in this section most resembling the old signal region uses $e\mu$ data. The version of the analysis with low-p_{T} tracks is still blinded in order to not bias ourselves by

Sample	Total Yield	Yield Passing Exclusivity
Elastic $\gamma\gamma \to W^+W^-$	11.72 ± 0.17	8.89 ± 0.15
$q\bar{q} \to W^+W^-$	85.35 ± 2.16	3.94 ± 0.42
DY $\mu^+\mu^-$	1911.36 ± 33.10	89.91 ± 6.37
DY $\tau^+\tau^-$	1.43 ± 0.86	0.00 ± 0.00
DD $\gamma\gamma \to \mu^+\mu^-$	1318.69 ± 34.71	391.25 ± 18.83
DD $\gamma\gamma \to \tau^+\tau^-$	0.88 ± 0.15	0.09 ± 0.05
SD $\gamma\gamma \to \mu^+\mu^-$	2580.24 ± 28.93	1942.18 ± 25.00
Elastic $\gamma\gamma \to \mu^+\mu^-$	807.53 ± 3.25	604.43 ± 2.79
Total	$6,717.2 \pm 56.15$	3040.69 ± 32.06
Data	6239.00 ± 78.99	1881.00 ± 43.37

Table 18.4: Yields of event picked data and simulation for $\mu\mu$ events with $m_{\ell\ell} > 160$ GeV. The ratio of the simulation to data yield is 1.08. While is is an overall agreement of better than 10%, the ratio of simulation to data for events that pass the exclusivity selection is 1.62. (Visualized in Figure 18.5)

looking at events that we expect to be signal. However, we will look at some events from our old signal region here. In Figure 13.8, the left-most bin, which requires 0 high-p_{T} tracks within ± 1 mm of the leptons shows about a 1.3:1 signal-to-background ratio. The main purpose of using low-p_{T} tracks is to reduce the background contamination, so we can look at events from the old signal region that *fail* the new exclusivity selection using low-p_{T} tracks. These events should be mostly background, so this is not an unblinding of the analysis.

Figure 18.7, shows the lepton kinematics for these events. They must have 0 tight, high-p_{T} tracks within ± 1 mm of the lepton vertex *and* fail the exclusivity selection detailed in Section 17.2 using the neural network for track selection. There are 85 data events that fall into this region, which leads to relatively large uncertainties in each bin, but the overall cross-section and shape modelling seems fairly accurate, with exact yields given in Table 18.5. With the current samples, the simulation underpredicts the data by about 25%, though this does not account for dissociative $\gamma\gamma \to W^+W^-$ events. If the $S_{\mathrm{Excl}} = 3.59$ scale factor is used, the simulation-to-data ratio is 87%. However, it is not clear whether that scale factor is applicable here, as the modelling of the low-p_{T} track spectrum in the elastic $\gamma\gamma \to W^+W^-$ sample has not been studied in detail. This comparison also does not include an estimation of events with fake leptons, events from tops, or events with gluon-induced WW. It also does not include the $\beta_{q\bar{q}\to WW}$ and β_{DY} values from the fit in Part II.

It is encouraging to see that this region is dominated by $q\bar{q} \to W^+W^-$ events, which is what we hoped for by introducing low-p_{T} tracks. Indeed, Figure 18.8 shows the number of low-p_{T} tracks near the lepton vertex. Most of the events have 0 high-p_{T} tracks within ± 0.65 mm of the leptons but are eliminated by having a non-zero number of low-p_{T} tracks. The yields presented here suggest that there would be fewer than 50 background events

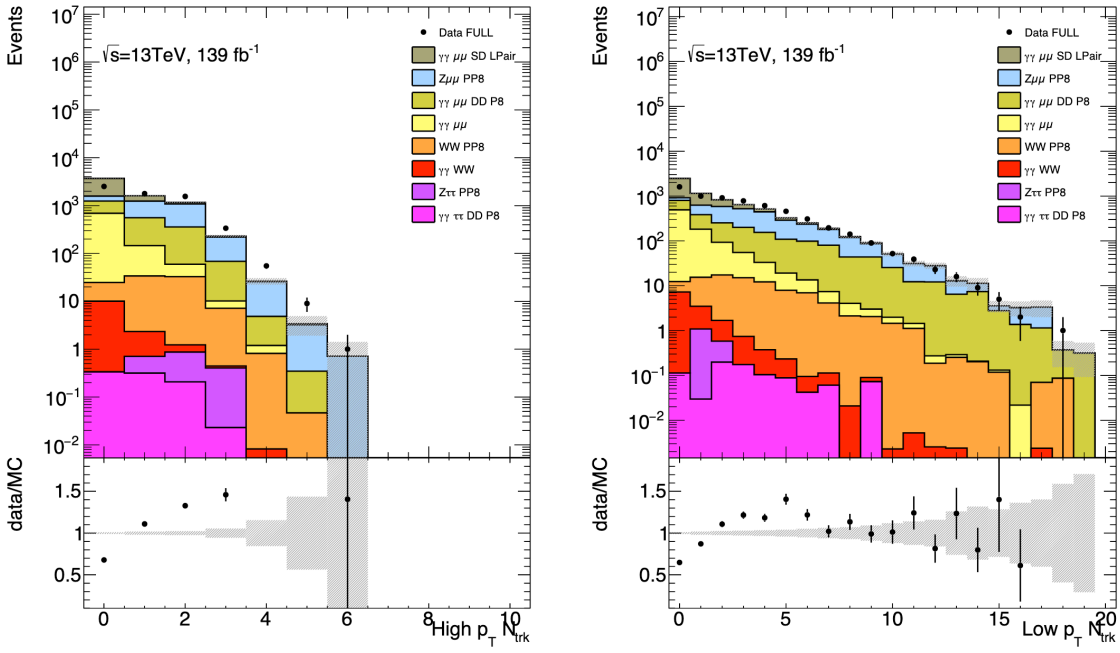


Figure 18.5: The data and simulation agreement for the number of high-p_{T} tracks (left) and low-p_{T} tracks (right) within ± 0.65 mm of the lepton vertex, selected with the neural network-based selection discussed in Section 4.1. These are $\mu\mu$ events with $m_{\ell\ell} > 160$ GeV, selected with $0 \leq n_{\mathrm{trk}} \leq 2$, but this selection was based on the number of high-p_{T} tracks that pass the "tight" requirement. The neural network based selection is different than the tight one, which is how there are data events with more than 2 high-p_{T} tracks.

remaining in the signal region, though this makes no attempts to account for the migration of fake events or the scaling of the inclusive WW cross-section observed in the fit. It also does not account for events that might migrate from the old control regions into the signal region, which might occur due to the reduced exclusivity window size.

18.3 Expected Results

While the analysis with low-p_{T} tracks is still blinded, we can look at a potential signal region in simulation. As explained in the introduction to this chapter, this region includes $e^{\pm}\mu^{\mp}$ events that have 0 low-p_{T} or high-p_{T} tracks that pass the track-selection NN cuts and have $p_{\mathrm{T},\ell\ell} > 20$ GeV. Plots of the lepton kinematics in these events are given in Table 18.9. These plots include only elastic $\gamma\gamma \to W^+W^-$ events, as S_{Excl} has not been applied. However, even without the dissociative components, the signal is expected to dominate. The expected yields are given explicitly in Table 18.6. In this table, we see that the expected signal to

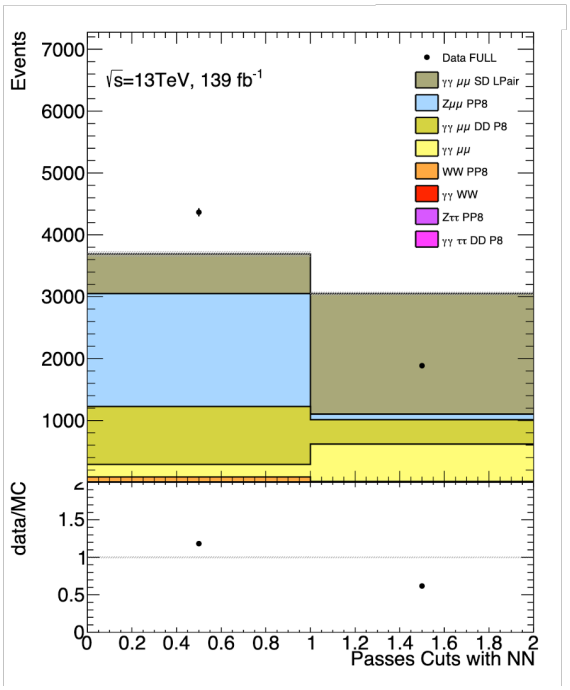


Figure 18.6: The data and simulation agreement for the number of events that pass and fail the event selection proposed in Section 17.2 using the neural network-based track selection. These are $\mu\mu$ events with $m_{\ell\ell} > 160$ GeV, which are dominated by $\gamma\gamma \to \mu^+\mu^-$ events, particularly, single-dissociative events. The simulation overpredicts the number of events with 0 tracks, and therefore, overpredicts the number of events that will pass the exclusvity selection.

background ratio is 7:1 after the application of the S_{Excl} scale factor from Part II. It is not immediately obvious that this scale factor is valid in the presence of low-p_{T} tracks. However, this table also does not include the signal-strength factor of $\mu = 1.33$ that was found in Part II. Similarly, the backgrounds are missing contributions from events with fake leptons, events from tops, or events with gluon-induced WW. It also does not include the $\beta_{q\bar{q}\to WW}$ and β_{DY} values from the fit in Part II.

The contribution of additional backgrounds to this signal region should be small, as most backgrounds involve QCD-induced events that should be reduced due to low-p_{T} tracks. Using Eq. 17.1, a conservative expected significance of this signal region is 9.8σ. When compared to the expectation of the old signal region using the same metric, this represents an increase in significance of 3.7σ.

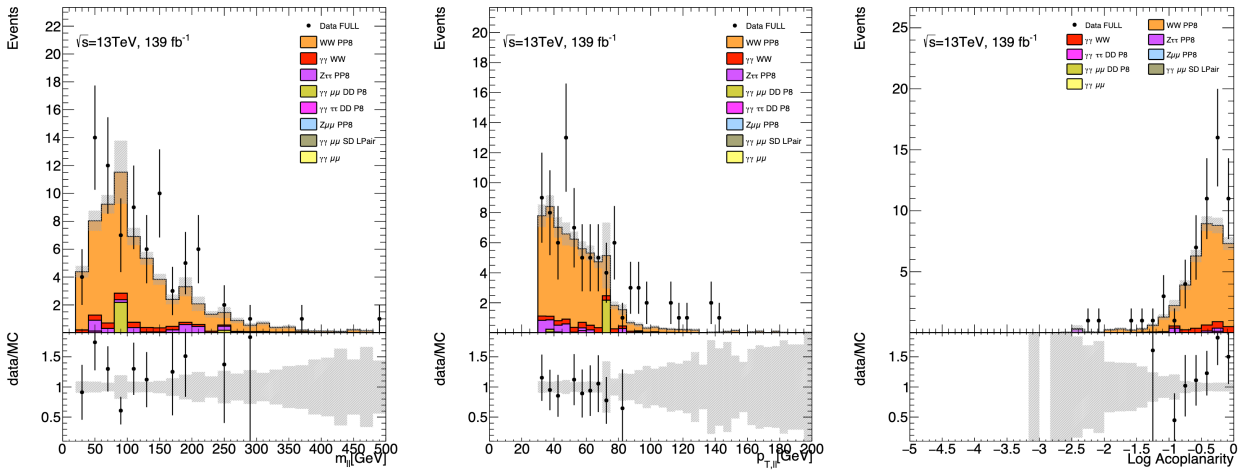


Figure 18.7: The data and simulation agreement for di-lepton $m_{\ell\ell}$ (left), $p_{\mathrm{T},\ell\ell}$ (middle), and acoplanarity (right) in former signal region events that now fail the newly suggested exclusivity selection using low-p_{T} tracks. The data comes from the event picked data selected with the requirements of Table 18.1, and the simulation is that laid out in Table 18.2.

Sample	Yield
Elastic $\gamma\gamma \to W^+W^-$	3.76 ± 0.09 (13.50 ± 0.36)
$q\bar{q} \to W^+W^-$	54.43 ± 1.52
DY $\mu^+\mu^-$	0.00 ± 0.00
DY $\tau^+\tau^-$	3.18 ± 1.09
DD $\gamma\gamma \to \mu^+\mu^-$	2.36 ± 2.18
DD $\gamma\gamma \to \tau^+\tau^-$	0.38 ± 0.15
SD $\gamma\gamma \to \mu^+\mu^-$	0.00 ± 0.00
Elastic $\gamma\gamma \to \mu^+\mu^-$	0.00 ± 0.00
Total	64.11 ± 2.88 (73.85 ± 2.90)
Data	85.00 ± 9.22

Table 18.5: Yields of the event picked data and simulation for events that were in the signal region from Part II, but that fail the newly suggested exclusivity selection in the presence of low-p_{T} tracks (the prediction with S_{Excl} applied to the $\gamma\gamma \to W^+W^-$ sample is given in parentheses). The ratio of the simulation to data yield is 0.75, though this does not include an estimate for single- or double-dissociative $\gamma\gamma \to W^+W^-$ events. If the elastic-only yield in the table above is multiplied by a factor of $S_{\mathrm{Excl}} = 3.59$, then the simulation to background ration "improves" to 0.87, though it is not clear that that scale factor is applicable here.

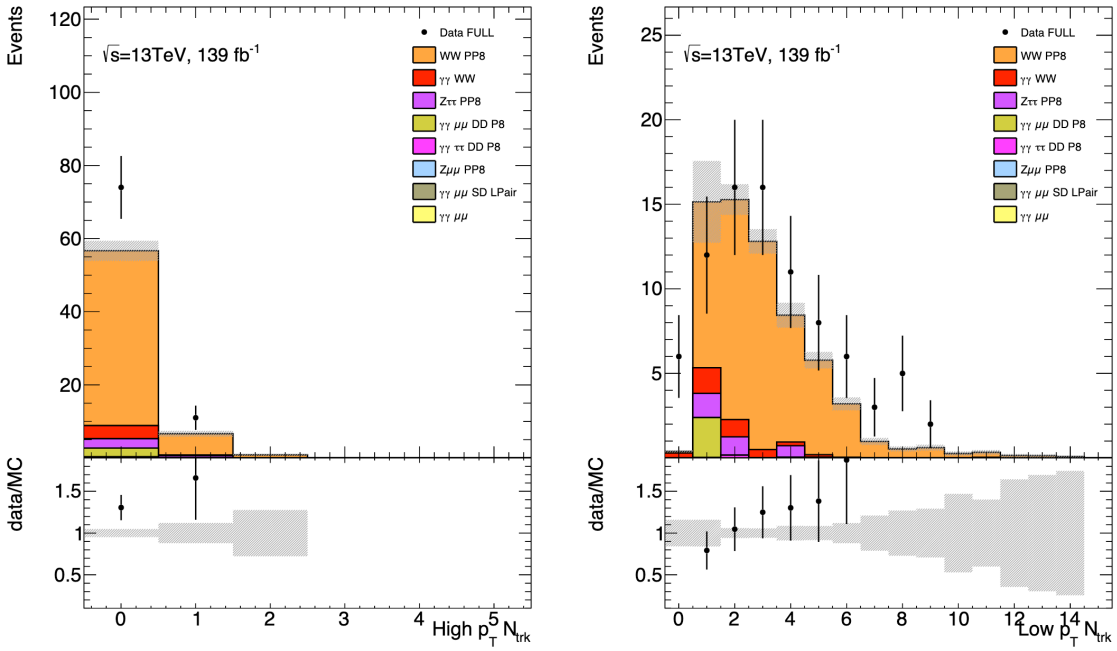


Figure 18.8: The data and simulation agreement for the number of high-p_{T} tracks (left) and low-p_{T} tracks (right) within ± 0.65 mm of the lepton vertex, selected with the neural network-based selection discussed in Section 4.1. These events were in the signal region from Part II but fail the newly suggested exclusivity selection in the presence of low-p_{T} tracks. This selection requires 0 high-p_{T} *and* low-p_{T} tracks, so it is possible for an event to pass with either 0 low-p_{T} tracks or 0 high-p_{T} tracks, but not both. The fact that there are some events with more than 0 high-p_{T} tracks is due to the fact that the neural network based track selection is used, which is different than the tight one required in the past.

18.4 Outstanding Issues for the $\gamma\gamma \to W^+W^-$ Re-Analysis

While Part III of this thesis did not present as complete an analysis as Part II, the results are highly promising. A few items remain for the analysis team:

1. A complete set of events for reprocessing has not yet been finalized. For example, events with subleading lepton p_{T} below 20 GeV will be included in the final analysis. This will lead to a slight change in the expected results from Section 18.3.

2. A complete set of simulated samples has also not been finalized. Whereas event picking reduced the storage space required for data, we still want large simulated samples to reduce the statistical uncertainties associated with our predictions. However, the vast

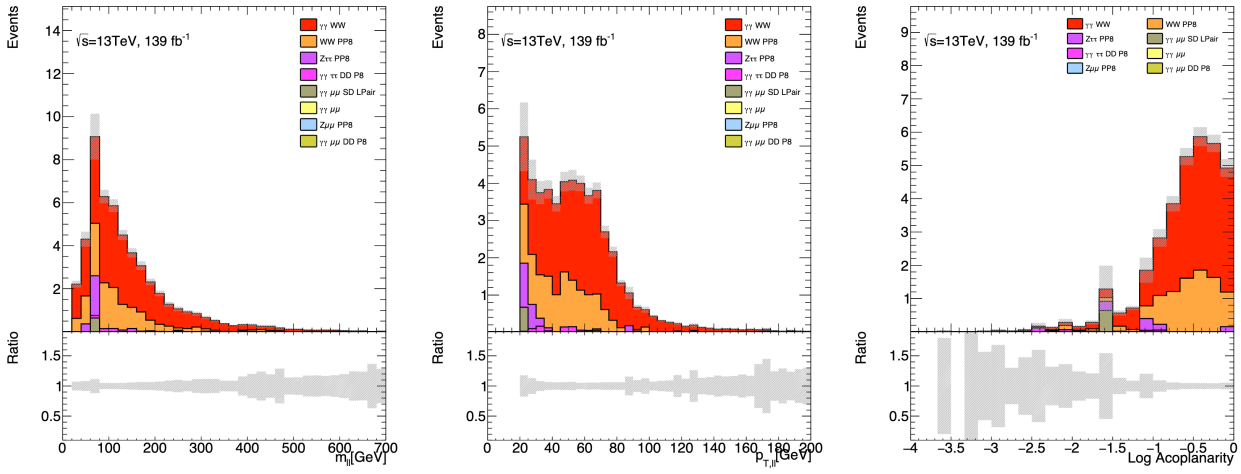


Figure 18.9: The simulated di-lepton $m_{\ell\ell}$ (left), $p_{\mathrm{T},\ell\ell}$ (middle), and acoplanarity (right) in a new signal region with low-p_{T} tracks. The simulated samples included are those in Table 18.2.

Sample	Yield
Elastic $\gamma\gamma \to W^+W^-$	33.99 ± 0.28 (122.02 ± 1.02)
$q\bar{q} \to W^+W^-$	13.57 ± 0.72
DY $\mu^+\mu^-$	0.00 ± 0.00
DY $\tau^+\tau^-$	2.21 ± 0.83
DD $\gamma\gamma \to \mu^+\mu^-$	0.00 ± 0.00
DD $\gamma\gamma \to \tau^+\tau^-$	0.80 ± 0.20
SD $\gamma\gamma \to \mu^+\mu^-$	0.70 ± 0.65
Elastic $\gamma\gamma \to \mu^+\mu^-$	0.08 ± 0.03
Total	51.35 ± 1.32 (139.38 ± 1.65)

Table 18.6: Yields of the simulation for events that fall into the new signal region, which must have 0 high-p_{T} and 0 low-p_{T} tracks that pass the neural network based track selection cuts. The prediction with $S_{\mathrm{Excl}} = 3.59$ applied to the $\gamma\gamma \to W^+W^-$ sample is given in parentheses.

majority of QCD-induced background events are not relevant in this analysis as they have many tracks. Simulation filters will be applied to restrict the number of charged particles allowed in simulated events that will be processed with low-p_{T} tracks.

3. The corrections explained in Chapters 7-10 must be rederived in to the presence of low-p_{T} tracks. The pileup correction can be derived in the same way as laid out in Chapter 7, and the underlying event (UE) correction can be derived in a similar way to that presented in Chapter 8. A set of data events must be selected and re-processed in order to extract a fake factor as in Chapter 9. Lastly, it is also not clear if the scale factor method should be used to account for the dissociative component of the signal. This method brought $\ell\ell$ to WW transfer uncertainty into the signal strength measurement in Part II, which was one of the largest in the analysis. The newest release of the SuperChic generator can produce the elastic $\gamma\gamma \to W^+W^-$ process, and it can produce elastic, single-dissociative (SD), and double-dissociative (DD) $\gamma\gamma \to \ell^+\ell^-$ events. Previously, LPAIR was the only generator that could perform all of elastic, SD, and DD $\gamma\gamma \to \ell^+\ell^-$ events, but it could not produce $\gamma\gamma \to W^+W^-$ events.

4. The analysis in Part II was designed to make a statistically significant observation of $\gamma\gamma \to W^+W^-$ and to measure its cross-section. We would also like to perform differential cross-section measurements, which is a measurement of the cross-section as a function of kinematic variables, such as $m_{\ell\ell}$ and $p_{\mathrm{T},\ell\ell}$. Assuming no statistically significant deviations from the Standard Model turn up, we would also like to place constraints on anomalous quartic and triple gauge couplings (aQGC and aTGC) through an effective field theory (EFT) framework. The differential cross-section measurements in particular will aid placing constraints, as some potential modifications to the Standard Model would cause $m_{\ell\ell}$- or $p_{\mathrm{T},\ell\ell}$-dependent enhancements[4].

[4]See Section 2.4.1

Chapter 19

Part III Conclusions

The introduction of low-p_{T} tracks to the $\gamma\gamma \to W^+W^-$ analysis represents a powerful means to suppress background processes. The key signal vs. background discriminant for photon-induced processes is the presence of underlying event tracks, and few background events that had no reconstructed tracks with $p_{\mathrm{T}} > 500$ MeV will also have no reconstructed tracks below that threshold.

Part III of this thesis detailed the implementation of a new algorithm for the reconstruction of low-p_{T} tracks. This algorithm can avoid some of the computational slow-down associated with high-pileup environments at the LHC through its two-pass structure and the possibility of using a region of interest for the second tracking pass. The re-analysis will also circumvent storage limitations by using a novel data-flow to ATLAS which permits a small list of events to be reprocessed with unique algorithms.

A new system of track selections was devised to reduce the expected fake rate of low-p_{T} tracks. These selections help ensure that only high-quality tracks are considered in the analysis. Furthermore, the exclusivity selection was re-evaluated. Previously, events were rejected if any tight track was within ± 1 mm of the lepton vertex. While more underlying event charged particles are introduced with low-p_{T} tracking, more pileup particles are reconstructed too, leading to a reduced window size and some restrictions on which low-p_{T} tracks are considered.

Initial studies of low-p_{T} tracking in data look promising, even when using the simulation corrections derived in Part II. Though these corrections will have to be rederived, we expect their effect to be about the same as that observed in the first round of the analysis. In particular, after applying the underlying event correction described in Chapter 8, the spectrum of the number of low-p_{T} tracks counted near the lepton vertices appears to be well modelled for both DY and $q\bar{q} \to W^+W^-$events.

With the introduction of low-p_{T} tracks, we expect a new signal region to retain about 90% of signal events and less than 20% of background events compared to the analysis presented in Part II. From studies performed in simulation alone, we hope to improve the statistical significance relative to the first round of the analysis by about 4σ.

Bibliography

[1] Murray Gell-Mann. "Symmetries of Baryons and Mesons". In: *Phys. Rev.* 125 (3 1962), pp. 1067–1084. DOI: 10.1103/PhysRev.125.1067.

[2] P.A. Zyla et al. (Particle Data Group). In: *Prog. Theor. Exp. Phys.* 2020.8 (2020). 083C01. DOI: 10.1093/ptep/ptaa104.

[3] E. Fermi. "On the Theory of Collisions between Atoms and Electrically Charged Particles". In: *Nuovo Cimento* 2 (1925), pp. 143–158.

[4] C. F. v. Weizsäcker. "Ausstrahlung bei Stößen sehr schneller Elektronen". In: *Zeitschrift fur Physik* 88 (1934), pp. 612–625. DOI: 10.1007/BF01333110.

[5] E. J. Williams. "Nature of the High Energy Particles of Penetrating Radiation and Status of Ionization and Radiation Formulae". In: *Phys. Rev.* 45 (10 1934), pp. 729–730. DOI: 10.1103/PhysRev.45.729.

[6] 2007. URL: https://commons.wikimedia.org/wiki/File:Onde%5C_electromagnetique.svg.

[7] Carlos A. Bertulani. "Photon exchange in nucleus–nucleus collisions". In: *Int. Journ. of Mod. Phys. A* 18.05 (2003), pp. 685–723. DOI: 10.1142/s0217751x03012357. arXiv: nucl-th/0209036.

[8] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin. "The photon PDF in events with rapidity gaps". In: *EPJC* 76.5 (2016). DOI: 10.1140/epjc/s10052-016-4100-2. arXiv: 1601.03772 [hep-ph].

[9] V A Khoze, A D Martin, and M G Ryskin. "Multiple interactions and rapidity gap survival". In: *J. Phys. G* 45.5 (2018), p. 053002. DOI: 10.1088/1361-6471/aab1bf. arXiv: 1710.11505 [hep-ph].

[10] H. Spiesberger. "QED radiative corrections for parton distributions". In: *Phy. Rev. D* 52.9 (1995), pp. 4936–4940. DOI: 10.1103/physrevd.52.4936. arXiv: hep-ph/9412286.

[11] A. D. Martin et al. "Parton distributions incorporating QED contributions". In: *The European Physical Journal C* 39.2 (2005), pp. 155–161. DOI: 10.1140/epjc/s2004-02088-7. arXiv: hep-ph/0411040.

[12] Richard D. Ball et al. "Parton distributions for the LHC run II". In: *Journal of High Energy Physics* 2015.4 (2015). DOI: 10.1007/jhep04(2015)040. arXiv: 1410.8849 [hep-ph].

[13] Valerio Bertone et al. "Illuminating the photon content of the proton within a global PDF analysis". In: *SciPost Physics* 5.1 (2018). DOI: 10.21468/scipostphys.5.1.008. arXiv: 1712.07053 [hep-ph].

[14] G. Altarelli and G. Parisi. "Asymptotic freedom in parton language". In: *Nuclear Physics B* 126.2 (1977), pp. 298–318. DOI: https://doi.org/10.1016/0550-3213(77)90384-4.

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

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

[17] Mateusz Dyndal and Laurent Schoeffel. "The role of finite-size effects on the spectrum of equivalent photons in proton–proton collisions at the LHC". In: *Phys. Lett. B* 741 (2015), pp. 66–70. DOI: 10.1016/j.physletb.2014.12.019. arXiv: 1410.2983 [hep-ph].

[18] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin. "Exclusive physics at the LHC with SuperChic 2". In: *Eur. Phys. J. C* 76.1 (2016), p. 9. DOI: 10.1140/epjc/s10052-015-3832-8.

[19] L. A. Harland-Lang et al. "A new approach to modelling elastic and inelastic photon-initiated production at the LHC: SuperChic 4". In: *Eur. Phys. J. C* 80.10 (2020), p. 925. DOI: 10.1140/epjc/s10052-020-08455-0.

[20] V. A. Khoze, A. D. Martin, and M. G. Ryskin. "Diffraction at the LHC". In: *Eur. Phys. J. C* 73.7 (2013). DOI: 10.1140/epjc/s10052-013-2503-x. arXiv: 1306.2149 [hep-ph].

[21] The ATLAS Collaboration. "Observation and Measurement of Forward Proton Scattering in Association with Lepton Pairs Produced via the Photon Fusion Mechanism at ATLAS". In: *Phys. Rev. Lett.* 125.26 (2020). DOI: 10.1103/physrevlett.125.261801. arXiv: 2009.14537 [hep-ex].

[22] ATLAS Collaboration. *Summary Plots, 2018*. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/index.html.

[23] R. Bauer et al. "Measurement of electromagnetically produced e+ e- pairs in distant S - Pt collisions". In: *Phys. Lett. B* 332 (1994), pp. 471–476. DOI: 10.1016/0370-2693(94)91283-1.

[24] C. R. Vane et al. "Electron-positron pair production in Coulomb collisions by 6.4-TeV sulfur ions". In: *Phys. Rev. A* 50 (3 1994), pp. 2313–2321. DOI: 10.1103/PhysRevA.50.2313.

[25] A. Belkacem et al. "Capture, ionization, and pair-production processes in relativistic heavy-ion collisions in the 1-GeV/nucleon energy range". In: *Phys. Rev. A* 56 (4 1997), pp. 2806–2818. DOI: 10.1103/PhysRevA.56.2806.

[26] "Production of e^+e^- pairs accompanied by nuclear dissociation in ultraperipheral heavy-ion collisions". In: *Phys. Rev. C* 70 (3 2004), p. 031902. DOI: 10.1103/PhysRevC.70.031902.

[27] A. Aktas et al. "Multielectron production at high transverse momenta in ep collisions at HERA". In: *Eur. Phys. J. C* 31 (2003), pp. 17–29. DOI: 10.1140/epjc/s2003-01326-x. arXiv: hep-ex/0307015.

[28] David M. South. "Searches for new physics at HERA". In: *2nd Workshop on Hadron Structure and QCD: From Low to High Energy*. Feb. 2006. arXiv: hep-ex/0602028.

[29] A. Aktas et al. "Muon pair production in ep collisions at HERA". In: *Phys. Lett. B* 583 (2004), pp. 28–40. DOI: 10.1016/j.physletb.2004.01.003. arXiv: hep-ex/0311015.

[30] A. Abulencia et al. "Observation of Exclusive Electron-Positron Production in Hadron-Hadron Collisions". In: *Phys. Rev. Lett.* 98.11 (2007). DOI: 10.1103/physrevlett.98.112001. arXiv: arXiv:hep-ex/0611040.

[31] T. Aaltonen et al. "Search for Exclusive Z-Boson Production and Observation of High-Mass $p\bar{p} \to p\gamma\gamma\bar{p} \to pl^+l^-\bar{p}$ Events in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. Lett.* 102.22 (2009). DOI: 10.1103/physrevlett.102.222002. arXiv: 0902.2816 [hep-ex].

[32] T. Aaltonen et al. "Observation of Exclusive Charmonium Production and $\gamma\gamma \to \mu^+\mu^-$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. Lett.* 102.24 (2009). DOI: 10.1103/physrevlett.102.242001. arXiv: 0902.1271 [hep-ex].

[33] ATLAS Collaboration. "Measurement of exclusive $\gamma\gamma \to l^+l^-$ production in proton–proton collisions at $\sqrt{s}$ 7 TeV with the ATLAS detector". In: *Phys. Lett. B* 749 (2015), pp. 242–261. DOI: 10.1016/j.physletb.2015.07.069. arXiv: 1506.07098 [hep-ex].

[34] ATLAS Collaboration. "Measurement of the exclusive $\gamma\gamma \to \mu^+\mu^-$ process in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett. B* 777 (2018), pp. 303–323. DOI: 10.1016/j.physletb.2017.12.043. arXiv: 1708.04053 [hep-ex].

[35] ATLAS Collaboration. "Observation of Centrality-Dependent Acoplanarity for Muon Pairs Produced via Two-Photon Scattering in Pb+Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV with the ATLAS Detector". In: *Phy. Rev. Lett.* 121.21 (2018). DOI: 10.1103/physrevlett.121.212301. arXiv: 1806.08708 [nucl-ex].

[36] CMS Collaboration. "Observation of proton-tagged, central (semi)exclusive production of high-mass lepton pairs in pp collisions at 13 TeV with the CMS-TOTEM precision proton spectrometer". In: *JHEP* 2018.7 (2018). DOI: 10.1007/jhep07(2018)153. arXiv: 1803.04496 [hep-ex].

[37] CMS Collaboration. "Exclusive $\gamma\gamma \to \mu^+\mu^-$ production in proton-proton collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 2012.1 (2012). DOI: 10.1007/jhep01(2012)052. arXiv: 1111.5536 [hep-ex].

[38] ATLAS Collaboration. "Observation of Light-by-Light Scattering in Ultraperipheral Pb+Pb Collisions with the ATLAS Detector". In: *Phys. Rev. Lett.* 123.5 (2019). DOI: 10.1103/physrevlett.123.052001. arXiv: 1904.03536 [hep-ex].

[39] ATLAS Collaboration. *Measurement of light-by-light scattering and search for axion-like particles with 2.2 nb^{-1} of Pb+Pb data with the ATLAS detector*. 2020. arXiv: 2008.05355 [hep-ex].

[40] CMS Collaboration. "Evidence for light-by-light scattering and searches for axion-like particles in ultraperipheral PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV". In: *Phys. Lett. B* 797 (2019), p. 134826. DOI: 10.1016/j.physletb.2019.134826. arXiv: 1810.04602 [hep-ex].

[41] ATLAS Collaboration. "Measurement of exclusive $\gamma\gamma \to W^+W^-$ production and search for exclusive Higgs boson production in *pp* collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector". In: *Phys. Rev. D* 94 (3 2016), p. 032011. DOI: 10.1103/PhysRevD.94.032011. arXiv: 1607.03745 [hep-ex].

[42] CMS Collaboration. "Evidence for exclusive $\gamma\gamma \to W^+W^-$ production and constraints on anomalous quartic gauge couplings in pp collisions at $\sqrt{s} = 7$ TeVand 8 TeV". In: *JHEP* 2016.8 (2016). DOI: 10.1007/JHEP08(2016)119. arXiv: 1604.04464 [hep-ex].

[43] CMS Collaboration. "Study of exclusive two-photon production of W^+W^- in pp collisions at $\sqrt{s} = 7$ TeV and constraints on anomalous quartic gauge couplings". In: *JHEP* 2013.7 (2013). DOI: 10.1007/jhep07(2013)116. arXiv: 1305.5596 [hep-ex].

[44] Gianfranco Bertone, Dan Hooper, and Joseph Silk. "Particle dark matter: evidence, candidates and constraints". In: *Physics Reports* 405.5-6 (2005), pp. 279–390. DOI: 10.1016/j.physrep.2004.08.031. arXiv: hep-ph/0404175.

[45] P. A. R. Ade et al. "Planck 2015 results. XIII. Cosmological Parameters". In: *Astronomy & Astrophysics* 594 (2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[46] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". In: *JETP Lett.* 5 (1967), pp. 24–27. DOI: 10.1070/PU1991v034n05ABEH-002497.

[47] Antonio Riotto and Mark Trodden. "Recent progress in baryogenesis". In: *Ann. Rev. Nucl. Part. Sci.* 49.1 (1999), pp. 35–75. DOI: 10.1146/annurev.nucl.49.1.35. arXiv: hep-ph/9901362.

[48] S. Bilenky. "Neutrino oscillations: From a historical perspective to the present status". In: *Nucl. Phys. B* 908 (2016), pp. 2–13. DOI: 10.1016/j.nuclphysb.2016.01.025. arXiv: 1602.00170 [hep-ph].

[49] J. Schechter and J. W. F Valle. "Neutrino masses in SU(2) x U(1) theories". In: *Phys. Rev. D* 22 (1980), p. 2227. DOI: 10.1103/PhysRevD.22.2227.

[50] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, Mar. 2014. ISBN: 978-1-107-03473-0.

[51] Stephen P. Martin. "A Supersymmetry Primer". In: *Advanced Series on Directions in High Energy Physics* (1998), pp. 1–98. DOI: 10.1142/9789812839657_0001. arXiv: hep-ph/9709356.

[52] Apostolos Pilaftis. "Heavy Majorana Neutrinos and Baryogenesis". In: *Int. J. Mod. Phys. A* 14.12 (1999), pp. 1811–1857. DOI: 10.1142/s0217751x99000932. arXiv: hep-ph/9812256.

[53] R.D. Peccei and Helen Quinn. "CP Conservation in the Presence of Pseudoparticles". In: *Phys. Rev. Lett.* 38 (1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440.

[54] J. Polchinski. *String theory. Vol. 1: An introduction to the bosonic string*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Dec. 2007. DOI: 10.1017/CBO9780511816079.

[55] Carlo Rovelli and Francesca Vidotto. *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Nov. 2014.

[56] Genessis Perez, Marco Sekulla, and Dieter Zeppenfeld. "Anomalous quartic gauge couplings and unitarization for the vector boson scattering process $pp \to W^+W^-jjX \to l^+\nu_l l^+\nu_l jjX$". In: *EPJC* 78.9 (2018). DOI: 10.1140/epjc/s10052-018-6230-1. arXiv: 1807.02707 [hep-ph].

[57] Celine Degrande et al. *Monte Carlo tools for studies of non-standard electroweak gauge boson interactions in multi-boson processes: A Snowmass White Paper*. 2013. arXiv: 1309.7890 [hep-ph].

[58] Jacob J. Ethier et al. *SMEFT analysis of vector boson scattering and diboson data from the LHC Run II*. 2021. arXiv: 2101.03180 [hep-ph].

[59] E. Chapon, C. Royon, and O. Kepka. "Anomalous quartic $WW\gamma\gamma$, $ZZ\gamma\gamma$, and trilinear $WW\gamma$ couplings in two-photon processes at high luminosity at the LHC". In: *Phys. Rev. D* 81.7 (2010). DOI: 10.1103/physrevd.81.074003. arXiv: 0912.5161 [hep-ph].

[60] G. Belanger and F. Boudjema. "$\gamma\gamma \to W^+W^-$ and $\gamma\gamma \to ZZ$ as tests of novel quartic couplings". In: 288.1 (1992), pp. 210–220. DOI: https://doi.org/10.1016/0370-2693(92)91979-J.

[61] OPAL Collaboration. "Constraints on anomalous quartic gauge boson couplings from $\nu\nu\gamma\gamma$ and $qq\gamma\gamma$ events at CERN LEP2". In: *Phys. Rev. D* 70.3 (2004). DOI: 10.1103/physrevd.70.032005. arXiv: hep-ex/0402021.

[62] D0 Collaboration. "Search for anomalous quartic $WW\gamma\gamma$ couplings in dielectron and missing energy final states in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. D* 88.1 (2013). DOI: 10.1103/physrevd.88.012005. arXiv: 1305.1258 [hep-ex].

[63] O. Kepka and C. Royon. "Anomalous $WW\gamma$ coupling in photon-induced processes using forward detectors at the CERN LHC". In: *Phys. Rev. D* 78.7 (2008). DOI: 10.1103/physrevd.78.073005. arXiv: 0808.0322 [hep-ph].

[64] The CMS Collaboration. *Limits on anomalous triple and quartic gauge couplings*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMPaTGC#Limits_on_anomalous_triple_and_q.

[65] Ernest O. Lawrence and M. Stanley Livingston. "The Production of High Speed Light Ions Without the Use of High Voltages". In: *Phys. Rev.* 40 (1 1932), pp. 19–35. DOI: 10.1103/PhysRev.40.19.

[66] Vladimir Veksler. "A new method of accelerating relativistic particles". In: *Compt. Rend. Acad. Sci. URSS* 43 (1944), pp. 346–348.

[67] Edwin M. McMillan. "The Synchrotron - A Proposed High Energy Particle Accelerator". In: *Phys. Rev.* 68 (5-6 1945), pp. 143–144. DOI: 10.1103/PhysRev.68.143.

[68] Evans Lyndon and Bryant Philip. "LHC Machine". In: *JINST* 3.08 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.

[69] R W Assmann, M Lamont, and S Myers. "A Brief History of the LEP Collider". In: *Nucl. Phys. B, Proc. Suppl.* 109 (2002), 17–31. 15 p. DOI: 10.1016/S0920-5632(02)90005-8.

[70] CERN. *LHC the guide*. 2017. URL: https://home.cern/resources/brochure/cern/lhc-guide.

[71] G. Arnison et al. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV". In: *Phys. Lett. B* 122 (1983), pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2.

[72] M. Banner et al. "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider". In: *Phys. Lett. B* 122 (1983), pp. 476–485. DOI: 10.1016/0370-2693(83)91605-2.

[73] G. Arnison et al. "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider". In: *Phys. Lett. B* 126 (1983), pp. 398–410. DOI: 10.1016/0370-2693(83)90188-0.

[74] P. Bagnaia et al. "Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider". In: *Phys. Lett. B* 129 (1983), pp. 130–140. DOI: 10.1016/0370-2693(83)90744-X.

[75] CERN. *CERN Accelerator Complex*. 2008. URL: https://public-archive.web.cern.ch/en/research/AccelComplex-en.html.

[76] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3.08 (2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003.

[77] The CMS Collaboration. "The CMS experiment at the CERN LHC". In: *JINST* 3.08 (2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004.

[78] The LHCb Collaboration. "The LHCb Detector at the LHC". In: *JINST* 3.08 (2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005.

[79] The ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *JINST* 3.08 (2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002.

[80] The ATLAS Collaboration. *ATLAS Beamspot Public Results - Run 2*. 2018. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BeamSpotPublicResults.

[81] *Characterization of Interaction-Point Beam Parameters Using the pp Event-Vertex Distribution Reconstructed in the ATLAS Detector at the LHC*. Tech. rep. ATLAS-CONF-2010-027. CERN, 2010. URL: https://cds.cern.ch/record/1277659.

[82] The ATLAS Collaboration. *ATLAS Luminosity Public Results - Run 2*. 2018. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[83] The ATLAS Collaboration. "ATLAS data quality operations and performance for 2015–2018 data-taking". In: *JINST* 15.04 (2020), P04003–P04003. DOI: 10.1088/1748-0221/15/04/p04003.

[84] G. Avoni et al. "The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS". In: *JINST* 13.07 (2018), P07017–P07017. DOI: 10.1088/1748-0221/13/07/p07017.

[85] *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. Tech. rep. ATLAS-CONF-2019-021. CERN, 2019. URL: https://cds.cern.ch/record/2677054.

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

[87] The ATLAS Collaboration. "Charged-particle distributions at low transverse momentum in $\sqrt{s} = 13$ TeV pp interactions measured with the ATLAS detector at the LHC". In: *Eur. Phys. J. C* 76.9 (2016). DOI: 10.1140/epjc/s10052-016-4335-y. arXiv: 1606.01133 [hep-ex].

[88] The ATLAS Collaboration. "Study of the material of the ATLAS inner detector for Run 2 of the LHC". In: *JINST* 12.12 (2017), P12009–P12009. DOI: 10.1088/1748-0221/12/12/p12009. arXiv: 1707.02826 [hep-ex].

[89] The ATLAS Collaboration. "The ATLAS Inner Detector commissioning and calibration". In: *Eur. Phys. J. C* 70.3 (2010), pp. 787–821. DOI: 10.1140/epjc/s10052-010-1366-7. arXiv: 1004.5293 [physics.ins-det].

[90] The ATLAS IBL Collaboration. "Production and Integration of the ATLAS Insertable B-Layer". In: *JINST* 13.05 (2018), T05008. DOI: 10.1088/1748-0221/13/05/T05008. arXiv: 1803.00844 [physics.ins-det].

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

[92] The ATLAS Collaboration. "The silicon microstrip sensors of the ATLAS semiconductor tracker". In: *NIM A* 578.1 (2007), pp. 98–118. DOI: https://doi.org/10.1016/j.nima.2007.04.157.

[93] The ATLAS Collaboration. "Operation and performance of the ATLAS semiconductor tracker". In: *JINST* 9.08 (2014), P08009–P08009. DOI: 10.1088/1748-0221/9/08/p08009. arXiv: 1004.5293 [hep-ex].

[94] E. Abat et al. "The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: Design and performance". In: *JINST* 3 (2008), P02013. DOI: 10.1088/1748-0221/3/02/P02013.

[95] The ATLAS Collaboration. "Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties". In: *JINST* 12.05 (2017), P05002–P05002. DOI: 10.1088/1748-0221/12/05/p05002. arXiv: 1702.06473 [hep-ex].

[96] John David Jackson. *Classical Electrodynamics*. Wiley, 1998. ISBN: 978-0-471-30932-1.

[97] A. Yamamoto et al. "The ATLAS central solenoid". In: *NIM A* 584.1 (2008), pp. 53–74. DOI: https://doi.org/10.1016/j.nima.2007.09.047.

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

[99] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1996. URL: http://cds.cern.ch/record/331061.

[100] *ATLAS tile calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1996. URL: http://cds.cern.ch/record/331062.

[101] *ATLAS muon spectrometer: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331068.

[102] The ATLAS Collaboration. "Commissioning of the ATLAS Muon Spectrometer with cosmic rays". In: *EPJC* 70.3 (2010), pp. 875–916. DOI: 10.1140/epjc/s10052-010-1415-2. arXiv: 1004.5293 [hep-ex].

[103] S. Grinstein. "The ATLAS Forward Proton Detector (AFP)". In: *Nuclear and Particle Physics Proceedings* 273-275 (2016), pp. 1180–1184. DOI: https://doi.org/10.1016/j.nuclphysbps.2015.09.185.

[104] *Letter of Intent for the Phase-I Upgrade of the ATLAS Experiment*. Tech. rep. CERN-LHCC-2011-012. LHCC-I-020. CERN, 2011. URL: https://cds.cern.ch/record/1402470.

[105] The CERN Courier. *Roman Pots for the LHC*. 1998. URL: https://cerncourier.com/a/roman-pots-for-the-lhc/.

[106] Sarah Charley. "LHC smashes old collision records". In: *Symmetry Magazine* (2016). URL: https://www.symmetrymagazine.org/article/lhc-smashes-old-collision-records.

[107] Aranzazu Ruiz-Martinez and The ATLAS Collaboration. *The Run-2 ATLAS Trigger System*. Tech. rep. ATL-DAQ-PROC-2016-003. CERN, 2016. DOI: 10.1088/1742-6596/762/1/012003. URL: https://cds.cern.ch/record/2133909.

[108] The ATLAS Collaboration. "Operation of the ATLAS trigger system in Run 2". In: *JINST* 15.10 (2020), P10004–P10004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/10/p10004. arXiv: 2007.12539 [physics.ins-det].

[109] *ATLAS level-1 trigger: Technical Design Report*. Technical design report. ATLAS. CERN, 1998. URL: http://cds.cern.ch/record/381429.

[110] Peter Jenni et al. *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical design report. ATLAS. CERN, 2003. URL: http://cds.cern.ch/record/616089.

[111] The ATLAS Collaboration. *Trigger Operation Public Results*. 2019. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults.

[112] The ATLAS Collaboration. "Performance of electron and photon triggers in ATLAS during LHC Run 2". In: *Eur. Phys. J. C* 80.1 (2020). DOI: 10.1140/epjc/s10052-019-7500-2. arXiv: 1909.00761 [hep-ex].

[113] The ATLAS Collaboration. "Performance of the ATLAS muon triggers in Run 2". In: *JINST* 15.09 (2020), P09015–P09015. DOI: 10.1088/1748-0221/15/09/p09015. arXiv: 2004.13447 [physics.ins-det].

[114] The ATLAS Collaboration. "Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data". In: *Eur. Phys. J. C* 77.3 (2017). DOI: 10.1140/epjc/s10052-017-4756-2. arXiv: 1612.01456 [hep-ex].

[115] The CMS Collaboration. "Measurement of charged pion, kaon, and proton production in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 96.11 (2017). DOI: 10.1103/physrevd.96.112003. arXiv: 1706.10194 [hep-ex].

[116] ATLAS Collaboration. *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*. ATL-SOFT-PUB-2007-007. 2007. URL: https://cds.cern.ch/record/1020106.

[117] A Salzburger. *The ATLAS Track Extrapolation Package*. Tech. rep. ATL-SOFT-PUB-2007-005. ATL-COM-SOFT-2007-010. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1038100.

[118] Andreas Salzburger on behalf of the ATLAS Collaboration. "Optimisation of the ATLAS Track Reconstruction Software for Run-2". In: *J. Phys.: Conf. Ser.* 664.7 (2015), p. 072042. DOI: 10.1088/1742-6596/664/7/072042.

[119] The ATLAS Collaboration. "Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2". In: *Eur. Phys. J. C* 77.10 (2017). DOI: 10.1140/epjc/s10052-017-5225-7. arXiv: 1704.07983 [hep-ex].

[120] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting". In: *NIM A* 262.2 (1987), pp. 444–450. DOI: https://doi.org/10.1016/0168-9002(87)90887-4.

[121] The ATLAS Collaboration. "Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC". In: *Eur. Phys. J. C* 77.5 (2017). DOI: 10.1140/epjc/s10052-017-4887-5. arXiv: 1611.10235 [physics.ins-det].

[122] *Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s}$ = 13 TeV*. Tech. rep. ATL-PHYS-PUB-2015-051. CERN, 2015. URL: https://cds.cern.ch/record/2110140.

[123] *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s}$ = 13 TeV*. Tech. rep. ATL-PHYS-PUB-2015-018. CERN, 2015. URL: https://cds.cern.ch/record/2037683.

[124] Jovan Mitrevski on behalf of the ATLAS Collaboration. "Electron and Photon Reconstruction with the ATLAS Detector". In: *Nuclear and Particle Physics Proceedings* 273-275 (2016), pp. 2539–2541. DOI: https://doi.org/10.1016/j.nuclphysbps.2015.09.452.

[125] The ATLAS Collaboration. "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s}$ = 13 TeV". In: *Eur. Phys. J. C* 79.8 (2019). DOI: 10.1140/epjc/s10052-019-7140-6. arXiv: 1902.04655 [physics.ins-det].

[126] The ATLAS Collaboration. "Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016". In: *Eur. Phys. J. C* 79 (2019). DOI: 10.1140/epjc/s10052-019-6650-6. arXiv: 1810.05087 [hep-ex].

[127] W Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*. Tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003. CERN, 2008. URL: https://cds.cern.ch/record/1099735.

[128] *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*. Tech. rep. ATLAS-CONF-2012-047. CERN, 2012. URL: https://cds.cern.ch/record/1449796.

[129] The ATLAS Collaboration. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *Eur. Phys. J. C* 77.7 (2017). DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex].

[130] The ATLAS Collaboration. "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ s =13 TeV". In: *Eur. Phys. J. C* 76.5 (2016). DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].

[131] J. Illingworth and J. Kittler. "A survey of the hough transform". In: *Computer Vision, Graphics, and Image Processing* 44.1 (1988), pp. 87–116. DOI: https://doi.org/10.1016/S0734-189X(88)80033-1.

[132] A. Hrynevich on behalf of the ATLAS Collaboration. "ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2". In: *JINST* 12.06 (2017), pp. C06038–C06038. DOI: 10.1088/1748-0221/12/06/c06038.

[133] Steven Schramm. *ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz-boosted Objects*. Tech. rep. ATL-PHYS-PROC-2017-236. CERN, 2017. URL: https://cds.cern.ch/record/2291608.

[134] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". In: *JHEP* 2008.04 (2008), pp. 063–063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[135] David Krohn, Jesse Thaler, and Lian-Tao Wang. "Jet trimming". In: *J HEP* 2010.2 (2010). DOI: 10.1007/jhep02(2010)084. arXiv: 0912.1342 [hep-ph].

[136] The ATLAS Collaboration. "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 96.7 (2017). DOI: 10.1103/physrevd.96.072002. arXiv: 1703.09665 [hep-ex].

[137] The ATLAS Collaboration. "Jet reconstruction and performance using particle flow with the ATLAS Detector". In: *Eur. Phys. J. C* 77.7 (2017). DOI: 10.1140/epjc/s10052-017-5031-2. arXiv: 1703.10485 [hep-ex].

[138] The ATLAS Collaboration. "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 79.11 (2019). DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex].

[139] *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*. Tech. rep. ATLAS-CONF-2016-035. CERN, 2016. URL: https://cds.cern.ch/record/2200211.

[140] The ATLAS Collaboration. "Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment". In: *Eur. Phys. J. C* 76.5 (2016). DOI: 10.1140/epjc/s10052-016-4110-0. arXiv: 1512.05955 [hep-ex].

[141] The ATLAS Collaboration. "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 78.11 (2018). DOI: 10.1140/epjc/s10052-018-6288-9. arXiv: 1802.08168 [hep-ex].

[142] Andy Buckley et al. "General-purpose event generators for LHC physics". In: *Physics Reports* 504.5 (2011), pp. 145–233. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].

[143] 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". In: *JHEP* 2014.7 (2014). DOI: 10.1007/jhep07(2014)079. arXiv: 1405.0301 [hep-ph].

[144] David J. Lange. "The EvtGen particle decay simulation package". In: *NIM A* 462.1 (2001), pp. 152–155. DOI: https://doi.org/10.1016/S0168-9002(01)00089-4.

[145] Manuel Bähr et al. "Herwig++ physics and manual". In: *Eur. Phys. J. C* 58.4 (2008), pp. 639–707. DOI: 10.1140/epjc/s10052-008-0798-9.

[146] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76.4 (2016). DOI: 10.1140/epjc/s10052-016-4018-8.

[147] V.M. Budnev et al. "The two-photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation". In: *Phys. Rep* 15.4 (1975), pp. 181–282. DOI: https://doi.org/10.1016/0370-1573(75)90009-5.

[148] H. Jung et al. "ThePEG: Toolkit for High Energy Physics event generation". In: *Proceedings of the workshop: HERA and the LHC workshop series on the implications of HERA for LHC physics* (2009). arXiv: 0903.3861 [hep-ph].

[149] Carl Schmidt et al. "CT14QED parton distribution functions from isolated photon production in deep inelastic scattering". In: *Phys. Rev. D* 93.11 (2016). DOI: 10.1103/physrevd.93.114015. arXiv: 1509.02905 [hep-ph].

[150] J.A.M. Vermaseren. "Two-photon processes at very high energies". In: *Nucl. Phys. B* 229.2 (1983), pp. 347–371. DOI: https://doi.org/10.1016/0550-3213(83)90336-X.

[151] Paolo Nason. "A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms". In: *JHEP* 2004.11 (2004), pp. 040–040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[152] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *JHEP* 2007.11 (2007), pp. 070–070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[153] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 2010.6 (2010). DOI: 10.1007/jhep06(2010)043. arXiv: 1002.2581 [hep-ph].

[154] Tom Melia et al. "W^+W^-, WZ and ZZ production in the POWHEG BOX". In: *JHEP* 2011.11 (2011). DOI: 10.1007/jhep11(2011)078. arXiv: 1107.5051 [hep-ph].

[155] Paolo Nason and Giulia Zanderighi. "W^+W^-, WZ and ZZ production in the POWHEG-BOX-V2". In: *Eur. Phys. J. C* 74.1 (2014). DOI: 10.1140/epjc/s10052-013-2702-5. arXiv: 1311.1365 [hep-ph].

[156] Hung-Liang Lai et al. "New parton distributions for collider physics". In: *Phys. Rev. D* 82.7 (2010). DOI: 10.1103/physrevd.82.074024. arXiv: 1007.2241 [hep-ph].

[157] The ATLAS Collaboration. "Measurement of the $Z/\gamma*$ boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *JHEP* 2014.9 (2014). DOI: 10.1007/jhep09(2014)145. arXiv: 1406.3660 [hep-ex].

[158] Jonathan Pumplin et al. "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis". In: *JHEP* 2002.07 (2002), pp. 012–012. DOI: 10.1088/1126-6708/2002/07/012. arXiv: hep-ph/0201195.

[159] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76.4 (2016). DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[160] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J. C* 75.5 (2015). DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[161] Stefan Höche et al. "QCD matrix elements and truncated showers". In: *JHEP* 2009.05 (2009), pp. 053–053. DOI: 10.1088/1126-6708/2009/05/053. arXiv: 0903.1219 [hep-ph].

[162] Enrico Bothmann et al. "Event generation with Sherpa 2.2". In: *SciPost Physics* 7.3 (2019). DOI: 10.21468/scipostphys.7.3.034. arXiv: 1905.09127 [hep-ph].

[163] Tanju Gleisberg and Stefan Höche. "Comix, a new matrix element generator". In: *JHEP* 2008.12 (2008), pp. 039–039. DOI: 10.1088/1126-6708/2008/12/039. arXiv: 0808.3674 [hep-ph].

[164] S Schumann and F Krauss. "A parton shower algorithm based on Catani-Seymour dipole factorisation". In: *JHEP* 2008.03 (2008), pp. 038–038. DOI: 10.1088/1126-6708/2008/03/038. arXiv: 0709.1027 [hep-ph].

[165] Stefan Höche et al. "QCD matrix elements and truncated showers". In: *JHEP* 2009.05 (2009), pp. 053–053. DOI: 10.1088/1126-6708/2009/05/053. arXiv: 0903.1219 [hep-ph].

[166] Stefan Höche et al. "A critical appraisal of NLO+PS matching methods". In: *JHEP* 2012.9 (2012). DOI: 10.1007/jhep09(2012)049. arXiv: 1111.1220 [hep-ph].

[167] Stefan Höche et al. "QCD matrix elements + parton showers. The NLO case". In: *JHEP* 2013.4 (2013). DOI: 10.1007/jhep04(2013)027. arXiv: 1207.5030 [hep-ph].

[168] Federico Buccioni et al. "OpenLoops 2". In: *Eur. Phys. J. C* 79.10 (2019). DOI: 10.1140/epjc/s10052-019-7306-2. arXiv: 1907.13071 [hep-ph].

[169] F. Cascioli, P. Maierhöfer, and S. Pozzorini. "Scattering Amplitudes with Open Loops". In: *Phys. Rev. Lett.* 108.11 (2012). DOI: 10.1103/physrevlett.108.111601. arXiv: 1111.5206 [hep-ph].

[170] Ansgar Denner, Stefan Dittmaier, and Lars Hofer. "Collier: A fortran-based complex one-loop library in extended regularizations". In: *Comput. Phys. Commun* 212 (2017), pp. 220–238. DOI: 10.1016/j.cpc.2016.10.013. arXiv: 1604.06792 [hep-ph].

[171] Richard D. Ball et al. "Parton distributions for the LHC run II". In: *JHEP* 2015.4 (2015). DOI: 10.1007/jhep04(2015)040. arXiv: 1410.8849 [hep-ph].

[172] Giuseppe Bozzi et al. "Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC". In: *Nucl. Phys. B* 737.1-2 (2006), pp. 73–120. DOI: 10.1016/j.nuclphysb.2005.12.022. arXiv: hep-ph/0508068.

[173] Daniel de Florian et al. "Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC". In: *JHEP* 2011.11 (2011). DOI: 10.1007/jhep11(2011)064.

[174] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason. "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction". In: *JHEP* 2007.09 (2007), pp. 126–126. DOI: 10.1088/1126-6708/2007/09/126.

[175] Emanuele Re. "Single-top Wt-channel production matched with parton showers using the POWHEG method". In: *Eur. Phys. J. C* 71.2 (2011). DOI: 10.1140/epjc/s10052-011-1547-z.

[176] Stefano Frixione et al. "Single-top hadroproduction in association with aWboson". In: *JHEP* 2008.07 (2008), pp. 029–029. DOI: 10.1088/1126-6708/2008/07/029.

[177] *ATLAS Pythia 8 tunes to 7 TeV data*. Tech. rep. ATL-PHYS-PUB-2014-021. CERN, 2014. URL: https://cds.cern.ch/record/1966419.

[178] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[179] *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model*. Tech. rep. ATL-PHYS-PUB-2016-017. CERN, 2016. URL: https://cds.cern.ch/record/2206965.

[180] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys. B* 867.2 (2013), pp. 244–289. DOI: 10.1016/j.nuclphysb.2012.10.003.

[181] S. Agostinelli et al. "GEANT4 – a simulation toolkit". In: *NIM A* 506.3 (2003), pp. 250–303. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.

[182] The ATLAS Collaboration. "The ATLAS Simulation Infrastructure". In: *Eur. Phys. J. C* 70.3 (2010), pp. 823–874. DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568 [physics.ins-det].

[183] The ATLAS Collaboration. *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*. Tech. rep. ATL-PHYS-PUB-2010-013. CERN, 2010. URL: https://cds.cern.ch/record/1300517.

[184] The ATLAS Collaboration. "Measurement of charged-particle distributions sensitive to the underlying event in $\sqrt{s} = 13$ TeV proton-proton collisions with the ATLAS detector at the LHC". In: *JHEP* 2017.3 (2017). DOI: 10.1007/jhep03(2017)157.

[185] Jahred Adelman et al. on behalf of the ATLAS Collaboration. "ATLAS FTK Challenge: Simulation of a Billion-fold Hardware Parallelism". In: *Procedia Computer Science* 66 (2015), pp. 540–545. DOI: https://doi.org/10.1016/j.procs.2015.11.061.

[186] The ATLAS Collaboration. "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data". In: *JINST* 14.12 (2019), P12006–P12006. DOI: 10.1088/1748-0221/14/12/p12006.

[187] ATLAS Collaboration. *Observation of photon-induced W^+W^- production in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector; public plots*. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2017-21/.

[188] The ATLAS Collaboration. "Measurement of distributions sensitive to the underlying event in inclusive Z boson production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 79.8 (2019). DOI: 10.1140/epjc/s10052-019-7162-0.

[189] The ATLAS Collaboration. "Measurement of event-shape observables in $Z \to \ell^+\ell^-$ events in pp collisions at $\sqrt{s} = 7$ TeV TeV with the ATLAS detector at the LHC". In: *Eur. Phys. J. C* 76.7 (2016). DOI: 10.1140/epjc/s10052-016-4176-8.

[190] The ATLAS Collaboration. "Measurement of distributions sensitive to the underlying event in inclusive Z-boson production in *pp* collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 74.12 (2014). DOI: 10.1140/epjc/s10052-014-3195-6.

[191] G. D'Agostini. "A multidimensional unfolding method based on Bayes' theorem". In: *NIM A* 362.2 (1995), pp. 487–498. DOI: https://doi.org/10.1016/0168-9002(95)00274-X.

[192] G. D'Agostini. "Improved iterative Bayesian unfolding". In: *Alliance Workshop on Unfolding and Data Correction*. Oct. 2010. arXiv: 1010.0632 [physics.data-an].

[193] Stefan Kallweit et al. *Accurate single- and double-differential resummation of colour-singlet processes with MATRIX+RadISH: W^+W^- production at the LHC*. 2021. arXiv: 2004.07720 [hep-ph].

[194] Massimiliano Grazzini, Stefan Kallweit, and Marius Wiesemann. "Fully differential NNLO computations with MATRIX". In: *Eur. Phys. J. C* 78.7 (2018). DOI: 10.1140/epjc/s10052-018-5771-7. arXiv: 1711.06631 [hep-ph].

[195] Massimiliano Grazzini et al. "W^+W^- production at the LHC: fiducial cross sections and distributions in NNLO QCD". In: *JHEP* 2016.8 (2016). DOI: 10.1007/jhep08(2016)140. arXiv: 1605.02716 [hep-ph].

[196] T. Gehrmann et al. "W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD". In: *Phys. Rev. Lett.* 113.21 (2014). DOI: 10.1103/physrevlett.113.212001. arXiv: 1408.5243 [hep-ph].

[197] Wojciech Bizoń et al. "Momentum-space resummation for transverse observables and the Higgs $p_\perp$ at N^{3}LL+NNLO". In: *JHEP* 2018.2 (2018). DOI: 10.1007/jhep02(2018)108. arXiv: 1705.09127 [hep-ph].

[198] Pier Francesco Monni, Emanuele Re, and Paolo Torrielli. "A new approach to the Higgs transverse-momentum resummation at NNLL+NNLO". In: *Phys. Rev. Lett.* 116.24 (2016). DOI: 10.1103/physrevlett.116.242001. arXiv: 1604.02191 [hep-ph].

[199] Thomas Gehrmann, Andreas von Manteuffel, and Lorenzo Tancredi. "The two-loop helicity amplitudes for $q\bar{q}' \to V_1V_2 \to 4$ leptons". In: *JHEP* 2015.9 (2015). DOI: 10.1007/jhep09(2015)128. arXiv: 1503.04812 [hep-ph].

[200] Stefano Catani and Massimiliano Grazzini. "An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC". In: *Phys. Rev. Lett.* 98.22 (2007). DOI: 10.1103/physrevlett.98.222002. arXiv: hep-ph/0703012.

[201] Stefano Catani et al. "Vector-boson production at hadron colliders: hard-collinear coefficients at the NNLO". In: *Eur. Phys. J. C* 72.11 (2012). DOI: 10.1140/epjc/s10052-012-2195-7. arXiv: 1209.0158 [hep-ph].

[202] P. de Jong. *Top Physics at the LHC*. 2009. arXiv: 0902.4798 [hep-ex].

[203] *Tagging and suppression of pileup jets with the ATLAS detector*. Tech. rep. ATLAS-CONF-2014-018. 2014. URL: https://cds.cern.ch/record/1700870.

[204] *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*. Tech. rep. ATL-PHYS-PUB-2015-026. 2015. URL: https://cds.cern.ch/record/2037717.

[205] The ATLAS Collaboration. "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector". In: *Eur. Phys. J. C* 76.11 (2016). DOI: 10.1140/epjc/s10052-016-4395-z. arXiv: 1510.03823 [hep-ex].

[206] *Modelling of Track Reconstruction Inside Jets with the 2016 ATLAS $\sqrt{s} = 13$ TeV pp dataset*. Tech. rep. ATL-PHYS-PUB-2017-016. 2017. URL: https://cds.cern.ch/record/2275639.

[207] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. CERN-OPEN-2012-016. 2012. URL: https://cds.cern.ch/record/1456844.

[208] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71.2 (2011). DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[209] D Barberis et al. "The ATLAS Eventindex: an event catalogue for experiments collecting large amounts of data". In: *J. Phys.: Conf. Ser.* 513.4 (2014), p. 042002. DOI: 10.1088/1742-6596/513/4/042002.

[210] William Patrick McCormack et al. *Low-p_{T} tracking for ATLAS in nominal LHC pileup*. Tech. rep. ATL-PHYS-PROC-2020-041. 2020. URL: https://cds.cern.ch/record/2718583.

[211] *dE/dx measurement in the ATLAS Pixel Detector and its use for particle identification*. Tech. rep. ATLAS-CONF-2011-016. 2011. URL: https://cds.cern.ch/record/1336519.

[212] Ke-Lin Du and M. N. S. Swamy. *Neural Networks and Statistical Learning*. London, UK: Springer-Verlag London Ltd., 2019. ISBN: 978-1-4471-7452-3.

[213] Wikpedia User Wiso. *Simple neural network*. URL: https://commons.wikimedia.org/wiki/File:Neural_network_example.svg.

[214] François Chollet et al. *Keras*. https://keras.io. 2015.

[215] Martin Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: http://tensorflow.org/.

[216] *Trigger Menu in 2017*. Tech. rep. ATL-DAQ-PUB-2018-002. CERN, 2018. URL: https://cds.cern.ch/record/2625986.

[217] Andreas Hoecker et al. "TMVA: Toolkit for Multivariate Data Analysis". In: *PoS* ACAT (2007), p. 040. arXiv: physics/0703039.

[218] *The Expected Performance of the ATLAS Inner Detector*. Tech. rep. 2008. URL: http://cds.cern.ch/record/1118445.

[219] The ATLAS collaboration. "A neural network clustering algorithm for the ATLAS silicon pixel detector". In: *JINST* 9.09 (2014), P09009–P09009. DOI: 10.1088/1748-0221/9/09/p09009. arXiv: 1406.7690 [hep-ex].

[220] *The Optimization of ATLAS Track Reconstruction in Dense Environments*. Tech. rep. ATL-PHYS-PUB-2015-006. 2015. URL: https://cds.cern.ch/record/2002609.

[221] *Measurement of delta-rays in ATLAS silicon sensors*. Tech. rep. ATLAS-CONF-2013-005. 2013. URL: http://cds.cern.ch/record/1509553.

[222] V Cindro et al. "The ATLAS Beam Conditions Monitor". In: *JINST* 3.02 (2008), P02004–P02004. DOI: 10.1088/1748-0221/3/02/p02004.

[223] S van der Meer. *Calibration of the effective beam height in the ISR*. Tech. rep. CERN-ISR-PO-68-31. ISR-PO-68-31. Geneva: CERN, 1968. URL: https://cds.cern.ch/record/296752.

[224] The ATLAS Collaboration. *Pixel-Cluster Counting Luminosity Measurement In ATLAS*. Tech. rep. ATL-DAPR-PROC-2016-001. 2016. DOI: 10.22323/1.282.1064. URL: https://cds.cern.ch/record/2232275.

[225] *CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update*. Tech. rep. CMS-PAS-LUM-13-001. 2013. URL: http://cds.cern.ch/record/1598864.

[226] Jason Gallicchio and Matthew D. Schwartz. "Quark and Gluon Tagging at the LHC". In: *Phys. Rev. Lett.* 107.17 (2011). DOI: 10.1103/physrevlett.107.172001. arXiv: 1106.3076 [hep-ph].

[227] Andrew J. Larkoski, Jesse Thaler, and Wouter J. Waalewijn. "Gaining (mutual) information about quark/gluon discrimination". In: *JHEP* 2014.11 (2014). DOI: 10.1007/jhep11(2014)129. arXiv: 1408.3122 [hep-ph].

[228] Christopher Frye et al. "Casimir meets Poisson: improved quark/gluon discrimination with counting observables". In: *JHEP* 2017.9 (2017). DOI: 10.1007/jhep09(2017)083. arXiv: 1704.06266 [hep-ph].

[229] Peter Berta et al. "Particle-level pileup subtraction for jets and jet shapes". In: *JHEP* 2014.6 (2014). DOI: 10.1007/jhep06(2014)092. arXiv: 1403.3108 [hep-ph].

[230] *Constituent-level pile-up mitigation techniques in ATLAS*. Tech. rep. ATLAS-CONF-2017-065. 2017. URL: http://cds.cern.ch/record/2281055.

[231] *Impact of Alternative Inputs and Grooming Methods on Large-R Jet Reconstruction in ATLAS*. Tech. rep. ATL-PHYS-PUB-2017-020. 2017. URL: http://cds.cern.ch/record/2297485.

[232] Andrew J. Larkoski et al. "Soft drop". In: *JHEP* 2014.5 (2014). DOI: 10.1007/jhep05(2014)146. arXiv: 1402.2657 [hep-ph].

[233] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet user manual". In: *Eur. Phys. J. C* 72.3 (2012). DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[234] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett. B* 659.1-2 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].

[235] Yu.L Dokshitzer et al. "Better jet clustering algorithms". In: *JHEP* 1997.08 (1997), pp. 001–001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323.

[236] M. Wobisch and T. Wengler. *Hadronization corrections to jet cross-sections in deep inelastic scattering*. 1998. arXiv: hep-ph/9907280.

[237] ATLAS Collaboration. *ATLAS Run 1 Pythia8 tunes*. Tech. rep. 2014. URL: https://cds.cern.ch/record/1966419.

[238] ATLAS Collaboration. *Summary of ATLAS Pythia 8 tunes*. Tech. rep. 2012. URL: https://cds.cern.ch/record/1474107.

[239] Christine Hu-Guo et al. "CMOS pixel sensor development: A fast read-out architecture with integrated zero suppression". In: *JINST* 4 (2009), P04012. DOI: 10.1088/1748-0221/4/04/P04012.

[240] Marco Battaglia et al. "R&D Paths of Pixel Detectors for Vertex Tracking and Radiation Imaging". In: *Nucl. Instrum. Meth. A* 716 (2013), pp. 29–45. DOI: 10.1016/j.nima.2013.03.040. arXiv: 1208.0251 [physics.ins-det].

[241] G. Contin. ""The MAPS-based ITS Upgrade for ALICE"". In: *Proceedings of Vertex2019 - The 28th International Workshop on Vertex Detectors* (2019).

[242] Zheng Li. *"Radiation Hardness/Tolerance of Si Sensors/Detectors for Nuclear and High Energy Physics Experiments"*. 2002. URL: https://www.slac.stanford.edu/econf/C020909/.

[243] Maurice Garcia-Sciveres and Norbert Wermes. "A review of advances in pixel detectors for experiments with high rate and radiation". In: *Reports on Progress in Physics* 81.6 (2018), p. 066101. DOI: 10.1088/1361-6633/aab064.

[244] ATLAS Collaboration. *"Technical Design Report for the ATLAS Inner Tracker Pixel Detector"*. Tech. rep. CERN-LHCC-2017-021. ATLAS-TDR-030. 2017. URL: http://cds.cern.ch/record/2285585.

[245] ATLAS Collaboration. *"Expected Tracking Performance of the ATLAS Inner Tracker at the HL-LHC"*. Tech. rep. ATL-PHYS-PUB-2019-014. Geneva, 2019. URL: http://cds.cern.ch/record/2669540.

[246] M. Hage-Ali and P. Siffert. ""CdTe nuclear detectors and application"". In: *Semiconductors and semimetals: Semiconductors for room temperature nuclear detector applications, Vol. 43* (1995).

[247] M Keil. *Operational Experience with the ATLAS Pixel Detector at LHC*. Tech. rep. ATL-INDET-PROC-2012-031. 2012. DOI: 10.22323/1.167.0049. URL: https://cds.cern.ch/record/1498736.

[248] Tamas Almos Vami. *Calibration and performance of the CMS pixel detector in LHC Run 2*. 2019. arXiv: 1909.12920 [physics.ins-det].

[249] János Karancsi. "Operational Experience with the CMS Pixel Detector". In: *JINST* 10.05 (2015), p. C05016. DOI: 10.1088/1748-0221/10/05/C05016. arXiv: 1411.4185 [physics.ins-det].

[250] P. Zambon et al. "A wide energy range calibration algorithm for X-ray photon counting pixel detectors using high-Z sensor material". In: *NIM A* 925 (2019), pp. 164–171. DOI: https://doi.org/10.1016/j.nima.2019.01.095.

[251] Huanjun Ding et al. "Characterization of energy response for photon-counting detectors using x-ray fluorescence". In: *Medical Physics* 41.12 (2014). DOI: 10.1118/1.4900820.

[252] Mini Das et al. "Energy calibration of photon counting detectors using x-ray tube potential as a reference for material decomposition applications". In: *Medical Imaging 2015: Physics of Medical Imaging*. Ed. by Christoph Hoeschen and Despina Kontos. Vol. 9412. International Society for Optics and Photonics. SPIE, 2015, pp. 263–268. DOI: 10.1117/12.2082979.

[253] David-Leon Pohl et al. "Obtaining spectroscopic information with the ATLAS FE-I4 pixel readout chip". In: *NIM A* 788 (2015), pp. 49–53. DOI: 10.1016/j.nima.2015.03.067. arXiv: 1408.2952 [physics.ins-det].

[254] Ph Bloch et al. "Silicon sensors for the CMS preshower detector". In: *NIM A* 479.2 (2002), pp. 265–277. DOI: https://doi.org/10.1016/S0168-9002(01)00900-7.

[255] A. Thompson et al. *"X-Ray Data Booklet (3rd Edition)"*. 2009. URL: https://xdb.lbl.gov/.

[256] Christina Streli, P. Wobrauschek, and Peter Kregsamer. "X-Ray Fluorescence Spectroscopy, Applications*". In: 1999, pp. 2478–2487. ISBN: 978-0-12-226680-5. DOI: 10.1006/rwsp.2000.0337.

[257] Arthur H. Compton. "A Quantum Theory of the Scattering of X-rays by Light Elements". In: *Phys. Rev.* 21 (5 1923), pp. 483–502. DOI: 10.1103/PhysRev.21.483.

[258] O. Klein and Y. Nishina. ""Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac"". In: *Z. Phys.* 52.11 (1929), pp. 853–868. DOI: 10.1007/BF01366453.

[259] Jesse W. M. Du Mond. "Compton Modified Line Structure and its Relation to the Electron Theory of Solid Bodies". In: *Phys. Rev.* 33 (5 1929), pp. 643–658. DOI: 10.1103/PhysRev.33.643.

[260] J. W. M. DuMond. "Breadth of Compton Modified Line". In: *Phys. Rev.* 36 (1 1930), pp. 146–147. DOI: 10.1103/PhysRev.36.146.

[261] B.K. Chatterjee, L.A. LaJohn, and S.C. Roy. "Investigations on compton scattering: New directions". In: *Radiation Physics and Chemistry* 75.12 (2006). New Directions in the Study of Interaction of Energetic Photons with Matter, pp. 2165–2173. ISSN: 0969-806X. DOI: https://doi.org/10.1016/j.radphyschem.2006.03.073. URL: https://www.sciencedirect.com/science/article/pii/S0969806X06001782.

[262] Andreas Zoglauer and Gottfried Kanbach. "Doppler broadening as a lower limit to the angular resolution of next-generation Compton telescopes". In: *X-Ray and Gamma-Ray Telescopes and Instruments for Astronomy*. Ed. by Joachim E. Truemper and Harvey D. Tananbaum. Vol. 4851. International Society for Optics and Photonics. SPIE, 2003, pp. 1302–1309. DOI: 10.1117/12.461177.

[263] A. Bähr, P. Lechner, and J. Ninkovic. "Sensor development at the semiconductor laboratory of the Max-Planck-Society". In: *JINST* 12.12 (2017), P12005–P12005. DOI: 10.1088/1748-0221/12/12/p12005.

[264] Maurice Garcia-Sciveres. *The RD53A Integrated Circuit*. Tech. rep. CERN-RD53-PUB-17-001. 2017. URL: https://cds.cern.ch/record/2287593.

[265] Timon Heim. "YARR - A PCIe based Readout Concept for Current and Future ATLAS Pixel Modules". In: *Journal of Physics: Conference Series* 898 (2017), p. 032053. DOI: 10.1088/1742-6596/898/3/032053.

[266] R. Redus et al. "Design and performance of the X-123 compact X-ray and gamma-ray spectroscopy system". In: *2006 IEEE Nuclear Science Symposium Conference Record*. Vol. 6. 2006, pp. 3794–3797. DOI: 10.1109/NSSMIC.2006.353819.

[267] Sparsh Mittal. "A Survey Of Architectural Techniques for Managing Process Variation". In: *ACM Computing Surveys* 48 (Feb. 2016). DOI: 10.1145/2871167.

[268] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 978-0-201-50397-5.

[269] A. Zee. *Quantum field theory in a nutshell*. USA: Princeton University Press, 2003. ISBN: 978-0-691-14034-6.

[270] A. Zee. *Group Theory in a Nutshell for Physicists*. USA: Princeton University Press, 2016. ISBN: 978-0-691-16269-0.

[271] Mark Thomson. *Modern particle physics*. New York: Cambridge University Press, 2013. ISBN: 978-1-107-03426-6.

[272] H. Rauch et al. "Verification of coherent spinor rotation of fermions". In: *Phys. Lett. A* 54.6 (1975), pp. 425–427. DOI: https://doi.org/10.1016/0375-9601(75)90798-7.

[273] S. A. Werner et al. "Observation of the Phase Shift of a Neutron Due to Precession in a Magnetic Field". In: *Phys. Rev. Lett.* 35 (16 1975), pp. 1053–1055. DOI: 10.1103/PhysRevLett.35.1053.

[274] A. Einstein. "On a Heuristic Point of View Concerning the Production and Transformation of Light". In: *Ann. Phys.* 17 (1905), pp. 132–148.

[275] J. J. Thomson. "Cathode Rays". In: *Philosophical Magazine* 44.269 (1897), pp. 293–316.

[276] G. P. Thomson and A. Reid. "Diffraction of Cathode Rays by a Thin Film". In: *Nature* 119.3007 (1927), p. 890. DOI: 10.1038/119890a0.

[277] John Ellis, Mary K. Gaillard, and Dimitri V. Nanopoulos. "A Historical Profile of the Higgs Boson". In: *The Standard Theory of Particle Physics* (2016), pp. 255–274. DOI: 10.1142/9789814733519_0014. arXiv: 1504.07217 [hep-ph].

Appendix A

Optimizing the Exclusivity Selection without Low-p_{T} Tracks

In the $\gamma\gamma \to W^+W^-$ analysis presented in Part II, a ± 1 mm exclusivity window around the lepton vertex was used, such that if an event had one or more tight-primary track with $d_0 < 1.0$ mm in the window, the event would not be in the signal region. This window size was used in the analysis detailed in [41], which used proton-proton LHC data collected at $\sqrt{s} = 8$ TeV and lower average pileup than was observed throughout Run 2. In contrast, the $\gamma\gamma \to \ell^+\ell^-$ analysis presented in [36], used a ± 0.5 mm window. However, while this analysis was performed with Run 2 data, it also used the ATLAS Forward Proton Detector (AFP) to identify photon-induced events. Additionally, it's goal was not to make a first observation of $\gamma\gamma \to \ell^+\ell^-$, but to make a precise, low-background measurement using a new detector element (the AFP). It was initially not entirely clear what event selection strategy should be used in the Part II analysis. Several strategies were explored, including:

1. The use of different window sizes, with a possibility for a pileup-dependent window size
2. Attempts to discriminate pileup from underlying event tracks
3. A machine-learning (ML) based approach to event selection.

A.1 Optimization without Machine Learning

In Chapter 15, a metric was introduced for exclusivity optimization studies when using only elastic $\gamma\gamma \to W^+W^-$ and $q\bar{q} \to W^+W^-$ samples (Equation 15.2). This metric had not been developed yet when the initial optimization studies for Part II were being performed; the S_{Excl} scale factor was not known, nor was the ratio of $q\bar{q} \to W^+W^-$ to other backgrounds. Instead another Poisson statistics-motivated metric was used:

$$\frac{\varepsilon_S}{\sqrt{\varepsilon_B}}. \tag{A.1}$$

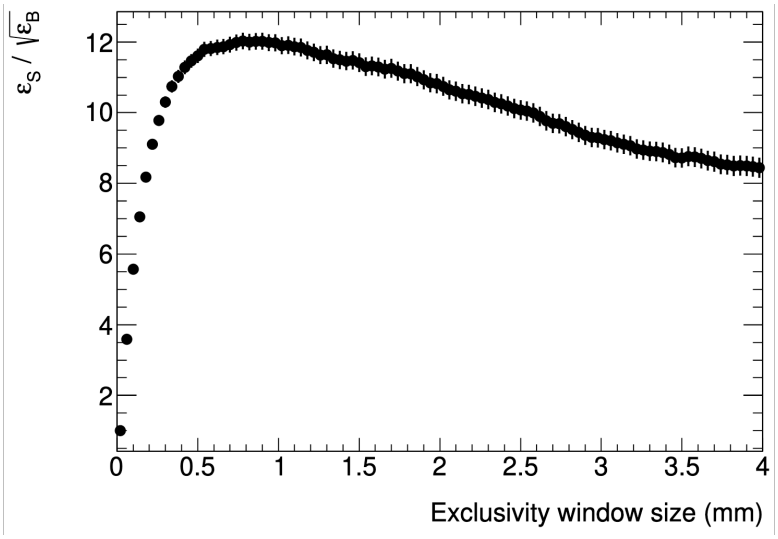


Figure A.1: The $\varepsilon_S/\sqrt{\varepsilon_B}$ metric as a function of window size. This study used only the elastic $\gamma\gamma \to W^+W^-$ sample and the default $q\bar{q} \to W^+W^-$ sample.

Here, ε_S and ε_B are the acceptance efficiencies for signal events and background events respectively. This metric attempts to capture the fact that statistical significance of the analysis should depend linearly on the signal yield and should be inversely proportional to the square root of the background yield, which is a statistical uncertainty in the analysis. The total number of signal and background events will take the form $N_{S,B} * \varepsilon_{S,B}$. When comparing the relative performance of two different event selection working points, the factors of $N_{S,B}$ cancel, leaving only $\varepsilon_{S,B}$.

Figure A.1 shows the spectrum of this metric as a function of window size. While the peak occurs slightly below 1 mm, the relative performance gain is expected to be negligible.

The use of three different window sizes, one for a pileup range of 0 - 25 collisions, one for 25 - 50, and one for 50+, was also considered. While the higher pileup bins typically favored a smaller window, the expected performance gain was small. For the sake of simplicity and consistency, a ± 1 mm window was determined satisfactory.

An approach was also explored where an event could be accepted if there was one track in the ± 1 mm window, and that track was closer to a reconstructed pileup vertex than to the lepton vertex. This also failed to produce a gain in performance. While this allowance would increase the signal yield, it also increases the background yield. Changes such as this tended to increase ε_B more than ε_S, resulting in no net gain statistical significance.

A.2 Exploration with Machine Learning

The use of a exclusivity window, which is a boolean, is logically equivalent to classifier based on the distance from the lepton vertex to the closest reconstructed track, which is a real number. In the case of the analysis presented in Part II, events are rejected if the distance to the closest track is < 1 mm. However, we can change that single variable classifier into one that uses multiple variables by using machine learning, similar to the true vs. fake track classifier presented in Section 4.1.

In this case, a TMVA [217] Boosted Decision Tree (BDT) is used. A BDT is somewhat similar to a neural network (NN) in the practical sense that it exploits multivariate correlations to make a classification prediction. However, where a NN forms linear combinations of the input variables in hidden layers, adjusting these combinations to minimize a loss function, a Decision Tree makes branching accept/reject choices based on adjustable cuts on the input variables. This would normally be just a boolean output, but the tree can be "boosted" by creating different trees on different subsets of the data, and then making a weighted final decision for events, based on information from these different trees.

A BDT was developed that took in 25 input features per event:

1. The distance from the lepton vertex to the 5 closest tracks
2. The uncertainty on the z parameter of the 5 closest tracks
3. The d_0 of the 5 closest tracks
4. The uncertainty on the d_0 parameter of the 5 closest tracks
5. A boolean for each of the 5 closest tracks, expressing whether the track passes the tight-primary selection.

The default training parameters for TMVA BDTs were used.

The blue line in Figure A.2 shows the ROC curve of this BDT. The red star indicates the working point associated with the ± 1 mm exclusivity window. The ROC curve essentially passes through this working point, so there is little to be gained by using the BDT.

A NN was also developed that takes in 11 variables per event:

1. The absolute distance from the lepton vertex to the two closest tracks
2. The d_0/σ_{d_0} for the two closest tracks
3. The number of IBL hits for the two closest tracks
4. The number of Pixel hits for the two closest tracks
5. The number of SCT hits for the two closest tracks

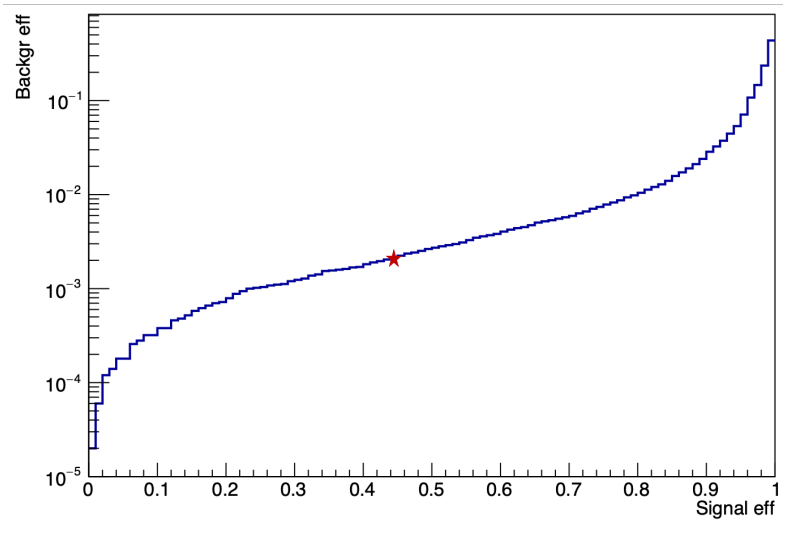


Figure A.2: Receiver Operator Characteristic (ROC) curve for the exclusivity BDT (blue). The x-axis depicts the signal acceptance efficiency, and the y-axis depicts the background acceptance efficiency. Down and to the right is better. The red star indicates the working point anticipated for the usual ± 1 mm exclusivity window.

6. A 0 or 1 value for whether the two closest tracks are on the same or opposite sides of the lepton vertex. This is motivated by the fact that if the tracks come from a pileup interaction, they should both be on one side of the leptons.

This was a Keras neural network, which had two hidden layers of 16 and 8 nodes, respectively, both using tanh activation functions. However, this similarly exhibited no gain in performance relative to the exclusivity window or the BDT.

Ultimately, the most signal-like events have no tracks near the lepton vertex, giving no additional information for a multivariate classifier to use to distinguish between signal and background events. When using a working point that accepts more signal, we will also accept background events where no tracks were reconstructed, either due to general efficiency losses or due to the truth-level underlying charged particles having too low of a p_{T}, resulting in no gain in significance. The use of low-p_{T} tracks is a vital component to improving the statistical significance of this analysis.

Appendix B

Identifying Merged SCT Clusters

Charged particle tracking with the ATLAS detector utilizes the ATLAS Inner Detector (see Section 3.2.2). When charged particles create charge deposits (hits) in adjacent pixels or adjacent strips, the hits are grouped together into clusters. By using the clusters' radial positions and matching clusters on two sides of SCT modules, pixel and strip clusters are interpreted as space-points, and sets of space-points are combined into track-seeds (see Section 4.1). A collection of track-candidates for the event is generated by extending seeds by adding additional clusters from additional inner detector layers using a combinatorial Kalman filter [120]. The final set of tracks for the event is chosen by the ambiguity solving process, where candidates are scored and rejected if they fail to pass a quality threshold. ATLAS tracking is robust: pions at mid-rapidity with p_{T} greater than 1 GeV are reconstructed with greater than 90% efficiency and negligible fake rate. However, performance deteriorates rapidly when particle density approaches the silicon channel segmentation [218, 219]. In high-density environments, such as high p_{T} jets or three-pronged τ-lepton decays, tracks are likely to share inner-detector hits, which leads to an increased likelihood of track candidate rejection unless care is taken to mitigate this effect.

Cluster-sharing is not just a track reconstruction effect. Single clusters (both pixel and strip) can be created by multiple generated particles in a single event [219, 220]. These multiple-particle clusters are called "merged". This merging leads to a decrease in technical efficiency, where, in the context of three-pronged τ decays, technical efficiency is defined as the proportion of events for which all three reconstructible pions are actually reconstructed [220]. Reconstructible here means that the pion traversed and left charge deposits in at least a minimum number of silicon layers. For example, Ref. [220] shows that, in generator-level reconstructible events with minimal strip cluster sharing, pixel cluster merging leads to a decrease in τ reconstruction efficiency from above 99% at low τ p_{T} to about 80% at a τ p_{T} of about 1 TeV.

In response to this efficiency loss, ATLAS has implemented a neural network to identify merged pixel clusters [219]. For the events referenced in the previous paragraph, when tracks are allowed to share pixel clusters that are identified as merged, the τ reconstruction efficiency once again starts at above 99% at low τ p_{T} but only decreases to about 90% at a

τ p_{T} of about 1 TeV.

Prior to the work discussed in this appendix, there was no similar attempt to identify merged strip clusters, with tracks sharing at most two strip clusters. Allowing more sharing improves technical efficiency because more track-candidates are accepted, but this comes at the cost of increased fake rate. The currently allowed level of sharing was chosen to prevent accepting too many fake tracks [220]. In general, a cluster should be used by the same number of reconstructed tracks as the number of particles that deposited energy in the cluster. Thus a multiple-particle cluster should be allowed to be shared, and a single-particle cluster should not be shared. This appendix describes an algorithm developed to identify strip clusters created by multiple particles and the effects of using this information in track-candidate ambiguity solving.

B.1 Identification of Merged Clusters in SCT

The primary inspiration for identifying merged strip clusters is an ATLAS study of δ-rays, which are low-energy electrons knocked out of the silicon lattice from primary particle energy loss [221]. These δ-rays can propagate in the detector, leading to broadened clusters. This study relied on comparing the observed width (W_{o}) of a silicon-detector cluster to the expected width (W_{c}) as calculated using the incidence angle of the track at the cluster in the plane perpendicular to the beam-line, the thickness of the module, and the Lorentz drift angle. Figure B.1 illustrates some of the key parameters of this analysis in the case of δ-rays.

The strip pitch of the ATLAS SCT is 80 μm, so W_{o} is a discrete variable in integral multiples of 80 μm. If 0 $\mu\mathrm{m} > W_{\mathrm{c}} - W_{\mathrm{o}} > -80$ μm, the cluster is as wide as expected; if -80 $\mu\mathrm{m} > W_{\mathrm{c}} - W_{\mathrm{o}} > -160$ μm, the cluster is one strip too wide, and so forth. This effect is also illustrated in Figure B.1, where the actual cluster width is shown in green, but the expected cluster width is one strip. One would expect that a cluster coming from one particle would have no extra strip, barring a δ-ray, diffusion, or other sources of charge sharing, while clusters from more than one generated particles (merged clusters) would have at least one extra strip. The effect of cluster merging on the number of extra strips is illustrated in Figure B.2. This plot shows the $W_{\mathrm{c}} - W_{\mathrm{o}}$ distribution for clusters along the tracks of simulated $|\eta| = 0$, high-p_{T} three-prong τ decays. The sample of τ's has a uniform distribution of p_{T}'s between 400 GeV and 1 TeV and no pile-up events. The figure is created by looping over all the reconstructed tracks in an event with a nested loop over the clusters associated to the respective tracks. Thus, for example, if a cluster is used by two tracks, it will appear in the figure twice, but with different $W_{\mathrm{c}} - W_{\mathrm{o}}$ values, as W_{c} depends on the independent incidence angles of the tracks. About 80% of merged clusters have at least one extra strip, while about 80% of single-particle clusters do not have an extra strip.

Because of this sharp contrast in $W_{\mathrm{c}} - W_{\mathrm{o}}$ between merged and single-particle clusters, this value will be treated as the discriminating cluster-variable. Figure B.3 shows ROC curves from integrating Figure B.2 from the left. Curves are provided for both 400 GeV τ's and 1 TeV τ's, but the p_{T} dependence is not strong. Though there is more merging at high

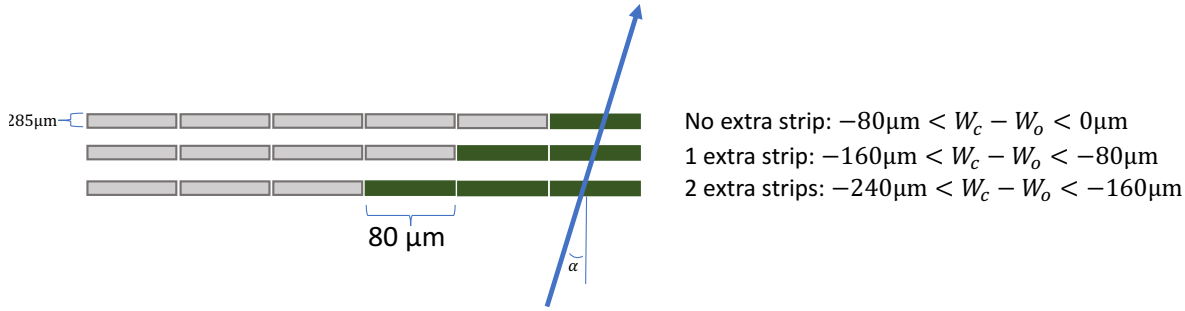


Figure B.1: Illustration of a strip cluster with no extra strips (top), one extra strip (middle), and two extra strips (bottom) and of parameters used to determine the presence of extra strips. The blue pointer, indicates the passage of a charged particle through a strip. The observed width, W_{o} is simply read out from the detector (shown in green). The expected width, W_{c}, is calculated from the incidence angle of the track at the cluster in the plane perpendicular to the beam-line, α, the Lorentz drift angle, $\lambda_{\mathrm{SCT}} = -4°$, and the thickness of the detector $t_{\mathrm{SCT}} = 285$ μm. Because the SCT strip pitch is 80 μm, extra strips correspond to $W_{\mathrm{c}} - W_{\mathrm{o}}$ in integral multiples of 80 μm.

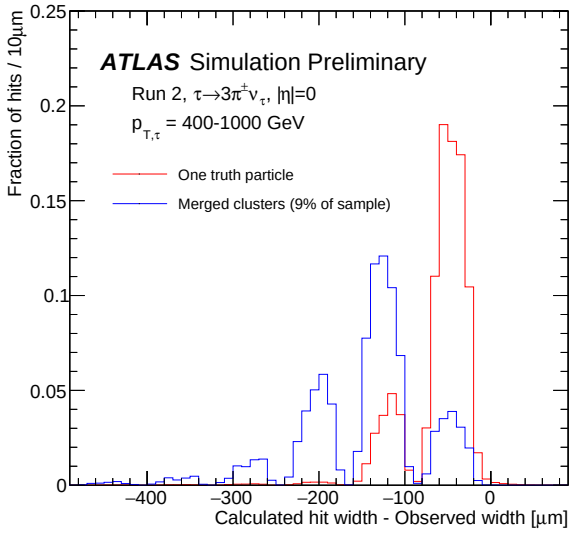


Figure B.2: Illustration of $W_{\mathrm{c}} - W_{\mathrm{o}}$ for clusters in a sample of high-p_{T} three-prong τ decays in the absence of pile-up. The red line is the distribution for single-particle clusters, and the blue line is the distribution for merged clusters.

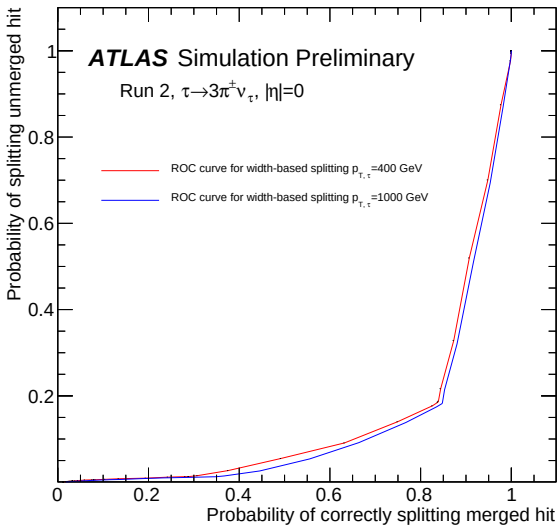


Figure B.3: ROC curves illustrating the discriminating power of $W_c - W_o$ alone in three-prong τ decays. Curves are provided for both 400 GeV τ's and 1 TeV τ's, but the p_T dependence is not strong. The kink at about (0.85, 0.2) corresponds to a cut at $W_c - W_o = -80$ μm.

p_T, $W_c - W_o$ does not depend on the τ p_T. For reference, the pixel cluster neural network is set at a working point where 85–90% of merged clusters are correctly identified, with $<$ 10% of single-particle clusters being incorrectly considered merged.

B.2 Results

To study the effect of identifying merged clusters in the SCT, we have allowed strip clusters to be shared by multiple tracks without penalty if they have at least one extra strip relative to the single track expected width. The change in efficiency for simulated three-prong τ decays is shown in Figure B.4. The green squares represent the default reconstruction efficiency at four different τ p_T values, and an interpolating line is added to guide the eye. The default reconstruction uses the pixel neural network. An event is considered reconstructed if all three of the pions have a reconstructed track matched to a track at generator level [122]. The blue triangles represent the efficiency when the merged strip cluster identification algorithm based on $W_c - W_o$ is implemented. In general, there is a 4–6% relative increase in performance. The red circles show the efficiency when the strip clusters are allowed to be shared based on the clusters' truth-level information, which is essentially a 100% correct to 0% incorrect merging-identification working point. The black inverted triangles show the efficiency when all cluster-sharing penalties are turned off. In this case, most track-candidates are accepted.

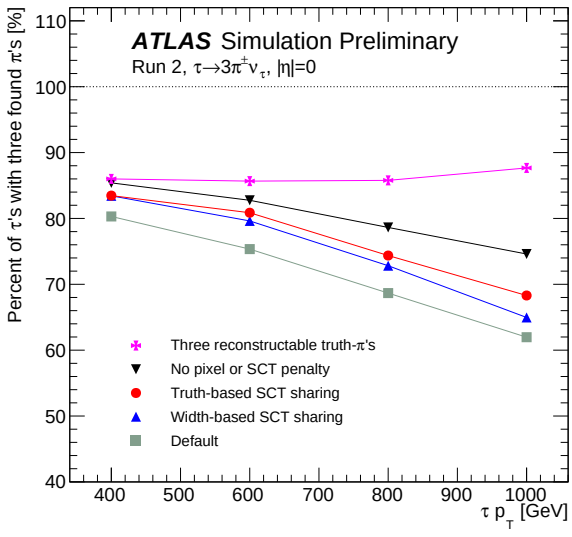


Figure B.4: Curves representing the reconstruction efficiency of the pions coming from a three-pronged τ decay as a function of τ p_{T}. An event is considered reconstructed if all three of the pions have a reconstructed track matched to a track at truth-level [122]. The different marker styles represent different settings in the ambiguity solving stage. Here, "default" uses the pixel neural network. Additionally, the pink crosses represent the proportion of events in which all three pions leave enough hits to be reconstructible.

The pink crosses represent the ideal case: it is the proportion of events where all three pions leave enough inner-detector hits to be reconstructible.

The width-based merging identification approaches the performance of generator-based merging identification, though the performances seem to be increasingly divergent at high τ p_{T}. Furthermore, the no-penalty case is certainly the most efficient. However, as expected from discussions above, this yields a high rate of duplicate tracks. Here, the rate of duplicates is studied as a proxy for the fake rate; if a single generator level pion is matched to more than one reconstructed track, it is considered to have a duplicate. This duplicate-track effect is illustrated in Figure B.5. The same symbols and colors are used as in Figure B.4. Without a cluster-sharing penalty in the ambiguity solving stage, the duplicate rate approaches 40%. While the default case typically stays below 3%, both width-based merging identification and generator-based merging identification stay below 5%.

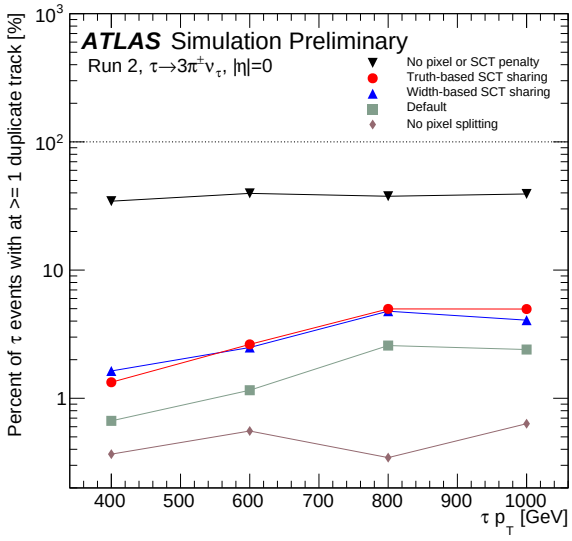


Figure B.5: Curves representing the proportion of events with at least one duplicate track as a function of τ p_{T}. An event has a duplicate track if more than one reconstructed track is matched to any pion at generator level. The different marker styles represent different settings in the ambiguity solving stage. "Default" uses the pixel neural network.

B.3 Conclusions and Outlook

Placing a threshold on the difference between a cluster's calculated width and its observed width, $W_{\mathrm{c}} - W_{\mathrm{o}}$, is an effective means of identifying merged strip clusters in simulation of the ATLAS detector. Allowing tracks to share strip clusters which have at least one extra strip based on this metric increases the reconstruction efficiency in dense environments, though there is a simultaneous increase in the duplicate track rate. These changes can improve searches and measurements at the LHC which use tracks inside high p_{T} τ's and jets.

Appendix C

Identifying Merged Tracks with Machine Learning

Tracking in high density environments, particularly in high energy jets and boosted τ's, plays an important role in many physics analyses at the LHC. In such environments, it is possible that two nearly collinear particles contribute to the same position measurements (hits) as they travel through the ATLAS pixel detector and SCT. To form tracks from hits, the pattern recognition in ATLAS combines groups of three hits into track seeds, which are extended into track candidates by adding hits from additional layers (see Section 4.1). An ambiguity solving procedure is performed to reach a final collection of tracks: track candidates receive a "track score" and candidates with scores below a certain threshold are rejected. As discussed in Appendix B, if a candidate shares hits with an accepted track, its score is reduced.

If the two particles are nearly collinear, it is possible that only a single track candidate will be created (a merged track), leading to a decrease in tracking efficiency. From the study of simulated high momentum $\tau \to 3\pi^{\pm}\,\nu_{\tau}$ decays in Appendix B, it is clear that a significant number of reconstructed tracks in these events are merged, and that there is a loss of tracking efficiency that cannot be recovered by accepting more track candidates. For example, at a τ p_{T} of 800 GeV, about 10% of events have a merged track, while a distinct track is accepted for all three pions in only about 65% of events.

This Appendix details a new technique: using a Boosted Decision Tree (BDT) to classify reconstructed tracks as merged. We will quantify the performance of the BDT and its application to reconstructed events and compare it to other reconstruction schemes.

C.1 Building the Boosted Decision Tree

Because this technique is applied to fully reconstructed tracks, it can access a complete array of track variables: global track parameters, cluster information for the pixel and SCT hits along the track, and local track parameters at the hits. The BDT uses 43 variables, shown

in Table C.1.

Figure C.1 illustrates the separation power of a select few of the 43 variables for merged and single particle tracks. Figure C.1 (top left) shows the normalized distributions of the dE/dx in the last layer of the pixel detector. Figure C.1 (top right) shows the normalized distributions of $|W_{\mathrm{c}} - W_{\mathrm{o}}|$, which is the difference in the expected cluster width based on track incidence angles at the detector element and the actual cluster width, for the cluster with the highest recorded charge on pixel Layer 2. Figure C.1 (bottom) shows the normalized distributions of $|W_{\mathrm{c}} - W_{\mathrm{o}}|$ for the recorded cluster on the outer side of the innermost layer of the SCT with the lowest pull on the track parameters. The tracks considered were reconstructed in samples of single $\tau \to 3\pi^{\pm}\ \nu_{\tau}$ events that were generated at $\eta = 0$. Four samples were used for training and testing, with τ p_{T}'s of 50, 400, 800, and 1000 GeV, each containing the same number of events. The variable values shown here are from τ's with p_{T} of 800 GeV. The variables considered show good discrimination power between the clusters of merged and single particle tracks.

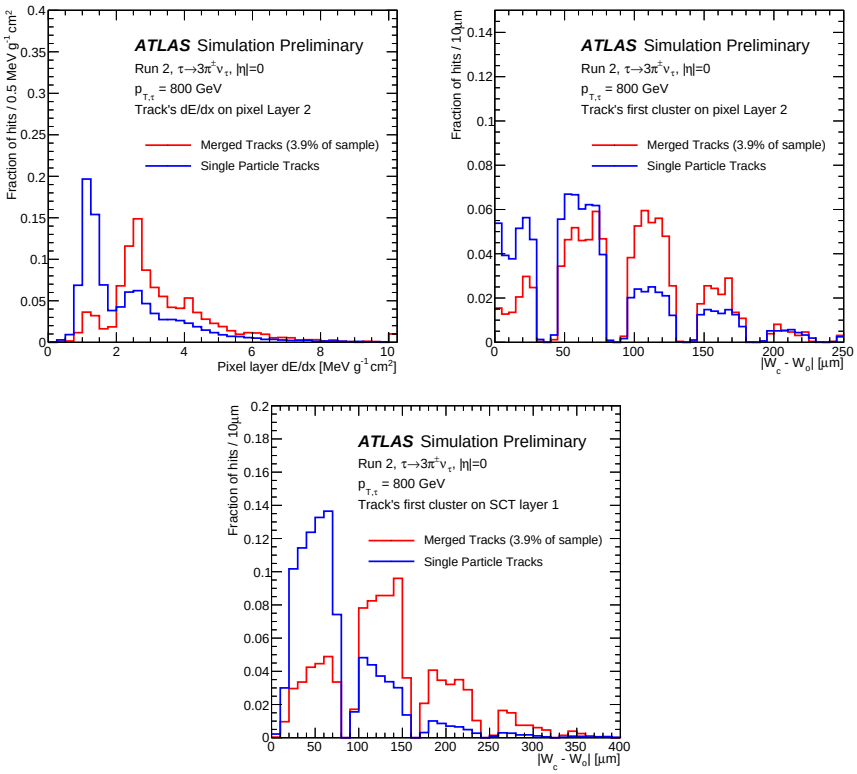


Figure C.1: Illustration of the separation power of select variables that are used in the BDT.

Variable	**Num. of Features**	**Explanation**
Track p_{T}, η, ϕ	3	–
Num. clusters on each pixel layer	4	–
Highest charge deposited in a cluster on each pixel layer	4	Sensor overlap may lead to 2 charge deposits on a single pixel layer.
dE/dx in each pixel layer	4	dE/dx of a track through a layer.
Boolean for whether a hit on each pixel layer is flagged as split	4	Determined by the pixel cluster neural network [219]
$\lvert W_{\mathrm{c}} - W_{\mathrm{o}}\rvert_{\mathrm{pix}}$ for pixel cluster with the highest charge on each pixel layer	4	W_{c} is the expected cluster width (the r-ϕ pitch) calculated from the incidence angle of the reconstructed track at the cluster in the plane perpendicular to the beam-line, the thickness of the module, and the Lorentz drift angle; W_{o} is the observed cluster width in integral multiples of 50 μm, which is the pixel pitch.
$\lvert L_{\mathrm{c}} - L_{\mathrm{o}}\rvert_{\mathrm{pix}}$ for pixel cluster with the highest charge on each pixel layer	4	L_{c} is the expected length in z-direction, and L_{c} is the observed length in z-direction.
Num. clusters on each SCT layer	4	–
$\lvert W_{\mathrm{c}} - W_{\mathrm{o}}\rvert_{\mathrm{SCT}}$ for SCT cluster with the lowest pull on each SCT layer	8	W_{o} is the observed cluster width in integral multiples of 80 μm, which is the SCT strip pitch. W_{c} is the expected cluster width.
Num. shared clusters on each SCT layer	4	A "shared" cluster is one used by multiple reconstructed tracks.

Table C.1: Description of features for a BDT to identify merged tracks. Note that there are four pixel layer and four double-sided SCT layers.

C.2 Training the BDT

Figure C.2 shows the normalized TMVA BDT [217] score distributions for the training sample and a ROC curve with points highlighting different potential cuts on BDT score, where any track with a BDT score above the cut value is flagged as "merged". The tracks considered come from samples of $\tau \rightarrow 3\pi^{\pm}\ \nu_{\tau}$ decays described in Section C.1. Signal is given a training value of 1 and background -1. From Figure C.2 (left), it is clear that the BDT finds a significant difference between merged and single particle tracks; the error bars represent simulation statistics only. True merged tracks are rare, representing about 3% of tracks in the samples used. Because of this, tracks selected with an aggressive cut on the BDT score will be predominantly single particle tracks misidentified as merged.

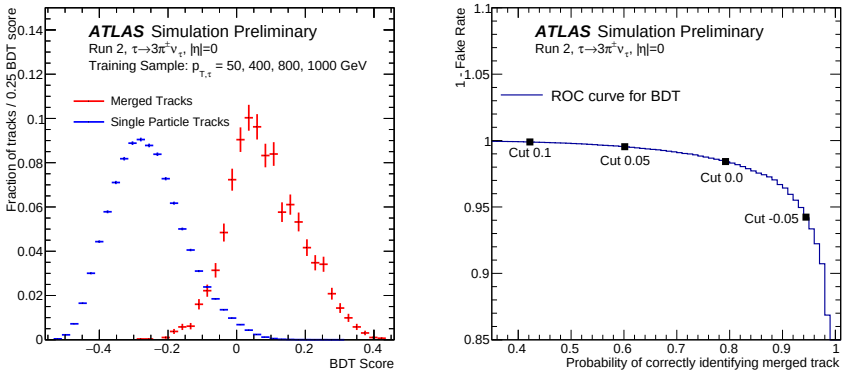


Figure C.2: (Left) Normalized BDT score distributions for the training sample; (right) ROC curve with points highlighting different potential cuts on BDT score.

C.3 Results: Impact on track counting

After training the BDT, we examine it's potential impact on the tracking efficiency. The BDT score is calculated for every track in our sample, and a track is flagged as merged if it has a BDT score above a selected cut value. A τ event is considered fully reconstructed if all three pions from its decay are reconstructed as tracks in the event. Both of the pions contributing to a merged track are considered to be reconstructed if that track is flagged as merged by the BDT. Figure C.3, shows how applying the BDT affects the efficiency for reconstructing τ events as a function of τ p_{T}, both for different BDT cut values (Figure C.3 (top right)) and as a function of track density in the event (Figure C.3 (bottom)).

In Figure C.3 (top left), the pink crosses show the technical reconstruction efficiency, where a pion will be considered "found" if it leaves at least 7 truth hits in the silicon layers of the inner detector. The green triangles show the ATLAS default track reconstruction per-

formance. The pale squares show reconstruction performance using the algorithm described in Appendix B. The inverted triangles show the reconstruction performance when the cluster sharing penalty is turned off. The blue diamonds show the performance when reconstructed tracks are considered to be two pions if they are flagged as merged by the BDT with a cut of 0.1, as described in Section C.2. The BDT is applied after reconstruction. The red circles show the performance if every truth-level merged track is counted as two pions. This filter is also applied after track reconstruction.

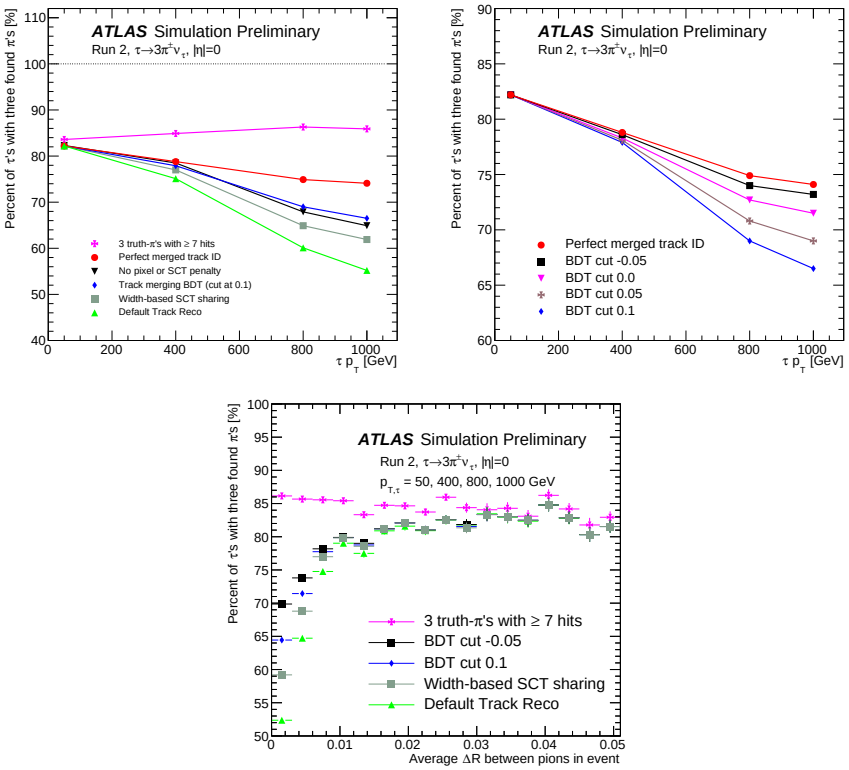


Figure C.3: (Top left) Simulated efficiency of τ reconstruction efficiency as a function of τ p_{T} using different reconstruction configurations (top left) and different BDT cuts (top right). (Bottom) shows the efficiency as a function of track density.

C.4 Mistag rate

Incorrectly tagging a single-particle track as merged would distort an otherwise correct measurement. For example, if merged tracks are counted as two particles with p_{T} equal to the original track, an event with an incorrectly split track could essentially have an extra high-p_{T} track that is unrelated to an actual particle. Thus it is important to find a balance between

increased reconstruction efficiency (as seen in Sec. C.3) and potential mistags. Figure C.4 shows the impact implementing the BDT has on the "duplicate rate" for the same sample used in Figure C.3. An event is considered to have at least one "duplicate" track if at least one pion is truth-matched [206] to more than one reconstructed track. Figure C.4 (left) shows the impact as a function of τ p_{T}, where the line and marker styles are the same as described for Figure C.3. Tracks reconstructed with the algorithm which allows track candidates to share SCT clusters if the cluster has an anomalous $|W_{\mathrm{c}} - W_{\mathrm{o}}|$ have a few percent increase in the duplicate rate over the default algorithm. The BDT was trained and tested on tracks reconstructed using that SCT cluster-sharing algorithm, so the "Track merging BDT" line should be compared to the "Width-based SCT sharing" line. Applying the BDT with a cut of 0.1 creates a negligible increase in the duplicate rate. Figure C.4 (right) illustrates the duplicate rate when using different BDT cuts.

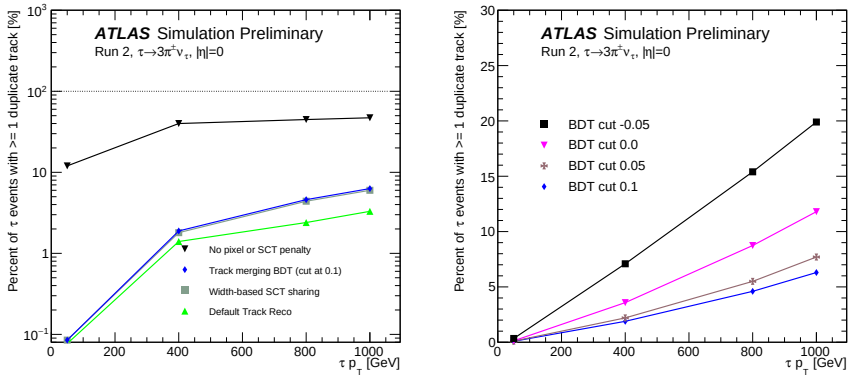


Figure C.4: (Left) Simulated rate of accepting a "duplicate" track as a function of τ p_{T} using different reconstruction configurations (left) and different BDT cuts (right).

C.5 Identifying the particle charges

A natural question to ask is whether the variables fed into the BDT contain enough information to perform a more sophisticated analysis than a simple merged/unmerged classifier. For example, we have created a BDT to distinguish whether the pions in a merged track have the same or opposite electric charge. This secondary BDT uses the same 43 variables as in Table C.1 and is trained on tracks which are flagged as merged by the initial BDT and considered merged at truth level. Figure C.5 demonstrates that it is possible to correctly identify the pion charges with over 85% accuracy for a wide range of τ p_{T}; C.5a shows a ROC curve for correctly identifying a merged track as coming from same-sign or opposite-sign π's, and C.5b shows the charge-tagging efficiency as a function of τ p_{T} for different charge BDT settings. A cut of 0 is the most efficient over a widest range of p_{T}'s.

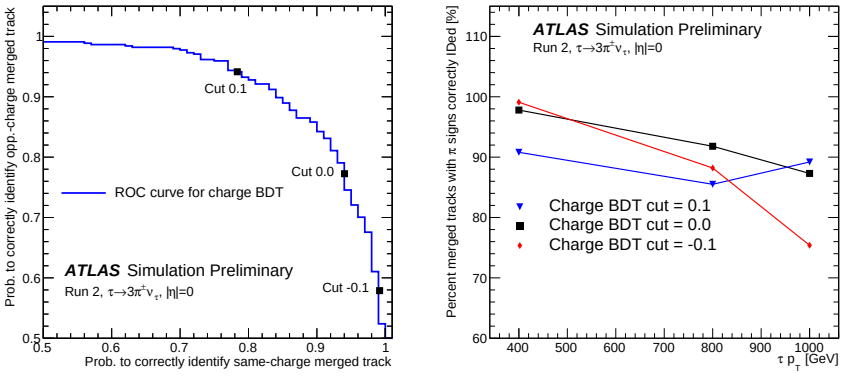


Figure C.5: (Left) ROC curve for the charge BDT; (right) efficiency for correctly identifying charges in a merged track as a function of τ p_{T}.

C.6 Conclusions and Outlook

This Appendix demonstrated a new technique for finding tracks in the ATLAS detector that are created by two nearly collinear particles. Using a BDT, one can find about 50% of merged tracks while increasing the duplicate rate at the sub-percent level. Making statements about the impact on measureables such as τ tagging or jet p_{T} resolution requires further study.

Appendix D

Development of a Pixel-Cluster Counting Algorithm for Luminosity Measurements

A critical component of all measurements with ATLAS is knowledge of the delivered luminosity. It is the multiplicative factor that converts an experimentally observed event yield into a theoretically calculated cross-section. In Run 2, ATLAS recorded 139 fb^{-1} of good data, with a 1.7% uncertainty on that value.

The luminosity determination in ATLAS is performed redundantly with several independent detector components to reduce the impact of systematic uncertainties and potential biases. Each subsystem is designed to monitor some quantity that scales linearly with luminosity. The main subsystems used are:

- The dedicated Luminosity Cherenkov Integrating Detector-2 (LUCID) [84]. This system is 20 photomultiplier tubes on both sides of the detector, located circularly around and just outside of the beampipe. Charged particles produce Cherenkov light when they pass through the quartz windows of the photomultiplier tubes. This detector system can be used to both count individual particles and integrate accumulated signal.

- The Beam Conditions Monitor (BCM) [222]. This is a diamond detector, which is similar manner to, but more radiation hard, than a silicon detector. This subsystem is also on both sides of the detector, at ± 1.8 m in z and a radius of 55 mm. It is designed to count the number of charged particles that pass through it.

- The Inner Detector is used to reconstruct tracks (see Section 4.1), and the number of tracks can be counted.

- The integrated signal created in the ATLAS calorimeters (see Chapter 3), such as the integrated photomultiplier currents in the tile hadronic calorimeter (TILE), should scale with luminosity.

Absolute luminosity calibrations are performed with the van der Meer scan method [223], using a special low-pileup run to determine the beam profile factors found in Equation 3.1.

This appendix details the initial development of another luminometer for ATLAS based on pixel cluster counting (PCC) [224]. Where to the track counting method used all layers of the Inner Detector, PCC uses the forward-most modules of the IBL to count the number of pixel clusters per event, where a pixel cluster is a group of adjacent hit pixels. Because only the forward modules of the IBL are used, the PCC algorithm detailed here uses different detector components than the track counting method. If each cluster corresponds to one primary particle, then the number of clusters should scale linearly with luminosity; in fact, the CMS Collaboration uses PCC as its baseline luminosity monitoring algorithm [225].

D.1 Determining Which Clusters to Count

Unfortunately, not all hits in the pixel detector, and therefore not all pixel clusters, are caused by primary particles. In any module, there is a background of clusters caused by other sources, such as secondaries from detector material, beam backgrounds, and non-collision backgrounds such as afterglow, which are out-of-time hits from radiologically activated detector material. The number of pixel clusters resulting from these sources is *not* expected to scale linearly with luminosity.

Luckily, we can distinguish between the background clusters and signal clusters created by primary particles. Primary particles originate at the interaction point, so they will pass through the forward modules of the IBL, which are about 30 cm away in the z-direction, at a shallow angle, leaving a long clusters. Therefore, this PCC method is based on fitting cluster-length distributions, such as that shown in Figure D.1. The per-module distribution is fit with a two-component function consisting of a Gaussian signal and an exponentially falling background. The Gaussian area of the fit scales linearly with luminosity, after correcting for inefficient/hot modules and beam spot position and width, which will be detailed below.

D.2 Afterglow

During LHC operation, the IBL detector material is activated by pp collision debris. The resulting low energy decay products, collectively known as afterglow, induce delayed signals in the detector and produce single or small pixel clusters, thereby contributing to the background component of the cluster-length distributions rather than to the Gaussian-shaped long-cluster signal.

A dedicated ATLAS run was performed to study this effect. Normally, there are 25ns between bunch crossings at the LHC, leading to a build up of activated material in the detector, but in the specialized run, one filled bunch crossing was followed by 7 empty crossings where detector data was collected. Figure D.2 shows the fit components of the cluster-length distribution for the filled bunch crossing and the subsequent empty crossings.

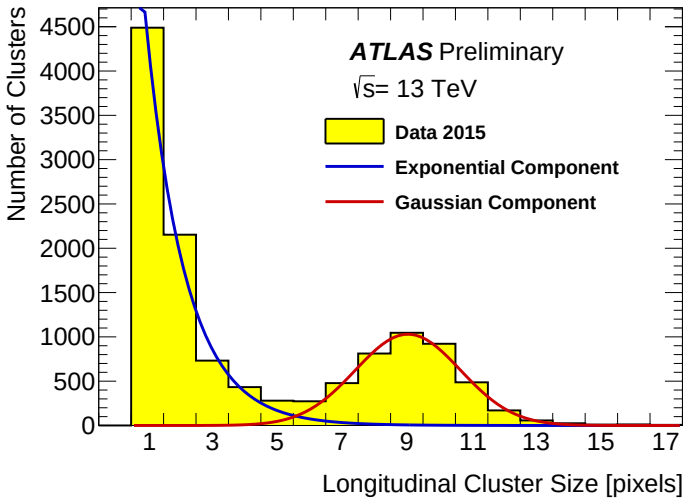


Figure D.1: Typical distribution of the pixel-cluster size in the direction parallel to the beam axis, in the second-most forward IBL module. The 12413 clusters that populate this plot originate from 10433 randomly-triggered events during ATLAS physics data-taking in 2015. The distribution is fit to the sum of a Gaussian signal, plus an exponentially-falling background of shorter clusters. The mean of the Guassian from the fit is a cluster length of 9-pixels, and its area is proportional to the number of charged particles originating from the luminous region. The exponential area contains contributions from secondary interactions and afterglow that will deviate from linear proportionality to luminosity.

The measured signal component is largest in the first bunch slot, and drops by 4 orders of magnitude within 25 ns. The measured background component, in contrast, decays much more slowly, with a time dependence suggestive of a mixture of radioactive decays. We have not attempted to identify those decays in this study.

D.3 Conversion of Gaussian Area to Luminosity

Several effects must be corrected for in order to convert the signal area to luminosity:

1. Module performance issues (for example a module being disabled for part of a run or a module being noisy) can introduce a bias. A data driven method is used to remove this potential bias. We fit the distribution of signal areas for all IBL 3-D modules at the same position along the beam line to a sinusoidal curve of the form

$$A * \cos\left(\frac{2\pi}{14}x + B\right) + C. \tag{D.1}$$

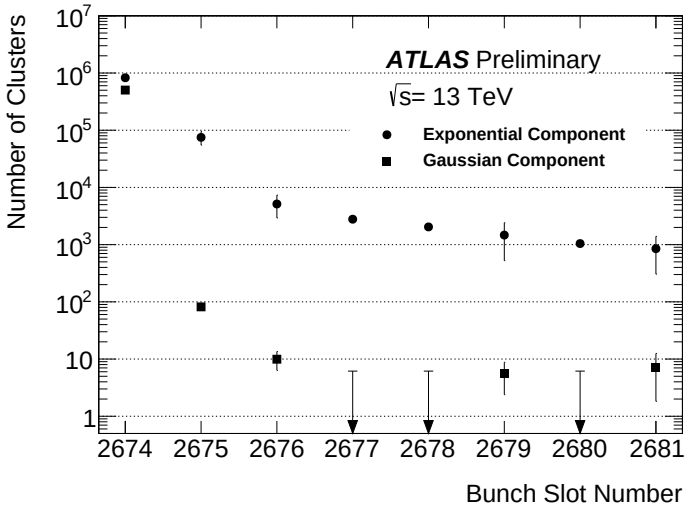


Figure D.2: Bunch-slot dependence of the signal and background components of the Gaussian (signal) and exponential (background) fit of a cluster length distributions. These signal and background levels are recorded in 8 consecutive 25 ns-long bunch slots. Only the first slot (numbered 2674) contained a colliding-bunch pair, while the remaining 7 were nominally empty.

Here, A is an overall normalization, and B and C account for the horizontal and vertical displacement of the beam spot relative to the IBL center. Rather than adding the individual module signal areas, we exclude outliers, re-fit, and use the sinusoidal normalization as the corrected sum of the module signal areas, rather than the sum of the 14 individual measurements. Figure D.3 illustrates such a fit.

2. If the luminous region is not longitudinally centered in the IBL, a forward-backward asymmetry appears in the signal areas, because the acceptance of each module depends on its distance from the average collision point. Figure D.4 shows the influence of beamspot position on this acceptance for the forward and backward modules and for all modules considered. The asymmetry can be used to calculate the longitudinal position of the beamspot through a linear transformation. Figure D.5 demonstrates the agreement between the asymmetry-derived position and the ATLAS offline calculation. Because the beamspot position influences module acceptance nonlinearly, the normalization of the sinusoidal fit from point 1 above will change as the beam spot moves. We correct for this purely geometric effect using a quadratic function of the asymmetry-derived beamspot position. Thus this correction is also data-driven, requiring no external information.

3. Even if the beam spot center does not move, a longer luminous region will produce a

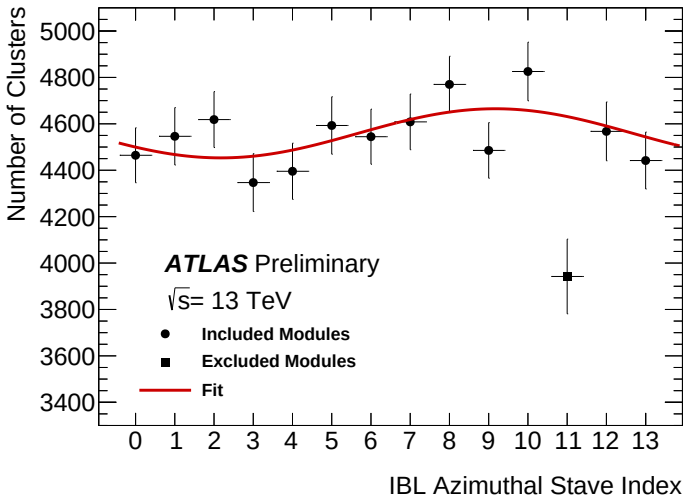


Figure D.3: Azimuthal dependence of the long-cluster (Gaussian) signal for the forward-most modules of the IBL. The sinusoidal dependence arises because in the plane perpendicular to the average beam direction, the luminous region is not perfectly centered on the IBL's central axis. In order to avoid being affected by occasional underperforming or noisy modules, the measured azimuthal dependence of the long-cluster signal is fit to a sinusoidal function. Outliers are excluded from the fit on the basis of internal consistency. In this example, the number 12 module was inefficient during the run considered and was excluded from the fit because its residual exceeded 4 sigma.

greater signal area, again due to geometric acceptance. The correction for luminous length uses a linear function of the beam spot width obtained from track reconstruction. Figure D.6 illustrates the correlation between beamspot length and signal-region clusters.

After corrections the measured luminosity is proportional to

$$\mathcal{L} \propto \frac{N_S}{1 + C_1 * R_o^2 + C_2 * \sigma_z}, \tag{D.2}$$

where N_S is the normalization of the sinusoidal to signal areas as in Figure D.3, R_o is the "offset ratio" as in Figure D.5, σ_z is the beam spot width, and C_1 and C_2 are geometric constants that can be determined from simulation.

Luminosity algorithms must be stable as a function of both time and pileup. Our validation of these conditions relies on comparison of different algorithms using different subdetectors and therefore different sources of potential bias. Figures D.7 and D.8 compare the new PCC algorithm to the luminometers currently used by ATLAS. The PCC luminosity

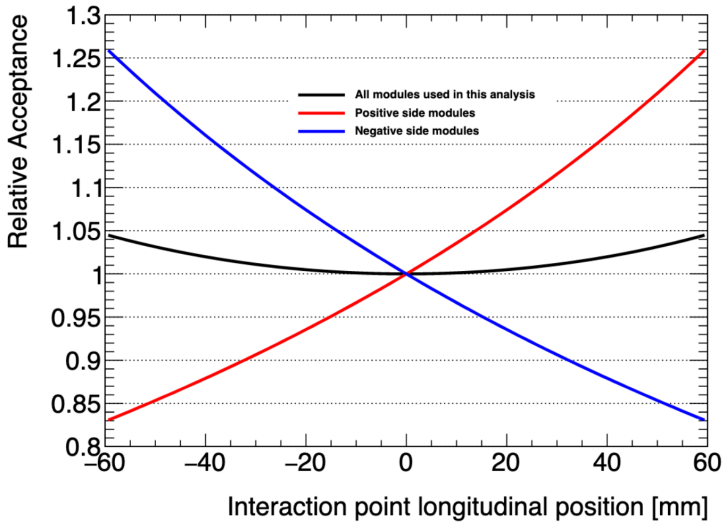


Figure D.4: Acceptance correction to the signal region in the modules used by the PCC algorithm as a function of the beamspot position.

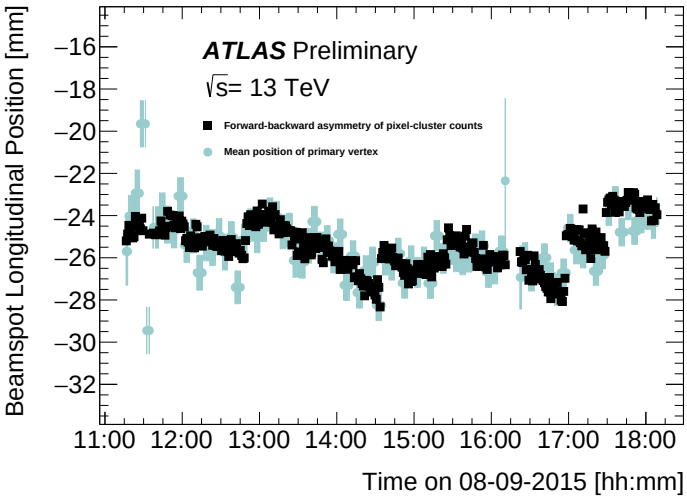


Figure D.5: Time evolution, during an ATLAS run, of the longitudinal position of the luminous region either as determined from the average position of reconstructed pp-collision vertices (light blue circles), or as inferred from the forward-backward asymmetry of the pixel-cluster counts in the IBL (black squares). The asymmetry is computed from the long-cluster signals in the 4 outermost IBL modules on either side of the interaction point. The pixel-cluster-based luminosity measurement can be corrected for this geometric effect by using the beamspot position derived from the measured asymmetry of the pixel-cluster signal.

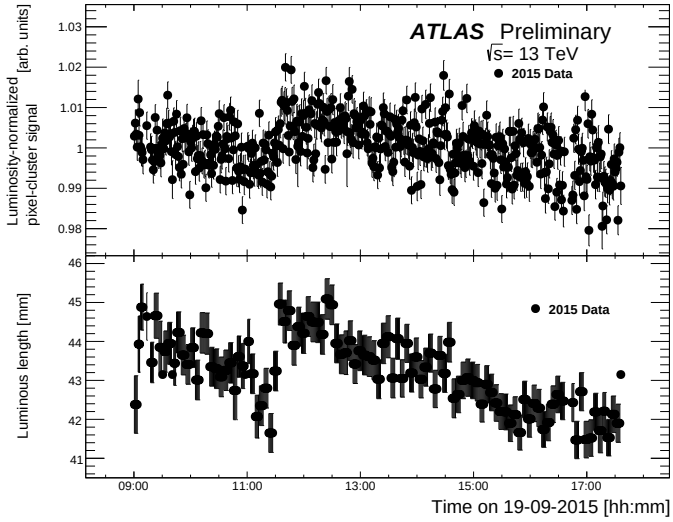


Figure D.6: Time evolution, during an ATLAS run, of the pixel-cluster luminosity signal normalized to the luminosity as measured by LUCID (top), and of the RMS luminous length inferred from the longitudinal distribution of reconstructed pp-collision vertices (bottom). The luminosity-normalized pixel-cluster signal exhibits a similar time dependence to that of the luminous length, because the acceptance of each module depends on its distance from the corresponding interaction vertex.

measurement is stable with respect to the other ATLAS measurements to within $\pm 1\%$ or better, and is therefore shown to be a competitive relative-luminosity monitoring method.

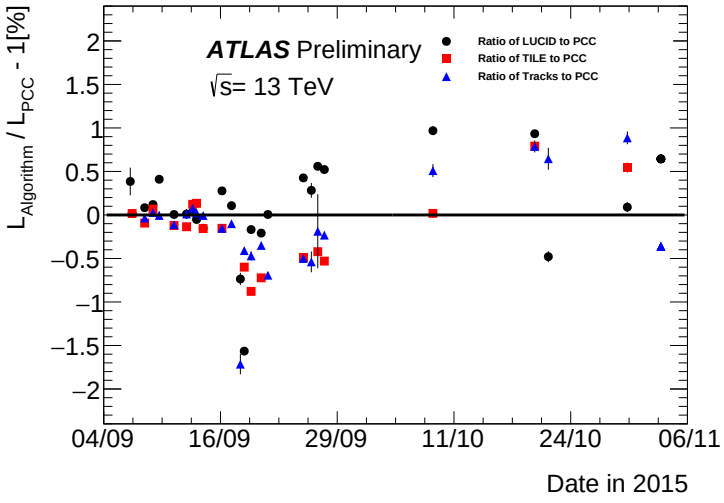


Figure D.7: Fractional difference in run-integrated luminosity between the PCC algorithm and the LUCID, TILE, and track-counting algorithms, for the subset of ATLAS runs for which the PCC data are available. The absolute scale of the PCC-based luminosity is cross-calibrated to the luminosity from the comparison algorithm averaged over the runs taken from 11 to 13 September 2015.

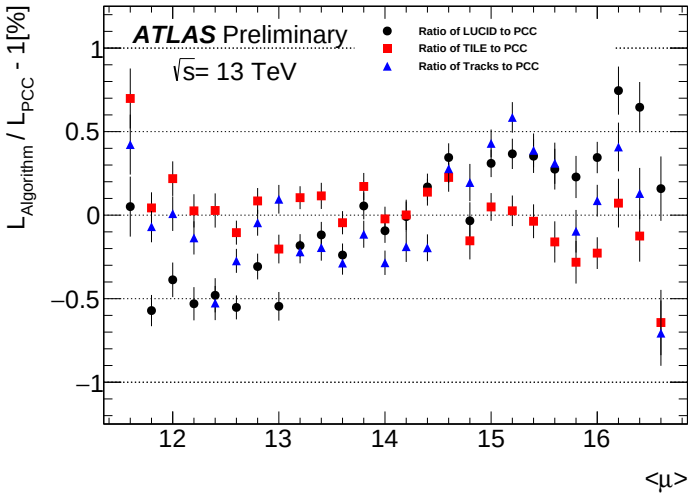


Figure D.8: Fractional deviation of the bunch-averaged pile-up parameter average, μ, obtained using different algorithms, from the PCC value, as a function of average μ during a physics run on September 25, 2015. The data are normalized such that all algorithms yield the same integrated luminosity in the run considered.

Appendix E

Quark-Gluon Jet Discrimination with Iterative Soft-Drop Multiplicity

Jets, which are collimated sprays of hadrons, were discussed in Section 4.4, but were otherwise not used in the $\gamma\gamma \to W^+W^-$ analysis, outside of a study on top quark processes (see Section 9.2). However, that does not imply that jets are not interesting. This Appendix details a study of variables for tagging jets as either quark-initiated or gluon-initiated. At a theoretical level, the concept of jet "initiation" is rather subtle, but here we can just imagine that a final state quark in a Feynman diagram will turn into a quark-initiated jet ("quark jet"), and a final state gluon turns into a gluon-initiated jet ("gluon jet").

When attempting to discriminate quark and gluon jets, one of the most powerful discriminants is the number of constituents inside the jet: gluon jets tend to produce more particles than quark jets [226]. However, there are two important issues must be confronted:

- Multiplicity observables are highly pileup dependent. Jets are constructed primarily from calorimeter information, and a pileup particle that passes through the calorimeter at the same location as the particles from an independent collision's jet will contribute to that jet's multiplicity. Track multiplicity is less dependent on pileup, due to point-of-origin information.

- Constituent multiplicity is not infrared (IR) safe, and the related quantity of track multiplicity is neither infrared nor collinear (IRC) safe [227]. This means that these observables are both sensitive to low-energy branchings of particles in the jet development process, and track multiplicity is sensitive to branchings where the branching particles are collinear with each other. Such branchings are associated with divergences in field theory calculations, so IR and IRC unsafe observables cannot be computed from first principles. Observables that *are* IRC safe are typically limited in power due to Casimir scaling [228].

A popular existing means of pileup mitigation is Constituent Subtraction (CS) [229, 230, 231]. This method removes constituents from jets if the constituent has energy that roughly

corresponds to the expected contribution of pileup in the area it spans.

An alternate strategy, which will be explored in this appendix, is "jet grooming". Grooming algorithms systematically remove soft and wide angle radiation from jets, altering their structure even in the absence of pileup. The algorithm explored here is called "Soft-Drop" [232]; it will be described in more detail in Section E.1.

When applying the Soft-Drop algorithm, we can count the number of branchings in a jet that *pass* the soft-drop condition. This constitutes a new IRC-safe multiplicity observable that can be used to discriminate quark and gluon jets, called Iterative Soft-Drop Multiplicity [228]. This will also be described in Section E.1.

E.1 Multiplicity Algorithm Definitions

In this simulation-level study, jets are initially formed from calorimeter clusters using the the anti-k_t algorithm [134] with radius variable $R = 0.8$. These jets are implemented with FastJet [233]. The multiplicity definitions considered throughout this appendix are explained in the Subsections E.1.1, E.1.2, E.1.4, and E.1.5.

E.1.1 Calorimeter-Cluster Multiplicity – Number of Clusters

After jet finding, the raw jet constituent multiplicity is simply the number of calorimeter cell clusters inside a jet.

E.1.2 Track Multiplicity

Tracks with $p_{\mathrm{T}} > 500$ MeV and $|\eta| < 2.5$ are associated to the jet via ghost association [234]. The number of badly reconstructed or combinatoric tracks is reduced while maintaining high efficiency by applying a loose selection on the number of hits on the track [122]. To reject pile-up, tracks are required to have $|z_0 \sin(\theta)| < 3$ mm, where z_0 is the impact parameter – the distance of closest approach to the primary vertex – along the z-axis. Additionally, the transverse impact parameter is required to be at most 2 mm.

E.1.3 Grooming

The jet grooming techniques considered here begin by re-clustering the constituents of the jet with the Cambridge–Aachen [235, 236] (C/A) algorithm in order to create a clustering tree with greater emphasis on angular ordering. This algorithm combines proto-jets with the smallest ΔR_{ij} until all proto-jets are separated by a set radial difference. The indices i and j refer to the proto-jets being compared. The soft-drop algorithm is then applied in two different ways, using either all of the jet constituent clusters as inputs or clusters surviving the CS algorithm.

E.1.4 Soft-Drop Declustered Multiplicity – Number of Clusters after Grooming

1. Undo the last clustering step.
2. If $z_{ij} < z_{\text{cut}}(\Delta R_{ij}/R)^{\beta}$, remove the softer proto-jet and repeat on the harder proto-jet, where R, the jet radius, is 0.8, and
$$z_{ij} = \frac{\min(p_{\text{T}i}, p_{\text{T}j})}{p_{\text{T}i} + p_{\text{T}j}}. \tag{E.1}$$
Otherwise, the algorithm terminates and both proto-jets are preserved.
3. In the end, if all other proto-jets have been dropped, the last remaining proto-jet is also removed from consideration. Thus, if the soft-drop condition is never satisfied, the multiplicity is zero. Otherwise, the soft-drop declustered multiplicity is the number of constituents inside the remaining proto-jets inside the jet.

For the soft-drop declustered multiplicity, the parameters $z_{\text{cut}} = 0.1$ and $\beta = 0$ were chosen to preserve IRC safety and to follow the standard in [232].

E.1.5 ISD Multiplicity – Number of Splittings while Grooming

1. Undo the last clustering step.
2. If $\Delta R_{ij} < \Delta R_{\text{cut}}$, the algorithm terminates.
3. If $z_{ij} > z_{\text{cut}}(\Delta R_{ij}/R)^{\beta}$, increment the ISD constituent multiplicity by one and iterate on the harder proto-jet. If not, iterate on the harder proto-jet without incrementing.

The parameters z_{cut}, β, and ΔR_{cut} are algorithm-specific. For the ISD Multiplicity, the parameters $z_{\text{cut}} = 0.007$, $\beta = -1$, and $\Delta R_{\text{cut}} = 0$ were chosen, which are the values found to maximize the quark-versus-gluon discrimination power while maintaining calculability in [228].

E.2 Simulated Datasets

Multijet events within the ATLAS detector are simulated using Pythia 8.2 [178] with the A14 set of tuned underlying event parameters [177]. Simulation was performed as described in Chapter 5 [182, 181]. Pileup was simulated with Pythia 8.1, using the MSTW2008LO PDF set [237] and A2 set of tuned underlying event parameters [238]. The simulation accounts for signal pulses in the detector caused by preceding bunch crossings, which can also contribute to the pileup contamination of jets. The mean number of interactions per bunch crossing in these events, μ, is 33. Events were required to have at least two jets. In order to study the regime of interest for data, in which there is a jet trigger, at least one of the jets must have $p_{\text{T}} > 500$ GeV.

E.3 Results

The various jet constituent multiplicity distributions are presented in Figure E.1 (left). Multiplicity definitions that involve grooming are peaked at lower values compared to the ungroomed definitions due to the systematic removal of jet constituents. While this is true too for the soft-drop declustered multiplicity, the movement of its peak is slight, and the lines for the ungroomed and groomed versions largely overlap. The main result on the pile-up dependence of the average multiplicities is shown in Figure E.1 (right). In order to show the curves on the same scale, each multiplicity definition is normalized to unity at $\mu = 30$, where μ quantifies the level of pile-up. All constituent multiplicity definitions shift towards higher values with increasing pile-up. For most definitions, the pile-up dependence is approximately linear, but calorimeter-cluster multiplicity after constituent subtraction exhibits a non-linear turn-on at low pile-up values. This may be due to poor pile-up estimation for the constituent subtraction algorithm in that regime. Prior to grooming, the track multiplicity is the most stable with respect to the level of pile-up, while calorimeter-based multiplicity is the most sensitive. With the parameters chosen here, ISD multiplicity is less robust against pile-up than soft-drop declustered multiplicity. The application of Constituent Subtraction is found to reduce the pile-up dependence of these calorimeter-based observables, which approaches or exceeds the stability of track multiplicity for the selections described here. The slopes found from fitting a line to the full range of the pile-up dependence are given in the legend of Figure E.1 (right).

To illustrate the discrimination power of these multiplicity definitions, receiver operating characteristic (ROC) curves for correct quark-jet tagging are shown in Figure E.2. The figure is divided into two rows, with the top showing ROC curves for low-μ and the bottom showing curves for high-μ events. ISD multiplicity is shown separately from the other calorimeter-based multiplicity definitions for clarity. This was originally proposed as an IRC-safe quark-versus-gluon jet discriminant that can evade the Casimir scaling limit [228], and its performance does exceed that limit. The figure presents ROC curves for the multiplicity definitions using both raw and constituent-subtracted reconstructed jets. While performing constituent subtraction clearly improves pile-up stability, the quark-versus-gluon discrimination seems to be worsened for calorimeter-cluster based multiplicity definitions and improved slightly for ISD multiplicity. In all cases, the discrimination power is worse for the higher-μ region.

E.4 Conclusions and Outlook

This appendix explored the pile-up dependence of various definitions of jet constituent multiplicity – observables that are particularly sensitive to pile-up – and examples of the quark-versus-gluon discrimination power of these various definitions in low- and high-μ events. Prior to performing constituent subtraction on jets, track multiplicity exhibits the most stability against pile-up, and calorimeter-cluster multiplicity exhibits the least; similarly, while ISD

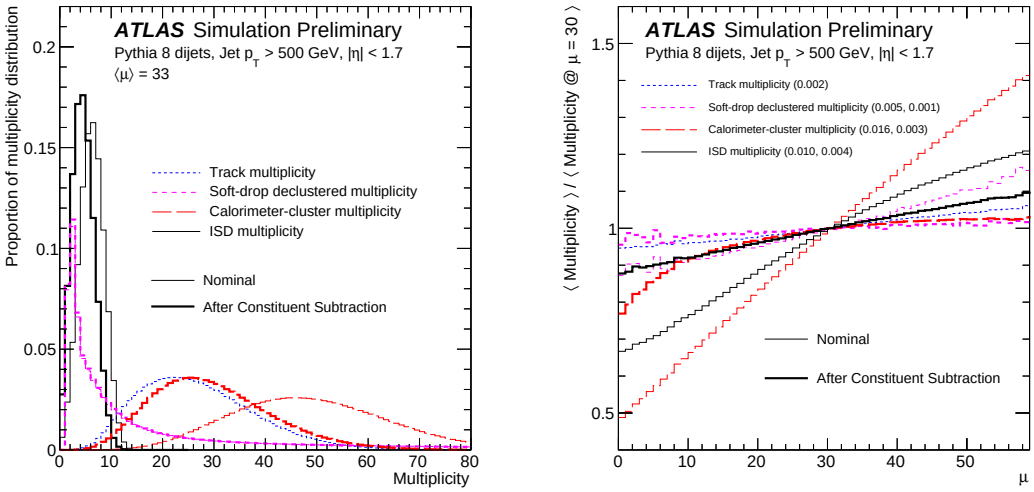


Figure E.1: (Left) Pile-up-inclusive multiplicity distributions of selected counting jet-measurables; the average μ value is 33. (Right) Pile-up dependence of the measurables. Track multiplicity, raw calorimeter-cluster multiplicity, soft-drop declustered cluster multiplicity, and ISD multiplicity are shown. For the calorimeter-based multiplicity definitions, the performance is shown for nominal jet constituents as well as for constituents after the application of Constituent Subtraction. The slopes of a straight line fit to the pile-up dependence curves are presented in the legend in the format (Nominal, After Constituent Subtraction). The full pile-up range was used for the fit.

multiplicity with the chosen parameters depends significantly on pile-up, calorimeter-cluster multiplicity after soft-drop declustering is more stable. Applying constituent subtraction significantly improves the stability against pile-up of all calorimeter-based constituent multiplicity definitions. Furthermore, constituent subtraction does not appear to have a strong effect on the quark-versus-gluon discrimination power of ISD multiplicity, though it has a mild effect on the calorimeter-cluster based definitions. Despite its pile-up dependence, ISD multiplicity surpasses the Casimir Scaling limit at moderate quark efficiency, as designed, performing better than other calorimeter-based multiplicity definitions. Grooming provides an interesting avenue for including a calorimeter-based definition of multiplicity in future quark-versus-gluon jet taggers. Traditionally, the track-based multiplicity has been used by ATLAS for quark-versus-gluon jet tagging; in the future, it may be possible to complement this information with a proxy for the neutral multiplicity as well.

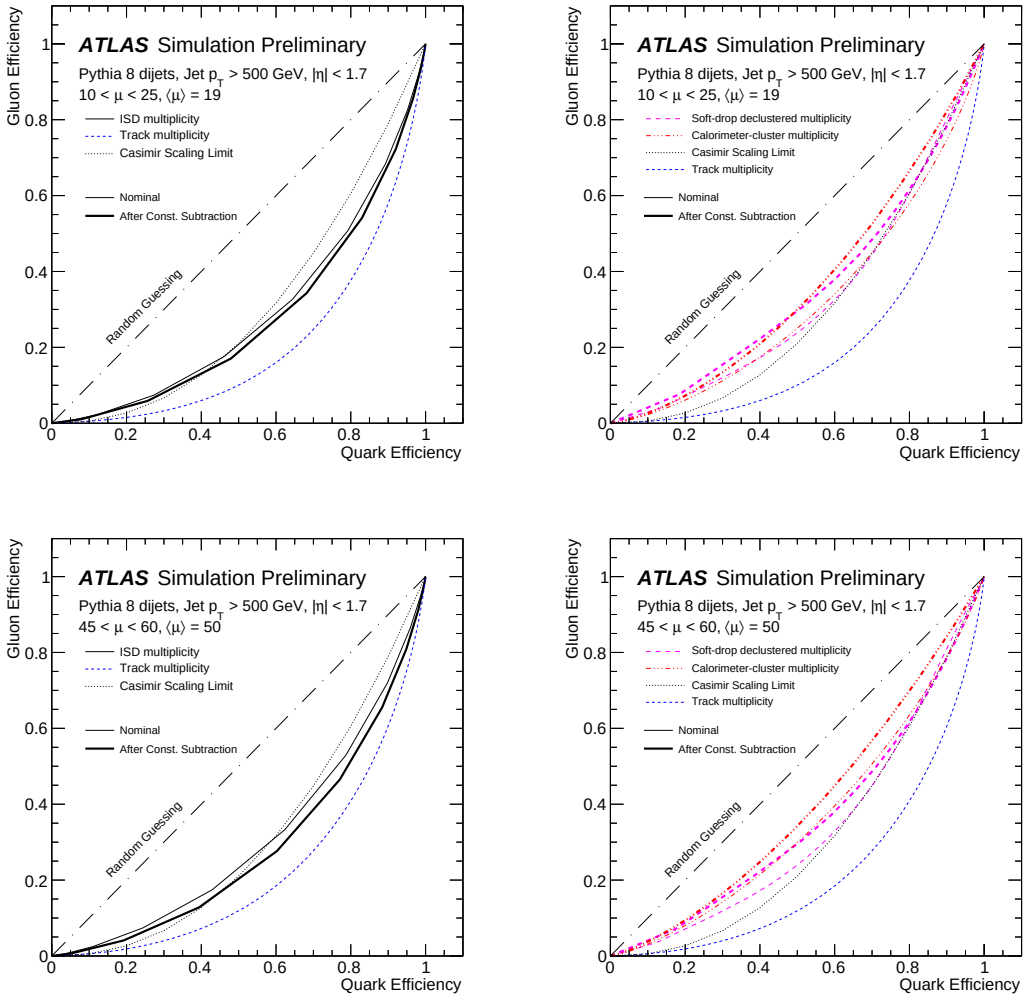


Figure E.2: ROC curves for quark-versus-gluon discrimination using the considered multiplicity definitions. ISD multiplicity is separated from the other calorimeter-based definitions for visual clarity, and lines for no-discrimination (straight), the Casimir Scaling limit (curved), and track-multiplicity based discrimination (blue dashes) are provided for reference. (Top left) ROC curve for ISD multiplicity using both raw and constituent-subtracted reconstructed jets in low-μ events. (Top right) ROC curve for calorimeter-cluster and soft-drop declustered multiplicity using both raw and constituent-subtracted reconstructed jets in low-μ events. (Bottom left) ROC curve for ISD multiplicity using both raw and constituent-subtracted reconstructed jets in high-μ events. (Bottom right) ROC curve for calorimeter-cluster and soft-drop declustered multiplicity using both raw and constituent-subtracted reconstructed jets in high-μ events.

Appendix F

Absolute Calibration of Silicon Sensors using Compton Scattering

F.1 Motivation and Requirements for Absolute Charge Calibration

Silicon pixel detectors are an important technology for experiments at the Large Hadron Collider (LHC) at CERN due to their radiation hardness and readout rate. As seen in Chapter 3, they often comprise the innermost elements of the detectors, being used for tracking and vertexing. Hybrid pixel and monolithic active pixel silicon (MAPS) detectors are currently prominent silicon technologies used by CMS, ATLAS, and ALICE [239, 240, 241].

The use of increasingly thin sensors in silicon trackers is fueled by multiple reasons: lower mass and higher radiation tolerance for hybrid detectors [242] and the increasing use of monolithic detectors [241, 243]. Furthermore, radiation damage decreases the charge collection for a given thickness, and the need to push silicon sensor radiation tolerance results in readout electronics sensitive to ever lower signals. Currently, the thicknesses of sensors used in these experiments' detectors range from 200 μm to 285 μm, depending on the experiment and location in the detector (as in Section 3.2.2). Planned upgrades to detectors will use even thinner sensors. For example, the ATLAS ITk hybrid pixel detector [244, 245] will include 100 μm thick planar sensors with an end-of-life average charge per minimum ionizing particle (MIP) of 6000 electron-hole (e-h) pairs or fewer at normal incidence. The equivalent value for the ALICE ITS [241] inner layer monolithic sensors, which will have a 25 μm thick active region, is 1600 e-h pairs or fewer. In this appendix, a conversion of 3.6 eV per e-h pair will be assumed [246].

To achieve maximum detection efficiency for lower-energy hits, operating thresholds are set to be as low as possible while avoiding noise occupancy. Throughout this appendix, the "threshold" can be interpreted as the energy deposit which would be detected 50% of the time. A slightly smaller energy deposit would be detected less frequently and a slightly larger

deposit more frequently. Pixel detectors are typically operated with thresholds of order 10 ENC, where ENC is the Equivalent Noise Charge obtained by converting the amplifier noise to units of input charge. Current pixel detectors in ATLAS and CMS typically operate with a hit threshold between 2000 e-h pairs and 5000 e-h pairs (7.2 - 18.0 keV), depending on the detector layer and year [247, 248, 249]. For the next generation of pixel detectors, thresholds may be in the range of 100 e-h pairs to 1700 e-h pairs (0.4 - 6 keV).

Silicon tracking detectors typically have an internal charge injection circuit for in situ tuning and threshold verification. However, this injection circuit must be validated with an external source to be considered accurate. The external source makes an absolute calibration possible by depositing a known amount of charge in the sensor. Throughout this appendix, the units for the nominal charge injected by the internal circuit will be denoted as $\tilde{e}$ in order to distinguish from true charge, which will be denoted as "e-h pairs". By performing a calibration, a function is obtained that converts nominal injected charge to true charge: $\mathcal{F} : \tilde{e} \rightarrow$ e-h pairs. Beyond absolute charge scale calibration, external sources may be necessary to check the uniformity of the injection circuit across a front-end. Additionally, there can be uncertainties in the charge collection from the sensor after irradiation, and any method other than generation of a known signal in the sensor itself will be subject to those uncertainties.

Historically, such absolute calibrations have relied on techniques such as X-ray absorption from fluorescence [250, 251, 252] or the absorption of radiation from radioisotope sources with known gamma- or X-ray peaks [253]. Studies are also done using the energy loss of minimum ionizing particles (MIPs) to provide qualitative comparisons and find outliers [254]. For a variety of reasons, these techniques are not always adequate below 6 keV:

- Irradiating a sensor with X-rays between 1 and 6 keV is possible using the K_α and K_β lines of the elements Na through Mn [255]. However, this requires the use of expensive equipment and the use of multiple elemental sources.

- There is no typically used radioisotope source with an energy peak below that of ^{55}Fe's 5.9 keV X-ray line. Other typical radioisotopes would be ^{241}Am (59.5 keV line) and ^{109}Cd (22 and 25 keV lines) [256].

- A MIP signal is not monochromatic and does not lend itself to characterization of the detector threshold behavior or the response function. MIP calibrations are also often not practical.

With these problems in mind, there are several practical goals and advantages of the calibration method explained in this appendix, which is based on Compton scattering. It should provide access to energies between 1 and 6 keV by using commonly available monochromatic photon sources. The calibration signal itself will be monochromatic, but a single setup will provide access to a continuous spectrum of energies. Additionally, its setup should be lower cost than that of an X-ray fluorescence setup.

F.2 Compton Scattering for Calibration

The Compton scattering of photons off the electrons within a silicon sensor can be used to deposit 1 - 6 keV of energy into the sensor. The energy of a photon after Compton scattering off an electron is [257]:

$$E_{\gamma'} = \frac{E_\gamma}{1 + (E_\gamma / m_e c^2)(1 - \cos\theta)}, \tag{F.1}$$

where $E_{\gamma'}$ is the scattered photon's energy, E_γ is the initial energy, and θ is the scattering angle of the photon. The scattered electron acquires the energy $E_\gamma - E_{\gamma'}$, and this energy will be promptly reabsorbed as a localized energy deposit within the sensor. For example, if photons from an ^{241}Am source (peak at 59.5 keV) are scattered at angles in the range 5° to 90°, then between 0.03 and 6.2 keV will be deposited. A plot of the energy deposited in the sensor as a function of the scattering angle for photons from an ^{241}Am source is shown as the solid line in Figure F.1 (left). The energies of the radioisotope ^{55}Fe and of select X-ray fluorescence lines are also shown for comparison, as are the energies of Compton-scattered ^{109}Cd photons. Depending on the physical setup used, larger scattering angles could be used to access higher energies. Alternatively, a higher energy radioisotope could be used to produce the initial photons.

The differential cross-section for Compton scattering is given by the Klein-Nishina formula [258]. The total cross-section for the Compton scattering of a 59.5 keV photon is 5.47×10^{-25} cm^2. This varies depending on the energy as shown in Figure F.1 (right).

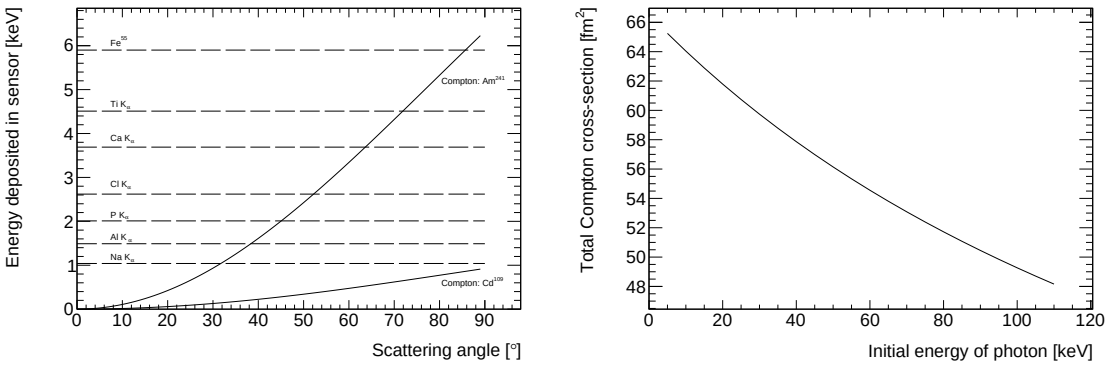


Figure F.1: (Left) Solid lines: energy imparted to a silicon sensor as a function of the Compton-scattering angle. This is just $E_\gamma - E_{\gamma'}$, with $E_{\gamma'}$ from Equation F.1. Lines shown for an ^{241}Am source (59.5 keV) and a ^{109}Cd source (22 keV). Dashed lines: select radioisotope and X-ray fluorescence lines. These are just for comparison and are not scattering-angle dependent. (Right) Total cross-section for Compton scattering as a function of the incoming photon energy.

To know the deposited energy, an X-ray detector can be placed to detect photons scattering into a small angular region. Any photons detected by the X-ray detector at a given scattering angle will have deposited approximately the same energy in the sensor. The total charge of each deposit associated with a particular scattering angle will not be known exactly due to the Doppler broadening of the Compton peak [259, 260, 261]. This broadening of the scattered energy peak is due to the fact that the electrons within the sensor are not at rest. For incident photons with energy of 59.5 keV, the FWHM of the peak of photons scattered at 90° is expected to increase by 0.6 keV [262]. Because of this effect, the Compton scattering method is best suited for an absolute calibration aggregated over many pixels, though the low hit rate, which will be discussed below, limits the usefulness of this method for a direct absolute calibration of the threshold DAC and noise.

The X-ray detector can be used to trigger data readout from a chip attached to the silicon sensor. In this way, a photon detected by the X-ray detector and an energy deposit in the silicon can be correlated. The position of the X-ray detector can be adjusted so that the calibration can be performed over a continuous range of energies. Three steps must occur for data collection to take place after the emission of a source X-ray:

1. The photon must be scattered.
2. The photon must be detected by the X-ray detector. Upon photon detection, the X-ray detector triggers chip readout.
3. The chip must have a hit. If there is a hit, its information is recorded.

If any of these steps fails, then no action is taken.

F.3 Device under study

A hybrid pixel detector was used in the development of this calibration technique, and the geometry of such a detector is shown in Fig F.2 [243]. An example of a Compton scattering leading to a charge deposit is also illustrated in the figure.

The sensor used to study this calibration method was a 150 μm silicon sensor manufactured by MPG-HLL [263], which had 100x25 μm^2 pixels. The sensor was bump-bonded onto an RD53A readout integrated circuit [264], which served as a readout chip. The total size of the chip is 20.0 mm by 11.6 mm, but only the differential front-end of the RD53A used, which leads to an active area of 6.8mm x 9.2mm. Readout and tuning were performed by using the YARR software framework [265]. The RD53A and its sensor are mounted on a custom-designed single chip Printed Circuit Board (PCB). The sensor, chip, and PCB unit will be referred to as the "silicon module assembly".

The front end of the RD53A is a charge sensitive amplifier (CSA) with continuous reset. The height of the analogue pulse the readout chip receives is proportional to the total charge deposited in the sensor. The pulse must rise above a set threshold to induce readout. This is the same threshold introduced in Section F.1. The pulse is digitized as a time over

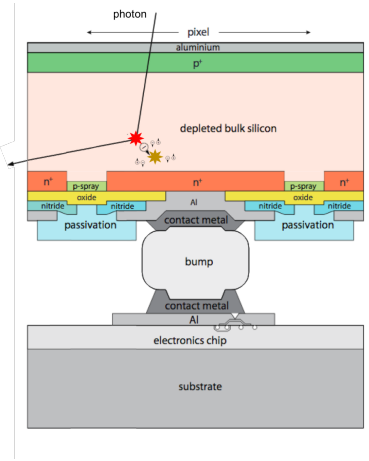


Figure F.2: Example geometry of a hybrid pixel detector. A passive silicon sensor is bump-bonded to a readout chip, which serves as the active electronics [243]. An incident photon is shown Compton scattering off of an electron at the point indicated by the red star. The electron deposits its energy within the sensor at the point indicated by the gold star, creating electron-hole pairs that drift within the depleted bulk.

threshold (ToT) reading, which is the integer number of clock cycles that the pulse stays above threshold. The RD53A has a 40 MHz chip clock. If the pulse were to barely exceed the threshold and dissipate before the next clock cycle, the ToT would be 1. This counts as detection. The ToT response for a given charge is manipulated by adjusting both the threshold and the signal fall time (or return to baseline). The settings for these parameters are based on input from the internal charge injection circuit. For the RD53A, ToT is stored as a 4-bit output in the datastream. Examples of the ToT response to injected charge for a particular tuning will be shown in Section F.6.

Every pixel in the RD53A has a charge-injection circuit, which is used for a pixel-by-pixel tuning of the threshold and discharge current. A simplified schematic of this circuit is shown in Figure F.3. Two input DC voltages can be selected for one terminal of a capacitor. Switching between these two voltages results in a charge proportional to the voltage difference being injected into the CSA. Most silicon pixel detectors will similarly have an internal injection circuit, and Compton scattering should provide the data needed to calibrate this circuit to an absolute charge scale.

For the RD53A in particular, the injection capacitor has a design capacitance of 8.5 fF. This cannot be measured for a chip with a sensor attached, such as the one used to study the Compton scattering method, but measurements on sensorless chips suggest an uncertainty for this value of about 0.3 fF. The differential nature of the injection is designed to eliminate uncertainties from the voltage distribution, and while there are temperature-related effects when temperatures change by 10C or more, the following study was performed

in an air conditioned room with stable temperature up to 5C. The DAQ system is essentially passive and should not induce any systematic effects in this study; however, there could be a systematic difference when comparing charge injected via the sensor rather than charge injected via the internal circuit. This is partially what this calibration method is designed to determine at the lower end of detectable charge.

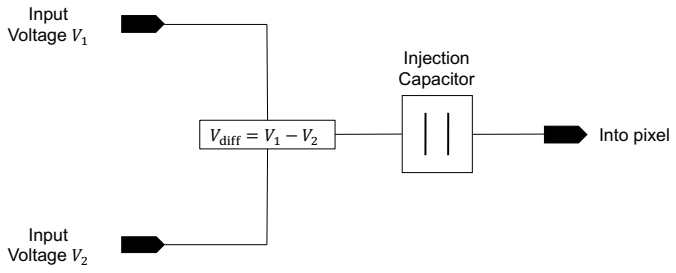


Figure F.3: Simple schematic of the calibration injection circuit in each pixel of the RD53A readout integrated circuit. Two input voltages are used to charge a capacitor, which then releases its charge into the pixel. This circuit is in parallel to the actual sensor.

By adjusting both the signal threshold and the signal fall rate, the 4-bit ToT output can be adjusted to cover a variety of energy ranges. The RD53A is capable of tuning to thresholds below 500 ẽ. For the differential front-end, the functional relationship between input charge and ToT is non-linear and pixel dependent.

In typical running, RD53A can self-trigger readout based on sensor hits. However, it can also be externally triggered. The data acquisition system (YARR) includes settings and scans based on such external triggering.

F.4 In-lab setup

To detect scattered photons, an Amptek X-123 x-ray spectrometer with a 5mm diameter CdTe detector element was used [266]. The spectrometer has auxiliary ports that allow for the output of a simple logic pulse whenever a photon is detected. This was the external trigger for the RD53A, as in the second step of data readout detailed at the end of Section F.1.

A picture and diagram of the experimental setup from above are presented in Figure F.4. The equipment is laid out on an optical bench with a 1" x 1" grid of holes. An ^{241}Am source with activity of 100 mCi is placed on one end of the grid; there is a 3mm aperture in a piece of brass shielding in front of the source that creates a beam of photons. The ^{241}Am spectrum has a 59.5 keV peak, with a measured FWHM of 0.55 keV. The photons pass through an additional brass collimator with a diameter of 3mm before falling on the silicon module assembly. The spectrometer is mounted on a single arm pivot with rotation point below the sensor, which allows the angle of the scattered-photon acceptance window to be easily

adjusted. During data taking, the setup is covered by a box to block out ambient light which could be a background to the silicon sensor and spectrometer. The scattering of photons from the collimator, non-sensor parts of the module assembly, and support structures are a background for the spectrometer.

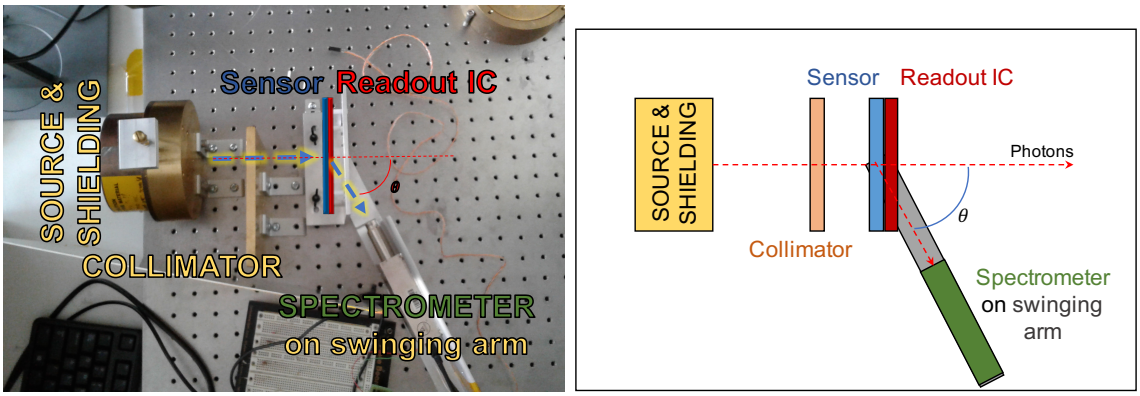


Figure F.4: (Left) Picture of the experimental setup. 59.5 keV photons are emitted from an Americium source through a hole in surrounding brass shielding, pass through a brass collimator, scatter off of the silicon sensor bump bonded to the readout integrated circuit, and are detected by the spectrometer. (Right) Diagram of setup.

F.4.1 Spectrometer readout

An example of the spectrometer's readout is shown in Figure F.5. The spectrum was acquired at an angle of 71.6° relative to the beamline. Here, two peaks can be seen, the larger one in light blue is the peak associated with Compton scattering. Because of the scattering angle, this peak is centered on an energy value of 55.1 keV. The smaller peak on the right is caused by the Thompson scattering of photons off of nuclei. These photons lose little energy, and as such, the peak is at 59.5 keV. The FWHM of the Compton peak, at 1.4 keV, is 2.5 times larger than that of the Thompson peak. The broadening of the Compton scattering peak is caused primarily by two factors. One is the Doppler broadening, which was discussed in Section F.2. Second, slightly different angles of scattering are permitted due to the aperture sizes of the shielding, collimator, and spectrometer. The geometries that lead to maximal and minimal scattering angles is portrayed in Figure F.6. Based around the center-line of the photon beam, the photons can be emitted within a distance of ±1.5 mm and they can be scattered within a distance of ±4 mm (this is somewhat restrained by the width of the front-end of the RD53A). Over the 150 mm distance between the shielding aperture and the sensor, this represents an angular range of ±2.1° for the photons incident on the sensor. Because the photon's incidence position can vary by ±4 mm with respect to the center-line, and the spectrometer aperture is ±2.5 mm, an additional angular smearing of ±2.4°

is expected when the spectometer is positioned at a 60° scattering angle. Therefore, ±4.5° of different scattering angles are allowed at the spectrometer positions to be explored in the following calibration. This effect alone would be expected to increase the FWHM by 0.8 keV.

There is some tradeoff between the broadening of the Compton peak and the rate of scattered photon detection. For example, by placing the spectrometer closer to the RD53A, its aperture would cover a larger solid angle. In this case the scattered photon acceptance would increase, but a larger range of Compton scattering angles would be detected. The chip-to-spectrometer distance in Figure F.4 was constrained by the swinging arm. The arm had a series of screw holes, and the spectrometer had to be fixed to one. The closest hole was chosen to maximize the solid angle for the given setup.

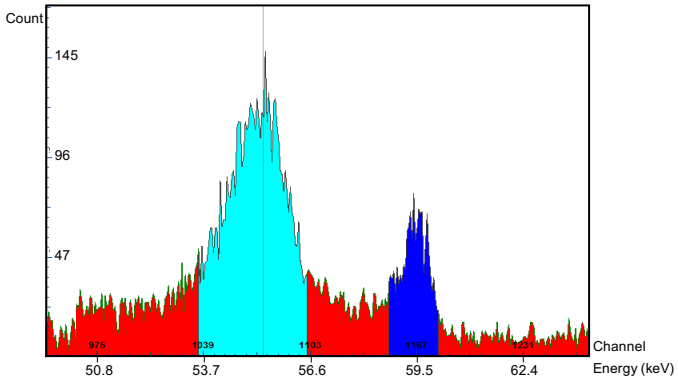


Figure F.5: Example photon spectrum for Compton scattering of photons at an angle of 71.6°. The peak associated with Compton scattered photons is in light blue; the peak associated with Thompson scattered photons is in dark blue.

F.5 Resolution

There is an intrinsic dispersion in the ToT response of each pixel to a given charge, both one injected internally and one deposited externally. In addition to manufacturing process variation making each pixel's transistors slightly different [267], charge deposits can happen at different times relative to the start of a clock cycle, influencing the number of cycles that will be counted, and there are non-uniformities in analog signal fall time. If the same charge was injected into a single pixel multiple times, a distribution of ToT values with non-zero width would be expected. This is demonstrated in Figure F.7. In this figure, four different pixels received 100 injections of 1000 ẽ from the internal circuit, and each pixel has a different distribution of ToT values. The figure also shows the distributions that occur for different injection fine delay settings. The fine delay fixes when the internal injection occurs relative

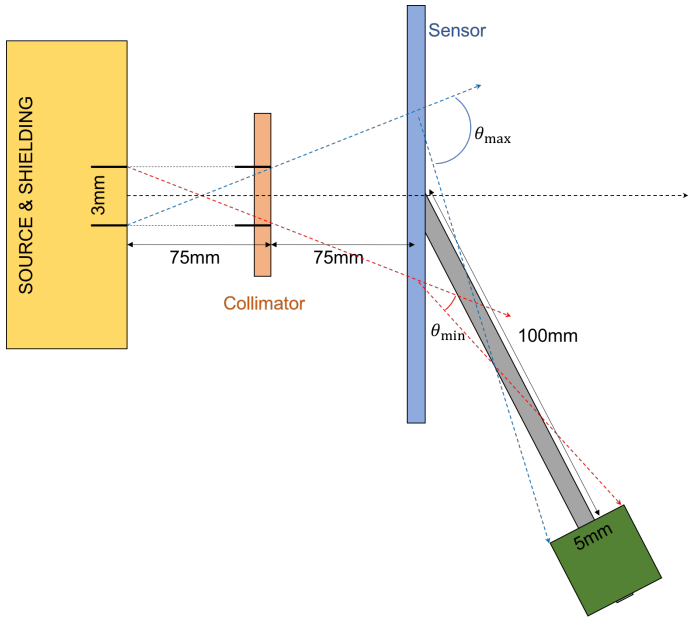


Figure F.6: Diagram of the scattering scenarios that lead to the geometric broadening of the Compton peak. Distances are not shown to scale here in order to make small distances and angles visible. The minimum scattering angle that the spectrometer will see, θ_{min}, is shown by the red dashed line, and the maximum angle, θ_{max}, is shown by the blue dashed line.

to the start of a clock cycle, and here demonstrates the impact of charge deposit timing on resolution.

Due to the intrinsic variation mentioned above, each pixel will have a unique charge → ToT response function. The knowledge of each pixel's function will be crucial for calibration.

F.6 Tuning and Observations

Tuning a chip is the process by which each pixel's threshold and charge to ToT response function is set. Many internal injections are performed at the desired threshold value, and potential pixel settings are scanned, allowing for the selection of the correct settings. Each pixel has 5 bits of threshold fine adjustment designed to allow the user to equalize the threshold of all pixels to the same value.

For the purposes of a calibration using Compton scattering, the RD53A was tuned to have a threshold of 450 ẽ and a slow signal fall rate of about 200 ns per 500 ẽ. A histogram of the threshold of each pixel after tuning is shown in Figure F.8 (left). As discussed in Section F.1, this would mean that we would expect a hit to be recorded 50% of the time if 450 ẽ were to be injected into a random pixel. The average pixel threshold achieved was 441 ẽ, with a pixel-to-pixel standard deviation of 24 ẽ. Similarly, a histogram of all pixels' average

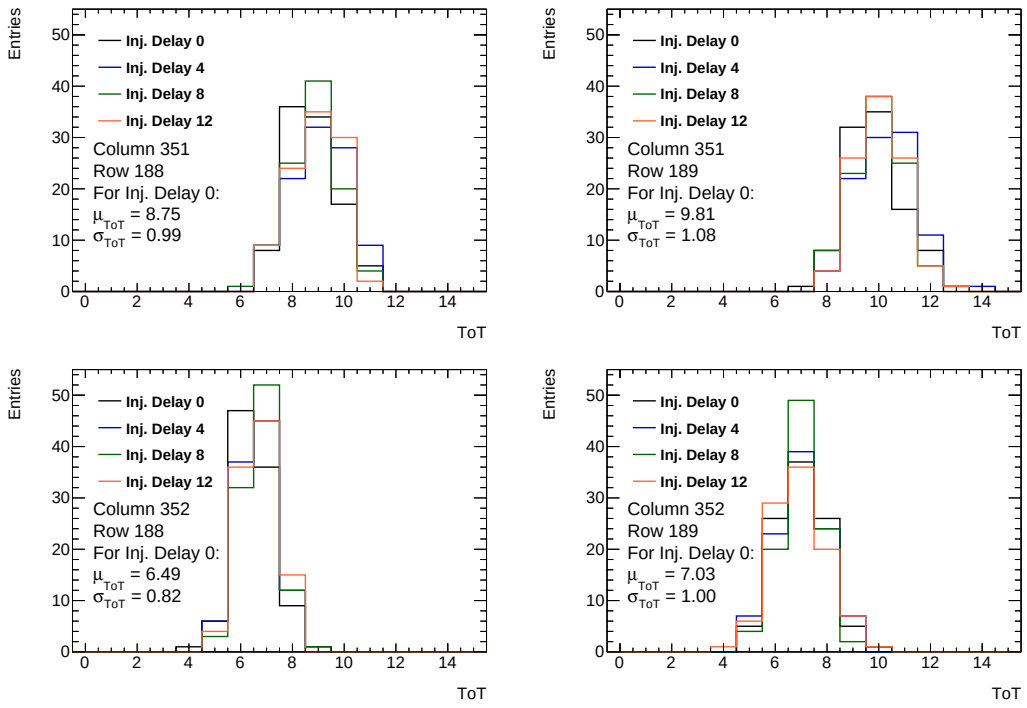


Figure F.7: ToT distributions for 100 internal injections of 1000 ẽ each into four different pixels. The injections are additionally shown for four different injection fine delay settings for each pixel. The fine delay affects when the internal injection occurs relative to the start of a clock cycle and is in units of 1/16 of a clock cycle. Here it is used demonstrate the effect of a Compton scattering happening at different times relative to the start of a clock cycle. The four pixels presented, in (column, row) format, are (351, 188), (351, 189), (352, 188), and (352, 189) from left to right.

ToT output due to a series of injections of 1000 ẽ is shown in Figure F.8 (right). Because of the slow return to baseline, the mean ToT is 7.8 clock cycles. The pixel-to-pixel dispersion manifests here as a standard deviation of 1.8 clock cycles. This can be compared to the single-pixel dispersion presented in Figure F.7. The asymmetry of the ToT distribution is due to the non-linearity of the ToT response to injected charge. The slope decreases with increasing charge, as can be seen in Figure F.9, so symmetric fluctuations in charge result in asymmetric fluctuations in ToT.

A unique ToT vs. injected charge response function exists for each pixel. For the purposes of this calibration, we want to find the function relating the most probable value of the ToT, here denoted as $(\text{ToT})_{\text{MPV}}$, to the injected charge. This will be a step function with integer values of ToT being possible. An alternate to this method would be to use the average ToT as a function of injected charge, but the average value is likely skewed upwards due

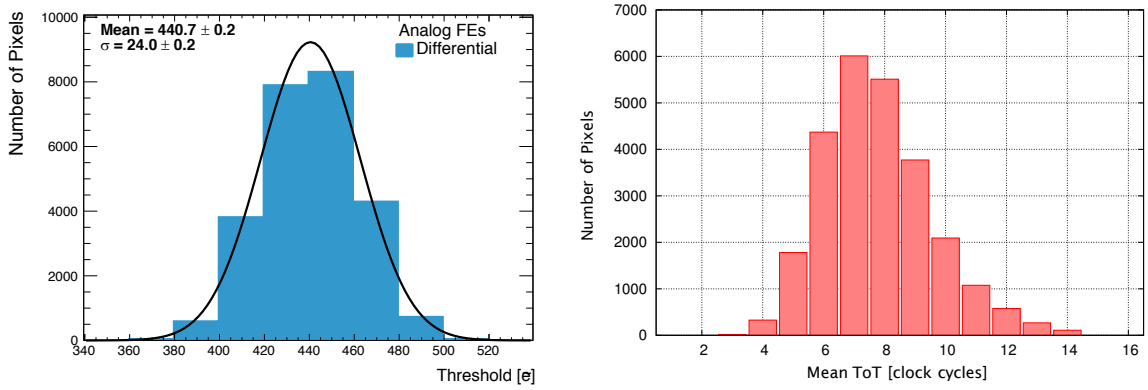


Figure F.8: (Left) Histogram of the per-pixel threshold achieved in a chip tuning. The average threshold achieved is 441 ẽ. (Right) Histogram of all pixels' average ToT for repeated injections of 1000 ẽ. The average ToT across all pixels for such an injection is 7.76 clock cycles.

to asymmetries in each pixel's ToT response. To approximate the per-pixel functions, a series of charge injections are performed from 500 ẽ to 1245 ẽ in steps of 5 ẽ for multiple fine delay settings. Three adjacent pixels' ToT response functions are shown in Figure F.9, illustrating the diversity of such functions found in the front-end. In the Compton scattering runs, each hit is associated with an individual pixel. The observed ToT will be compared to the expected most probable ToT, based on the mean charge deposit associated with the photon's scattering angle yields.

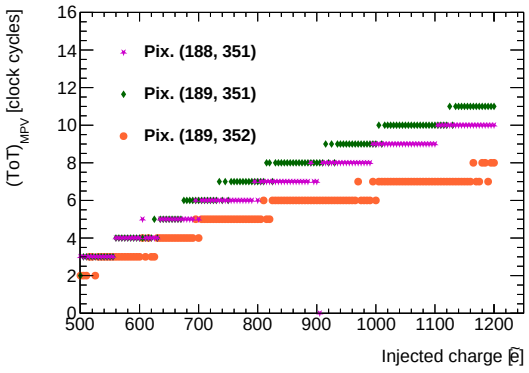


Figure F.9: Four adjacent pixels' ToT response functions. Points show the measured most probable ToT in the pixel for charge injections each ranging from 500 ẽ to 1245 ẽ in steps of 5 ẽ. The most probable value is found using multiple fine delay settings.

F.7 Noise and Backgrounds

As pixel threshold is lowered, noise fluctuations lead to recorded hits when no signal was present. Most commonly, this will result in a hit with ToT of 1, with an approximately exponentially falling distribution. The distribution of noise hits can be understood by recording data at random times with no photon source. Such a distribution is shown in Figure F.10 (left). It can be seen that the majority of noise hits have ToT = 1, with about 11% having ToT = 2, and 1% having a higher ToT. There is a noise hit in about 0.1% of randomly triggered readouts. Throughout these studies, fewer than 10 pixels in the differential front end had to be masked due to noise.

A significant fraction of the photons detected by the spectrometer are not Compton scattered by an electron in the sensor, and so do not deposit any charge in the sensor. Most of these photons have been scattered off of some other nearby object. Most of the time this occurs, there will be no coincident hit in the RD53A, meaning that the trigger returns an empty event in the third step of the data collection flow from Section F.1. However, some fraction of the time there will be a coincident hit, which may be noise or a hit associated with a Compton scatter at a random angle. To find the ToT distribution associated with this noise and background convolution, data is taken at random times with a photon source shining on the sensor. The resulting distribution is shown in Figure F.10 (right). Most of the hits with ToT = 1 here are noise, but hits with higher ToT tend to be from coincident X-ray hits. About half of the background hits have ToT = 15, which corresponds to overflow. These are hits where the pulse stays higher than threshold for a larger number of clock-cycles that can be encapsulated by the 4-bit ToT. Background hits comprise about 70% of this distribution, with noise hits as the other 30%. These percentages are based on comparing the number of hits with ToT $\leq$ 1, which are mostly noise, to the number of hits with ToT $\geq$ 2, which are mostly background.

In the data-taking run described in Section F.8, about 1/3 of hits that occurred were associated with hits with ToT 1 or 15. Based on Figure F.10 (right), about 1/5 of hits with ToT between 2 and 14 are expected to be background. In order to reduce background and noise contamination, the calibration will be performed using hits with ToT in the range [4,13].

F.8 Observations

The spatial distribution (in readout chip row and column coordinates) of recorded hits after 100 hours of running at a 56° scattering angle is shown in Figure F.11. These hits were read out based on a trigger caused by a photon being detected by the spectrometer, as detailed at the end of Section F.1. The hits shown have ToT in the range [3,14] in order to show fewer noise and background hits. The roughly circular distribution of hits is caused by the collimator, which was offset from the center of the module in order to illuminate the edge of the sensor for use in other studies. The exact alignment is not important for this work.

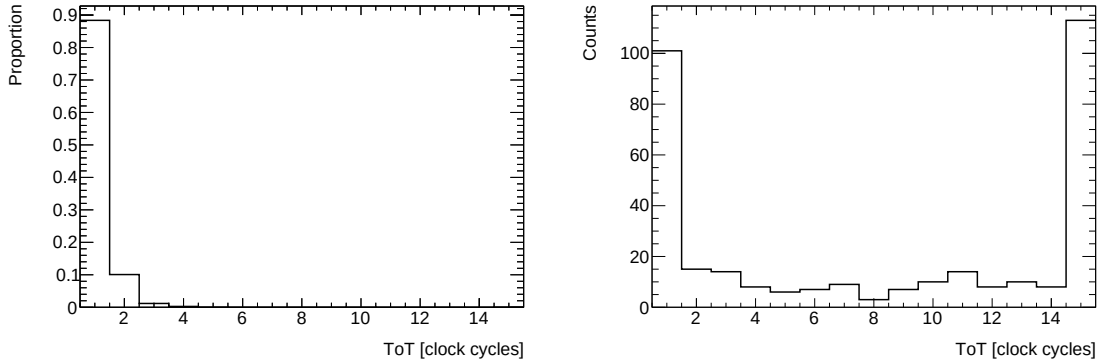


Figure F.10: (Left) The hit distribution that results from recording data at random intervals. There was no source of photons during this run, so these hits are considered to be pure noise. The majority of noise hits have ToT = 1, with about 12% having ToT = 2, and 1% having a higher ToT. (Right) The hit distribution for random data taking with an ^{241}Am source. This distribution is a combination of noise and background. Most hits with ToT = 1 are noise hits. Hits will higher ToT tend to be background, which will be mostly Compton scatters into random directions here.

The majority of pixels received only 1 hit during the run, if they saw a hit at all. 4% of hit pixels were hit twice throughout the run. Similarly, 95% of readout events had only one pixel activated. In 0.4% of events, two adjacent pixels were activated, creating a "cluster". The remaining events had multiple non-adjacent pixels activated. Such events are caused when a noise or background event is coincident with a Compton scatter. The distribution of noise and background would mimic that discussed in Section F.7.

F.9 Example Run and Proof of Principle

To test the method, data-taking runs were performed at four scattering angles: 51°, 56°, 63°, and 67°. Table F.1 shows the expected energy of charge deposits and the numbers of hits with ToT in the range [4,13] for each of the runs. Figure F.12 shows the distribution of per-pixel differences between the ToT of hits associated with Compton scattered photons and the most probable ToT expected from the calculated charge deposit (expected ToT). Only hits with ToT in the range [4,13] are considered in order to reduce noise and backgrounds. The expected ToT in each pixel is found using the response functions such as those in Figure F.9. The distributions of differences are fit to Gaussian functions, as shown in Figure F.12, and the mean and standard deviation results from the fits are presented in Table F.1.

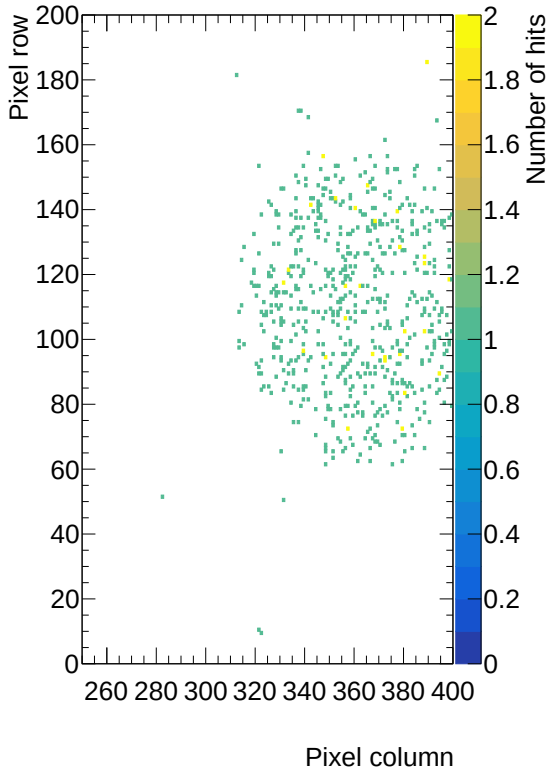


Figure F.11: Spatial distribution of hits with ToT in the range [3,14] in the readout chip after a 100 hour run at 56°. The hits were associated with a Compton scattered photon detected by the spectrometer. The majority of pixels had only 1 hit in the run, if they had any hit at all.

Angle [°]	Deposited Energy [keV]	Charge [e-h pairs]	Number of hits	μ_{fit} [clock cycles]	σ_{fit} [clock cycles]
51	2.5	690	347	1.5 ± 0.11	1.9 ± 0.08
56	3.0	830	697	1.0 ± 0.10	2.6 ± 0.07
63	3.7	1020	912	-0.08 ± 0.11	2.9 ± 0.08
67	4.1	1140	630	-1.0 ± 0.12	2.6 ± 0.08

Table F.1: Summary and results of the four runs used to test the Compton-scattering calibration method.

F.10 Precision and results

The Compton scattering calibration method is performed by comparing an observed ToT to an expected ToT. The expected ToT is derived by using the expected mean external

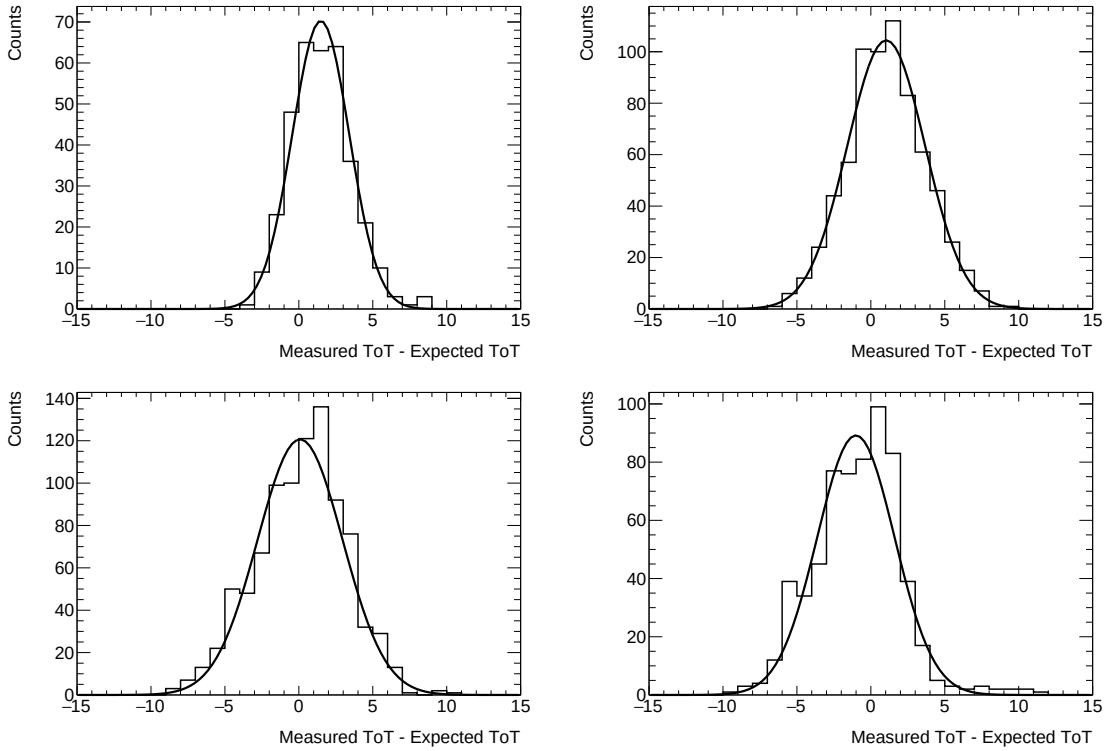


Figure F.12: Distributions of per-pixel differences between the ToT of hits from Compton scatters and the expected ToT. The expected ToT is based on an internal injection of the charge deposit expected from from the Compton formula. Top left, top right, bottom left, and bottom right show the data for the runs at 51°, 56°, 63°, and 67°, respectively. Fits to a Gaussian function are also shown for each distribution.

charge deposit as the argument in the charge to ToT response function. The charge to ToT response was derived using the uncalibrated internal injection circuit. Uncertainties on the function will result in uncertainty on the expected ToT. A precise calibration using this method requires accurate knowledge of both the external charge deposit and the expected charge to ToT conversion functions.

The external charge deposit is dependent on the scattering angle, and can be derived using Equation F.1. Every deposit caused by scattering at the same angle is not expected to be the same, as the scattered photon energy distribution has a non-zero width. However, the error on the centroid of this distribution comes from uncertainty on the angle. Here, an uncertainty of $\pm 1^\circ$ is taken on the mean scattering angle. The mean scattering angle is the angle of deflection of a photon that is emitted along the photon center-line and scatters into the center of the spectrometer in Figure F.6. The uncertainty comes from the measurement of the angle and translates to an uncertainty of 0.1 keV in deposited energy or 28 e-h pairs

in deposited charge.

The finite width of the Compton scattering peak, as observed in Figure F.5, is unavoidable and means that the exact charge deposit in each pixel hit cannot be known precisely. The width of the Compton scattering peak leads to the width of the distributions in Figure F.12. As noted in Section F.4.1, the FWHM of the Compton peak is 1.4 keV, which translates to a FWHM of 390 e-h pairs. Based on knowledge of the charge to ToT conversions functions, such a FWHM should translate into a widths of about 3 ToT units in the the distributions in Figure F.12. Further, the ToT measurement itself has imperfect resolution. As seen in Figure F.7, the injection of a single charge value can result in a ToT distribution with FWHM of up to 3 ToT units, which is primarily caused by noise due to the small charge scale. A combination of the energy width and ToT resolution lead to an expected standard deviation in the (observed - expected) distribution of about 2 ToT units, depending on the pixels involved. The standard deviations seen in table F.1 are slightly larger, but this is likely due to background hits, which will have an overall broadening effect. An uncertainty of 0.5 clock cycles will be taken on the centroids in order to account for potential biases the background introduces.

Combining the uncertainty on the Gaussian centroid and the uncertainty due to backgrounds yields a conservative uncertainty of $\pm$ 0.6 ToT unit on the expected ToT.

The mean values of the distributions in Figure F.12 are plotted against the expected charge deposit in Figure F.13 (a), along with the uncertainties on both the expected ToT and the average external charge deposit. A fit to line was performed, finding a slope of -0.54 ± 0.18 units of ToT per 100 e-h pairs difference in charge deposit for this particular tuning. An intercept of 5.4 ± 1.7 ToT units was found, which translates to a difference of 3.0 ± 0.9 units of ToT at the threshold injection of 440 e-h pairs. In Figure F.13 (b), the correction is applied to the response function of the average ToT across the entire differential front end to injected charge. The correction is not meant to be applied on a pixel-to-pixel basis, as it is derived by averaging over multiple pixels. It is also not predicted to be linear *a priori*, but a linear regression was performed because it fit the data well. Applying the correction outside the bounds of 690-1140 e-h pairs is necessarily an extrapolation, which would need to be validated with additional runs at other energies. This is particularly important when considering the threshold in this case. While Figure F.13 implies a threshold below 440 e-h pairs, the true correction may have non-linearity that has not been accounted for.

Interpreting these results is non-trivial, as a variety of factors or combinations of factors could influence the result. For example, the presence of a sensor on this chip makes it impossible to measure the capacitance of the injection capacitor. However, changing this value would affect overall offset of the points, and cannot significantly change the observed slope. There is also uncertainty relating to the charge collection for a chip with a sensor present; there may be systematic inefficiencies in charge collection or other unanticipated effects.

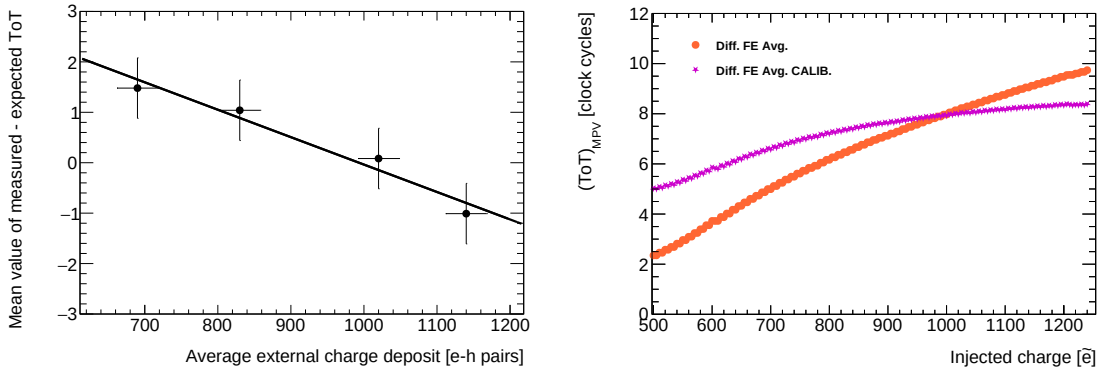


Figure F.13: (Left) Plot of the mean values of the measured - expected ToT distributions found in Figure F.12 as a function of the expected charge deposit. A linear trend line has been included with the four data points. (Right) The average ToT across the entire differential front end is plotted against injected charge both before and after using the correction found in (left).

F.11 Discussion and Conclusions

This appendix presents a new method for the calibration of thin silicon sensors with energies in the range of 1 - 6 keV using the Compton scattering of photons. The method was developed and tested using a 150 μm sensor bump-bonded onto an RD53A and an ^{241}Am source. The physical setup involved a spectrometer that can be easily repositioned to enable calibrations at any desired energy within the achievable range. An example calibration was demonstrated, finding a trend in the difference between the ToT response to the external charge deposit and corresponding internal charge injection. This demonstration also doubles as a first low-energy calibration of the differential front end of the RD53A chip, as previous low-energy calibrations have all been performed using the linear front-end with deposited energies above 2 keV. In some sense, the observed result suggests a malfunction in the charge injection circuit, as the preamplifier response does not appear problematic. However, there could also be effects related to charge collection from the sensor, which would be particularly important for small charge deposits. Validation of these results using an X-ray fluorescence setup is an important next step for this method.

In the four data taking runs used for the example calibration, about 5 Compton-scattering hits were observed per hour. That means that about 7% of the pixels exposed by the collimator were hit during a 100 hour run. Because of this, the four runs presented above were actually comprised of hits in largely non-overlapping sets of pixels. Based on the current setup, about 700 hours of running would be needed to expect 50% of pixels to be hit. However, this required time could be reduced by adjusting the setup. The hit rate will increase linearly with source activity. The trigger rate will increase quadratically

with distance from module to spectrometer (linearly with solid angle subtended), but this would lead to further broadening of the Compton scattering peak or create multiple peaks. Instead, using multiple spectrometers (or a segmented spectrometer) will multiply the rate accordingly without loss of resolution, and will allow taking data for multiple energies at the same time. In any case, data taking periods of at least a week long are likely needed in order to achieve adequate statistics. This method is thus not suitable as a quality control tool to be applied to a large number of modules, but rather as a means to characterize the response and absolute charge scale on a single representative device.

Compared to the Compton scattering method, traditional calibration methods provide superior speed. They also typically involve narrower energy peaks and smaller backgrounds. However, these methods are best applied above 6 keV. The Compton scattering calibration method provides access to a continuous spectrum of energies between 1 and 6 keV. The setup and materials are readily available in most labs and cost less than an X-ray fluorescence setup. An extension of this method could be to trigger on the Thompson peak instead of the Compton peak of Figure F.5, in order to reach very low energy deposits (of order eV and below) for experiments measuring low energy phonon signals instead of charge.

Appendix G

An Introduction to Particle Physics

This appendix contains a series of lecture notes designed to give the audience the background knowledge of particle Physics needed to understand a basic particle physics analysis. These notes are aimed at the level of an undergraduate, and were inspired by some outreach activities I was involved in as an graduate student. This is a living document, and I hope the lectures will evolve as I give and improve them. If you're reading this and find something that's wrong, please let me know! The content is primarily adapted from [50, 268, 269, 270, 271, 96].

G.1 Mathematical Matters

Throughout this lecture, I will try not to take too much mathematics as assumed knowledge other than Calculus. As a physicist, it is easy to talk about things like fields (in the Physics sense) and groups (in the math sense), while forgetting that these concepts are very abstract and essentially meaningless to most people. As such, I will try ground the discussion in concepts that are familiar from everyday life.

Fundamentally, Physics is an attempt to construct models that describe physical processes. These models should not only agree with what has previously been observed; they should also be able to predict future observations. These goals are somewhat nebulous if defined independently, but they are concretely expressed in the language of mathematics. For example, while it is nice to say that the force that causes an apple to fall is the same as the force that binds the Moon to the Earth, it is even nicer to accurately predict the apple's or Moon's position and velocity as a function of time.

Most Physics models are designed to describe and predict how the state of a system evolves in time. By a "system", I mean some collection of objects that will interact with each other. This might be the Solar system, blocks connected by a spring on a table, or the electron and a proton of a Hydrogen atom. The "state" of the system is a description of how the system is arranged at any given moment. For example, if you knew the relative position of the Sun and all the planets, as well as the speed and direction of those bodies at a single

given time, that would be the "state" of the Solar system. By "evolves in time", I just mean how the state changes as time goes on. In other words, if you want to predict when the next solar eclipse will occur, you'll need to be able to accurately predict the positions of the Sun, Earth, and Moon for the next several years based on your current knowledge of state of the Sun-Earth-Moon system.

G.1.1 Units

Because we are trying to describe physical systems, we must have a common language to discuss quantities such as lengths, times, masses, and energies. In physics, the standard units for these quantities are meters, seconds, kilograms, and joules, respectively. These units are defined such that the quantities that we encounter from day-to-day are close to one (where I use "close" loosely to mean within a few orders of magnitude). For example, a car might travel at 14 m/s (50 km/hr) and a person might eat a burger with a mass of 0.1 kg. We should take a moment to note that the typical unit for energy, the joule, can be thought of as a unit of the form $mass \cdot \frac{distance^2}{time^2}$ or $\mathrm{kg} \cdot \frac{\mathrm{m}^2}{\mathrm{t}^2}$. This can been explained by remembering that the formula for kinetic energy is $\frac{1}{2}mv^2$ or that Einstein famously told us that $E = mc^2$. Of course, one could equally say that mass is a unit of the form $energy \cdot \frac{time^2}{distance^2}$.

For particle physicists, these units are not quite as useful. Particles have masses that are around 10^{-30} kilogram and often travel with speeds around 10^8 m/s. Because of this, we will use a system called "natural units" throughout this thesis (unless units are otherwise explicitly stated). In this system, we define two quantities as 1:

$$c = 2.998 \times 10^8 \ \frac{\text{meters}}{\text{second}} = 1$$

and

$$\hbar = 1.055 \times 10^{-34} \text{ joules} \cdot \text{seconds} = 1.$$

The first of these is the speed of light, and the second is a constant called "Plank's constant". It appears in many contexts in quantum mechanics; for example, the energy of a photon, E_γ, is a function of its frequency, ω: $E_\gamma = \hbar\omega$.

There is also a common unit of energy that appears in particle physics called the electron-volt or eV. This is the amount of kinetic energy an electron acquires when it is accelerated over one Volt of electric potential difference:

$$1\text{eV} = 1.602 \times 10^{-19} \text{ joules}.$$

Masses in particles physics are often quoted in units of MeV or GeV; for example, the proton mass is around 1 GeV. This is an appropriate unit for mass thanks to Einstein's $E = mc^2$ formula and the fact that $c = 1$! Momenta, which carry units of $mass \cdot speed$ are also often quoted in MeV or GeV.

G.1.2 Vectors, Rotations, and Groups

Physicists describe quantities that involve "direction" by using vectors. For example, if you throw a ball, its velocity is best expressed using a vector. Once the ball leaves your hand, it is travelling away from you, up from the ground, and potentially to the left or right if there's a breeze. In three-dimensional (3D) space, there are three directions the ball can travel in, or three "degrees of freedom". In this case, three values must be used to adequately express the vector. Generally, physicists define a coordinate system to keep track of their directions. Cartesian Coordinates are a common choice. One direction is chosen as "x", the direction 90° to the left is called "y", and the direction 90° up from both x and y is called "z". Thus the velocity of a ball, which can be visualized as an arrow pointing in some direction in the coordinate space, can be written as $\vec{v} = \{v_x, v_y, v_z\}$.

The defining feature of a vector is the way it behaves under rotations. When a rotation occurs, the degrees of freedom of a vector must be transform into each other in a standardized way. This is not hard to visualize. Imagine that you call the direction to your right x, and the direction in front of you y. If you point an arrow to the right, and its length is 1 meter, then you might write $\vec{x}_a = \{1\text{m}, 0, 0\}$. If you then point the arrow in front of you, then $\vec{x}_a = \{0, 1\text{m}, 0\}$. The x-component *became* the y-component. If you pointed the arrow diagonally to the front and right, then $\vec{x}_a$ might be $\{\frac{1}{\sqrt{2}}\text{m}, \frac{1}{\sqrt{2}}\text{m}, 0\}$. There the former x-component is shared between the new x- and y- component. The important thing is that the arrow is still the arrow; its length and shape did not change- only its orientation changed. An example of a non-vector would be a list counting the numbers of apples, bananas, and oranges on a shelf. You can write this as $\{1\text{a}, 2\text{b}, 3\text{o}\}$, but there is no way to transform an apple into a banana.

A vector is a type of tensor. A three-component vector in 3D space is called "rank 1" tensor. A "rank 0" tensor is called a "scalar". A scalar has only one component and is invariant under rotations. For example, temperature is a scalar- it has no direction. The rank of the tensor can be thought of as the number of rotations that must be done on the tensor to transform it correctly. A higher-rank generalization of a vector would be something like a stress tensor. This tensor keeps track of the x-, y-, and z-directional forces on the x-facing sides of an item, the forces on the y-facing sides, and the forces on the z-facing sides. The rank can also be thought of as the number of indices needed to keep track of the degrees of freedom of the tensor. In the stress tensor, the values can be indexed as xx, xy, xz, yx, yy, and so forth, where the first index refers to the face, and the second refers to the direction of the force. There must be one rotation per index. To keep the stress tensor the same if you rotate the item, then the forces must be rotated with the item.

Physicists and Mathematicians refer to the collection of all possible rotations in 3D space as the group "$SO(3)$". A "group" is a collection of elements with a defining operation that can combine two of the elements. The elements that comprise a set are totally arbitrary; they could be numbers, rotations, or something like all possible orderings of the set of numbers {1 2 3}. The operation that combines two elements must respect the principles of closure, associativity, and invertability, and there must be an identity element. This sounds very

abstract (and to some extent it is), but there are many groups that everyone is familiar with. For example, the set of all integers

$$\ldots, -2, -1, 0, 1, 2, \ldots$$

together with addition defines a group called "$\mathbb{Z}$". In this context, the identity element is 0: adding 0 to any number does not change the number. Closure means that any sum of integers is also an integer. Associativity means that $(x + y) + z = x + (y + z)$ for any three integers. Lastly, invertability means that for every element, there is another element that can be combined with that element to return the identity. Every integer's inverse is just its negative: $2 + (-2) = 0$ and $10 + (-10) = 0$.

For rotations, or $SO(3)$, the elements of the group are the rotations themselves. The element-combining operation is just performing rotations one after the other. Here, the identity element is performing no rotation, closure is satisfied because any two successive rotations is also a rotation, associativity is guaranteed by maintaining the order of rotations, and any rotation can be inverted by just rotating in the opposite direction by the same amount. The "SO" in the group's name stands for "special orthogonal", which refers to the fact that rotations preserve length and orientation. The metric for the preserved length is the usual length of a vector: $s = \sqrt{x^2 + y^2 + z^2}$. Orientation is preserved in the sense that rotating your right hand will never make it your left hand- that would require a spatial inversion, which is a change in orientation. The "(3)" in the name refers to the fact that 3D rotations can be represented by 3x3 matrices. The intimate connection between vectors and rotations comes from the fact that vectors are the objects these 3x3 matrices act upon.

There are many groups that are important in physics that we will encounter going forward. For example, one could imagine existing in 4-dimensional space, where the group of rotations would be $SO(4)$. One could also imagine living on a 2D plane, where the group of rotations would be $SO(2)$, which can be represented by 2x2 matrices that transform x- and y-components of 2D vectors into each other. Rotating in a plane is equivalent to spinning around a single axis. There are actually multiple groups that describe this action. The most important alternate to $SO(2)$ is called $U(1)$. $U(1)$ is the group of 1x1 "unitary" matrices, or more simply, all complex numbers of the form $1 * e^{i\phi}$. All complex numbers, z, can be represented in the form $z = x + iy$ or equivalently $z = Ae^{i\psi}$, where $A = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1} \frac{y}{x}$. Complex numbers can be drawn in a 2D Cartesian Coordinate system where the x-axis is the real component, and the y-axis is the imaginary component; this is what the $x + iy$ notation indicates. The $Ae^{i\psi}$ notation highlights that the complex number has a magnitude, or distance from the origin, "A", and some angle from the positive real axis, "ψ". Multiplying a complex number by an element from $U(1)$ simply changes the angle of the complex number, spinning it around the origin. Hence the action of $SO(2)$ and $U(1)$ are the same.

The connection between $SO(2)$ and $U(1)$ is called an "isomorphism" (from the Greek for "equal form"). There exists a mapping of each element in $U(1)$ to an element in $SO(2)$:

$$U(1) : e^{i\phi} = \cos\phi + i\sin\phi \rightarrow SO(2) : \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}.$$

The elements in $U(1)$ act on single complex numbers, and the elements of $SO(2)$ act on 2-component vectors.

The rotation group $SO(3)$ is locally isomorphic to a group called "$SU(2)$", which is the special unitary group of 2x2 matrices. All of the elements of both groups can be parameterized by three angles, θ, ϕ, ψ, and there exists a mapping between them:

$$SO(3): \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \rightarrow$$

$$SU(2): \pm \begin{pmatrix} e^{i\frac{\theta}{2}} & 0 \\ 0 & e^{-i\frac{\theta}{2}} \end{pmatrix} \begin{pmatrix} \cos\frac{\phi}{2} & i\sin\frac{\phi}{2} \\ i\sin\frac{\phi}{2} & \cos\frac{\phi}{2} \end{pmatrix} \begin{pmatrix} e^{i\frac{\psi}{2}} & 0 \\ 0 & e^{-i\frac{\psi}{2}} \end{pmatrix}. \tag{G.1}$$

We should note the $\pm$ sign in this mapping. For every *one* element in $SO(3)$, there are *two* in $SU(2)$. However, all of the the elements that are next to each other in $SO(3)$ are also next to each other $SU(2)$ in the sense that two rotations are nearby by having slight differences in rotation angle or direction. This is why I referred to the isomorphism as "local" earlier; in fact, $SU(2)$ is a double-cover of $SO(3)$.

As a group that is isomorphic to the rotation group, $SU(2)$ is itself an abstraction of rotations. The rotations created by elements of $SU(2)$ transform two complex numbers into each other, rather than the three real numbers transformed by elements of $SO(3)$. It is interesting to realize that the 2-component complex vector must be rotated by 4π to return to its initial state, rather than the 2π we are accustomed to for 3-component real vectors. This can be intuited from the fact that $SU(2)$ is a double-cover or seen directly from the mapping in Eq. G.1.

Physically, the group $SU(2)$ appears when describing the Weak force- which we will get to soon- and the spin of particles. Spin is an internal property that all particles possess (the name comes from how some particles behave in a magnetic field). The spin state of some classes of particles can be expressed using 2-component complex vectors that are rotated by elements of $SU(2)$ (maybe the attentive reader remembers the concept of "spin up" and "spin down" electrons from Chemistry; the spin vector expresses that information in its two components). If such a particle is rotated in 3D space, then its momentum and position, which are regular 3D vectors, rotate in accordance with an element from $SO(3)$, and its spin is rotated using the corresponding element from $SU(2)$. The fact that the spin of such particles takes two full spatial rotations to return to its initial state is an experimentally confirmed fact [272, 273]. Spin is just as much a property of particles as color might be a property of a ball; if, for whatever reason, the color of the ball changed from blue to red and back to blue as you rotated the ball twice, then the color could be represented by an $SU(2)$ vector.

Speaking of color, there are are some particles that actually do have an internal property that physicists have called "color". Sadly, it is not related in any way to the normal concept of color. The color property of particles is expressed using 3-component complex vectors of

Group	N	Shape
SO(2)	1	2x2
SO(3)	3	3x3
SU(2)	3	2x2
SO(4)	6	4x4
SU(3)	8	3x3

Table G.1: The number of matrices, N, in the set of generating matrices for the important Lie Groups discussed in this section. The shapes of the matrices are also made explicit.

the group called $SU(3)$, which is the special unitary group of 3x3 matrices. Whereas $U(1)$ is related to $SO(2)$ and $SU(2)$ is related to $SO(3)$, $SU(3)$ is not related to $SO(4)$.

On a technical note, all of the groups discussed above are examples of "Lie groups" (named after Sophus Lie). For each of the above groups, there exists a small set of N matrices, T_n, such that every element in the group can be generated by exponentiating a linear combination of the T_n. The number of matrices, N, varies from group to group. A linear combination of matrices is just a matrix sum where each matrix can have a different weight, θ_n:

$$A = \sum_{n=0}^{n=N} \theta_n T_n.$$

Thus, any element of the group can be expressed as e^A. This exponential is best conceptualized by remembering the Taylor expansion: $e^X = \sum_{k=0}^{\infty} \frac{X^k}{k!}$.

The numbers of matrices in the generating set for each of the above groups is written in Table G.1. We can note here that $SO(3)$ and $SU(2)$ have the same N as expected from their isomorphism.

For example, for $SO(3)$ the 3 matrices are:

$$SO(3) : S_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \; S_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \; S_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{G.2}$$

and for $SU(2)$ the matrices are:

$$SU(2) : \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \; \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{G.3}$$

These matrices are known as the "Pauli matrices". Technically, the generating matrices for $SU(2)$ are $i\sigma$, but the Pauli matrices are conventionally defined without the factor of i.

The choice of the parameters θ_n can be thought of as rotation angles. To generate the element of $SO(3)$ corresponding to a rotation around the z-axis by θ_z degrees, we find (after

working through the math):

$$e^{0*S_x+0*S_y+\theta_z*S_z} = \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{G.4}$$

as expected.

For the matrices in Eq. G.2 and G.3, we should note that the matrices *do not commute*. That is to say for any two different matrices, $[S_1, S_2] = S_1S_2 - S_2S_1 \neq 0$. The bracket notation there merely indicates the difference found after the equal sign. If the generating matrices of a group do not commute, that groups is said to be "Non-Abelian". All of the groups discussed above are Non-Abelian except for $SO(2)$, or equivalently $U(1)$. This may seem like an unnecessary tangent now, but it has important consequences in particle physics, as we will see later.

The generating set of matrices for each group satisfies a particular equation relating to how the matrices commute. For example, the matrices for $SO(3)$ satisfy

$$[S_i, S_j] = \epsilon_{ijk}S_k, \tag{G.5}$$

where $\epsilon_{ijk} = 1$ for $\{ijk\} = \{123\}, \{231\}, \{312\}$, $\epsilon_{ijk} = -1$ for $\{ijk\} = \{321\}, \{213\}, \{132\}$, and $\epsilon_{ijk} = 0$ for any other choice of $\{ijk\}$. The S_i for $SO(3)$ are 3x3 matrices, but the 2x2 matrices generating $SU(2)$ from Eq. G.3 *also* satisfy Eq. G.5 (technically they satisfy $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$, but that's a matter of normalization). The commutation relations of a Lie Group's generating matrices is a defining feature of the group.

G.1.3 Relativity

Most theories in modern field theory respect the principles of Special Relativity, though it is interesting to note that Einstein intuited relativity from the classical field theory of Electrodynamics. The foundational principles behind Special Relativity are that the speed of light is both a universal constant and a universal speed limit, and that the laws of physics are the same in all inertial references frames. An inertial reference frame is one that is moving at constant speed and direction for all time. In other words, an experiment conducted on a spaceship moving at half the speed of light will yield the same results as an experiment conducted in a spaceship that is standing still. The consequences of relativity can be difficult to understand, and many tomes have been written about flashlights on trains, twins aging at different speeds, and planes flying through barns. We can eschew such metaphors though; the most fruitful way to approach relativity in the context of particle physics is through the lens of group theory.

In classical physics, space and time are totally separate: time plods along at a universal, regular pace, and space can be rotated but not distorted by elements of $SO(3)$. In relativity, space and time are not so distinct. The directions of space can be rotated into each other in the usual manner, but space can be "rotated" into time and time can be "rotated" into space.

These space-to-time "rotations" are called "boosts". They occur when reference frames are changed, or in other words, by speeding up or slowing down how fast you move when making observations. Because space and time are on equal footing and can be transformed into each other, it is now possible to write a 4-dimensional vector:

$$x^\mu = \{x_0, x_1, x_2, x_3\} = \{ct, x, y, z\}.$$

Here we have introduced the usual notation that vectors (and tensors) that exist in space-time are indexed with Greek letters, where the 0^{th} component of the vector is the time component. Because $c = 1$, we normally just write ct as t.

The usual group for 4-dimensional rotations is $SO(4)$. This group preserves a vector's orientation and its length, where the length-squared is defined as $s^2 = x_0^2 + x_1^2 + x_2^2 + x_3^2$. We can note here that $SO(3)$ is a subgroup of $SO(4)$, meaning that if one axis is held fixed, all the remaining rotations are just the rotations of 3D space. Sadly, $SO(4)$ is not the group that contains all relativistic rotations and boosts, but it is related. The correct group is known as the Lorentz Group. The Lorentz Group preserves orientation, but it preserves the length metric $s^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2$. Similar to $SO(4)$, the Lorentz Group has 6 generating matrices. Three correspond to spatial rotations:

$$J_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \; J_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \; J_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{G.6}$$

and three correspond to boosts:

$$K_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \; K_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \; K_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{G.7}$$

If we perform an exponential mapping to convert the generators into group elements, we might find, for example, that a spatial rotation around the z-axis by an angle θ_z gives:

$$R_z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta_z & \sin\theta_z & 0 \\ 0 & -\sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{G.8}$$

which is reminiscent of Eq. G.4. If we take the exponential $e^{\beta_x * K_1}$, we find:

$$e^{\beta_x * K_1} = \begin{pmatrix} \cosh\beta_x & \sinh\beta_x & 0 & 0 \\ \sinh\beta_x & \cosh\beta_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{G.9}$$

This is a boost in the x-direction. The parameter β_x is called the rapidity of the boost. The rapidity is related to the speed of the boost by:

$$\cosh\beta_x = \frac{1}{\sqrt{1-v^2}},\ \ \sinh\beta_x = \frac{v}{\sqrt{1-v^2}}. \tag{G.10}$$

The speed divided by the speed of light is denoted by v. Remember that $c = 1$, so all speeds are now fractions. While the rapidity can range from $-\infty$ to ∞, the $\cosh\beta_x$ ranges from 1 to ∞, and $\sinh\beta_x$ ranges from $-\infty$ to ∞.

The elements of the Lorentz group, all possible spatial rotations and boosts, are known as Lorentz Transformations. Because they can be expressed as 4x4 matrices in space-time, they are often denoted as $\Lambda_{\mu\nu}$, where the μ and ν indices denote the position in matrix. Similar to $SO(3)$, there are objects that are scalars, vectors, and tensors under the Lorentz group. Scalars do not transform under transformations, vectors transform as $x_\mu \to \Lambda_{\mu\nu}x_\nu$, and tensors transform as $F_{\mu\nu} \to \Lambda_{\mu\sigma}\Lambda_{\nu\rho}F_{\sigma\rho}$ (one transformation for each of the indices). I have introduced here Einstein Index Notation, a notation where repeated indices indicate summation of the form: $a_\mu b_\mu = a_0b_0 - a_1b_1 - a_2b_2 - a_3b_3$.

In 3D space, the product of two vectors is a scalar, such as the length-squared of a vector $s^2 = x^2 + y^2 + z^2$. The length of a vector does not depend on how much you rotate it. It is also possible to build Lorentz scalars from products of Lorentz vectors and tensors. For example, the quantities $x_\mu x_\mu$ and $F_{\mu\nu}F_{\mu\nu}$ are Lorentz scalars. Lorentz scalars can be spotted if there are no non-repeated indices.

Aside from scalars, vectors, and tensors, there is another important type of object that transforms under the Lorentz group. To understand this type of object, we must examine the structure of the Lorentz group. The commutation relations of the group are:

$$[J_i, J_j] = \epsilon_{ijk}J_k \tag{G.11}$$

$$[J_i, K_j] = \epsilon_{ijk}K_k \tag{G.12}$$

$$[K_i, K_j] = \epsilon_{ijk}J_k. \tag{G.13}$$

This tells us that the products of rotations in orthogonal directions give rotations as expected, that the product of a rotation and a boost is a boost, which is not unexpected, and that the product of two orthogonal boosts will result in a rotation. This last point is little strange but it is part of how relativity distorts space and time. The main takeaway here is that rotations are connected to boosts, and boosts are connected to rotations.

It is possible to do something interesting with the generators of the Lorentz group though. Where we normally divide them into rotation and boost generators, we can alternatively build mixed generators as:

$$J_i^+ = \frac{1}{2}(J_i + iK_i),\ \ J_i^- = \frac{1}{2}(J_i - iK_i).$$

The i subscripts run from 1-3 as in Eqs. G.6 and G.7. With these mixed generators we find:

$$[J_i^+, J_j^+] = \epsilon_{ijk}J_k^+ \tag{G.14}$$

$$[J_i^-, J_j^-] = \epsilon_{ijk} J_k^- \tag{G.15}$$

$$[J_i^+, J_j^-] = 0. \tag{G.16}$$

This tells us that there are secretly two separate $SU(2)$s hiding in the Lorentz group! The $SU(2)$ subgroups may be separate, but this is not saying that space and time are separate, after all, the generators of the subgroups combine boosts and rotations.

Because we know the structure of the subgroups, we can choose different generating matrices to *represent* them, and this will not change the structure of the group or its number of elements. All of the boosts and rotations will still be there. Luckily, we've already encountered matrices that represent $SU(2)$, which can be found in Eq. G.3. This raises the question of whether we can have two sets of those 2x2 matrices to serve as our generators. We would still have 6 generators (3×2), and Eqs. G.14 and G.15 are still satisfied. However, in this case Eq. G.16 would be violated.

But there's one last trick up our sleeves. There's another value that can satisfy all of Eqs. G.14-G.16: 0. This is the "trivial" representation of $SU(2)$. It might seem silly to map every rotation to 0, but there *are* values that don't transform under rotations: scalars.

Armed with the knowledge that we can represent one of the two $SU(2)$s in the Lorentz group as 0, we can represent the other with the 2x2 matrices generated by the matrices in Eq. G.3. If our Lorentz transformations are represented by 2x2 matrices, then the objects that they transform have 2 components. These 2-component objects are known as "spinors".

This little diversion was really just a long-winded way to finish up our list of objects that transform under the Lorentz group: scalars (1 component), spinors (2 components), vectors (4 components), and tensors (number of components depends on the tensor rank).

As a grounded example of how relativity works, we can consider the case of muons created by cosmic rays. Muons are particles that are very similar to the electron, but heavier and unstable. If they are sitting at rest, they decay into an electron in about 2.2 μs. They are created in the hundreds of kilometers above the earth when protons from space strike the molecules in the upper atmosphere and about 10,000 pass through every m^2 of the surface of the earth every minute. If relativity was not the correct description of space-time, then even if the muons travelled at the speed of light, they would only travel about 660 m before decaying ($2.2 \times 10^{-6}\mathrm{s} \cdot 2.998 \times 10^{8}\mathrm{m/s} = 660\mathrm{m}$). This is much shorter than the height of the atmosphere. The muons can survive from the upper atmosphere down to the earth due to the phenomena of time dilation or length contraction, depending on your reference frame, which are examples of space and time being boosted into each other. From the point of view of the muon, it is at rest and the surface of the earth is hurtling towards it at nearly the speed of light. However, the muon sees the atmosphere as significantly shorter than we measure it. From the point of view of the surface of the earth, the atmosphere is it's normal height, but the lifetime of the muon has been extended dramatically. In other words, the atmosphere is tall in its rest frame, and the muon is short-lived in its rest frame, but when you perform boosts between the frames, these values transform into each other loosely speaking. In a reference frame that is moving down slower than the muon, the atmosphere would appear shorter and the muon lifetime would apper longer.

G.2 The Basics of Particle Physics

G.2.1 Fields and Waves

A long time ago, people realized that certain rocks seem to push and pull each other even when held at a distance. This was called magnetism. They also noticed that their hair might point towards wool jackets in the winter. This was called electricity. Eventually, they also realized that the moon was held in orbit around the earth by some invisible force called gravity. To describe such phenomenon where a force seems to be exerted over empty space, we need to introduce the concept of a "field".

Fields are functions that have a value everywhere in space and at all points in time. Fields come in many forms; for example, there are scalar fields, vector fields, spinor fields, and tensor fields.

A scalar field has a scalar value everywhere in space. Temperature is a good example of a scalar field. If you stuck a thermometer at random places throughout a room, you would measure a single value at each place, and that value will probably vary slightly from place to place. If you held the thermometer in a single place for any length of time, you would measure different temperature values at different times.

A vector field has some direction everywhere in space, and its direction can be rotated as discussed in the previous section. Wind is a good example of a vector field. If you go outside you will find that at every location you stick your anemometer, the wind has a strength and direction, both of which can vary from place to place and from time to time. You will also find that if the wind is blowing from left-to-right, once you turn to the left, the wind is now blowing at your face; wind obeys the rules of rotations.

Where a scalar field has a scalar value everywhere, and a vector field has a vector value everywhere, a spinor field has a spinor value everywhere. Similarly, a tensor field has a tensor value everywhere.

The ultimate goal of any theory that describes a phenomenon with a field is to find a set of differential equations governing the *dynamics* of the field. This is to say, we want to know among other things how the value of a field at a particular place changes as a function of time, how the value at a particular place affects the values nearby, and how the field responds to sources, boundaries, and other fields. The dynamical equation will describe how the field's values propagate through space and time, telling us how the value at one place at one point in time will influence the value somewhere else later, if it has any influence at all. In the case of temperature for example, we might want to know how lighting a candle will affect the temperature throughout the room. The candle is a source of heat, and intuition correctly tells us that the temperature will be higher near the candle and lower further away. If we had a dynamical equation for heat, we could describe how quickly the temperature falls of as a function of distance from the candle or the rate at which the temperature in the corner of the room increases.

One of the most important dynamical equations in Physics, and one that shows up

frequently in field theories, is the Wave Equation:

$$\frac{d^2}{dt^2}u(t,x) = c^2\frac{d^2}{dx^2}u(t,x). \tag{G.17}$$

This equation can be understood in simple terms by imagining that the function $u(t,x)$ is the vertical displacement of a guitar string. The dependence on x means that the displacement varies from place to place along the string, and the dependence on t means that the displacement can change as a function of time. If the string is still and no one plucks it, then $u(t,x) \equiv 0$ - which is to say that it stays in a straight line forever.

The second time derivative of a segment of the string, say at x_0, is its acceleration: $\frac{d^2}{dt^2}u(t,x_0)$. The second space derivative of the string at a given time, say at t_0, is its spatial concavity: $\frac{d^2}{dx^2}u(t_0,x)$. The wave equation tells us that the acceleration of a segment of a string is related to the way the string is shaped at a given moment. If you pull the string up, then the segment near your finger will have negative concavity, so it experiences a force ($F = ma$) downwards. When you release the string, it accelerates downwards until it eventually has positive concavity, at which point it will accelerate upwards. This cycle will repeat, leading to harmonic motion.

The most common solution to this equation in physics is the "plane wave". This solution can be expressed as:

$$u(t,x) = Ae^{-i\omega t \pm ikx}. \tag{G.18}$$

Here $k = \frac{\omega}{c}$. The variable A is the height of the wave, ω is the angular frequency, k is the inverse of the wavelength, and c can be thought of as the speed of the wave. The sum of any two solutions to the wave equation is also a solution; this is known as the "superposition principle". In fact, it is possible to build any arbitrary solution to the wave equation by summing up plane wave solutions. Strictly speaking, this is done by integrating over modes, which are uniquely identified by the real numbers ω and k. The process of building solutions from plane wave sums is known as Fourier analysis.

G.2.2 Electromagnetism

By the end of the 19$^{\text{th}}$ century, Electromagnetism was the most well-developed and stringently tested field theory in physics. A little more than a century's worth of experiments had informed the development of four coupled first-order differential equations that governed the dynamics of the Electric and Magnetic fields in a vacuum, known as Maxwell's Eqations:

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \\
\nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{c^2 \partial t} &= \mu_0 \mathbf{J} \\
\nabla \cdot \mathbf{B} &= 0 \\
\nabla \times \mathbf{E} - \frac{\partial \mathbf{B}}{\partial t} &= 0.
\end{aligned} \tag{G.19}$$

The equations have been written here in SI units, hence the inclusion of c. Here, $\mathbf{E}$ and $\mathbf{B}$ represent the Electric and Magnetic fields, where the bold face script indicates that these are vectors under traditional 3D rotations. The ρ indicates a charge density, which is a scalar, and $\mathbf{J}$ indicates a current density, which is a vector. Charge has no direction, but current corresponds to a charge flow, and thus has direction. The values ϵ_0 and μ_0 are the vacuum permittivity and permeability respectively, which are defined through the formula $\epsilon_0\mu_0 = \frac{1}{c^2}$. They govern how the Electric and Magnetic fields propogate through free space. Lastly, the dynamics of the fields are expressed using the operator ∇, which is effectively a vectorial version of spatial differentiation, where $\nabla = \{\frac{\partial}{\partial x}, \frac{\partial}{\partial z}, \frac{\partial}{\partial z}\}$.

The dot $(\cdot)$ notation indicates the "divergence" of the field, which is a scalar: $\nabla \cdot \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}$. The cross-product $(\times)$ notation indicates the "curl" of the field, which is a vector: $\nabla \times \mathbf{B} = \{\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z}, \frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x}, \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\}$.

Qualitatively, the first equation in Eq. G.19 tells us that the Electric field points directly away from (towards) a positive (negative) charge. The second tells us that the Magnetic field curls around a current, which is a flow of charge. Similarly, if an Electric field changes with time, then a curling Magnetic field is induced. The third equation tells us that the Magnetic field is pure curl; it never has a net flow into or out of an enclosed volume. The last equation tells us that a curling Electric field is induced by a Magnetic field that changes with time. For example, if you move a bar magnet in and out of a loop of wire, a current is actually induced in the wire due to the induced Electric field.

There are many techniques to solve these equations for situations involving point charges, planes of charge, arbitrarily charged sphere, infinite lines of current, loops of current, solenoids, circuits, and on and on. However, the propagation of these fields through free space is what interests us the most in these contexts. In particular, if there are no sources around ($\rho = 0$ and $\mathbf{J} = \mathbf{0}$), we can rearrange these equations to find separate second-order partial differential equations for the two fields. First we can take the curl of the last equation in Eq. G.19:

$$\nabla\times(\nabla \times \mathbf{E}) - \nabla\times\frac{\partial \mathbf{B}}{\partial t} = \nabla\times 0 = 0.$$

An identity from Calculus and the fact that $\nabla \cdot \mathbf{E} = 0$ when $\rho = 0$ tell us that $\nabla\times(\nabla \times \mathbf{E}) = -\nabla^2\mathbf{E} = -(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})\mathbf{E}$. Similarly, we know that the curl of $\mathbf{B}$ is related to the time derivative of $\mathbf{E}$, so by putting the pieces together:

$$\nabla^2\mathbf{E} - \frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = 0.$$

The same type of manipulation can be performed for the Magnetic field:

$$\nabla^2\mathbf{B} - \frac{1}{c^2}\frac{\partial^2\mathbf{B}}{\partial t^2} = 0.$$

A glance and Eq. G.17 tells us that the the dynamics of the Electric and Magnetic fields in free space are governed by wave equations that propagate at the speed of light! In fact, these waves *are* light.

Solutions to these wave equations can be given by sums of plane waves solutions as discussed in the previous section. For many years in the late 19^{th} and early 20^{th} centuries, these plane wave solutions were viewed strictly as a mathematical tool, and it was believed that the energy carried by Electromagnetic waves was a function of the intensity of the light (its brightness) alone. It was Einstein who first identified the plane wave modes as particles, which came to be known as photons. The energy carried by a photon is a function of its frequency. He made this connection by studying the photoelectric effect: light above a certain frequency threshold can eject electrons from a material, which happens only if a certain amount of energy is deposited in the material at once [274]. Below the threshold, no electrons are ejected, regardless of the intensity of the light.

To sum up, the existence and dynamics of particles can be expressed by the use of fields. The particles are identified with particular modes that solve the dynamical field equations. In fact, the fields themselves can be thought of as a sum of superimposed particles.

G.2.3 Particles

The identification of plane waves modes with energetic particles is a quantum mechanical phenomenon. For a plane wave with spatial dependence $e^{ik\cdot x}$, after performing the two space derivatives in the wave equation, we are left with the equation of motion of a harmonic oscillator. The total energy of a quantum harmonic oscillator can be expressed as a sum of quanta, each with energy $E = \hbar\omega$. In quantum field theory, these quanta are interpreted as particles. Any field is essentially a sum of an infinite number of harmonic oscillators, one for each possible value of k, which is a 3D vector.

Beyond explaining the photoelectric effect and certain other outstanding thermodynamic problems at the time, the identification of particles with plane wave modes provides a convenient interpretation to the puzzle of wave-particle duality. It had been known since the time of Newton that light has wave-like properties, such as its diffraction and interference. The photoelectric effect had demonstrated that light has particle-like properties, but it was unclear how something could be both a wave and a particle. Similarly, from its discovery, electrons were known to be particles [275]. However, electrons can also diffract and interfere like waves [276]. If particles *are* the wave modes of fields, that explains why they behave like waves and can still carry energy and momentum. Just as the energy of a particle is given as a function of its frequency ($E = \hbar\omega$), its momentum is given as a function of its wave number: $\mathbf{p} = \hbar\mathbf{k}$.

I mentioned briefly in Section G.1.2 that every particle has a property called spin, and in particular some particles have spin that can be expressed as a 2-component vector that transforms under $SU(2)$. In other words, such particles have two spin-related degrees of freedom, and they are called "spin-$\frac{1}{2}$". Some particles have scalar spin; their one spin-related degree of freedom does not transform under rotations, and they are called "spin-0". And some particles have vector spin; their three spin-related degrees of freedom transform as vectors under $SO(3)$, and they are called "spin-1". It is important to note that all massless particles with spin > 0 have only 2 spin degrees of freedom, though that is not to say that

they are all represented by vectors in $SU(2)$. Spin-1 massless particles are still represented by $SO(3)$ vectors, but certain modes are unphysical. In some sense, this can be thought of as a consequence of relativity. Rotations of spin should be performed in the particle's rest frame, but a massless particle always travels at the speed of light and therefore has no rest frame, rendering some rotations inaccessible.

As an important bit of nomenclature, particles with integer spin are called "Bosons", and particles with half-integer spin are called "Fermions".

G.2.4 The Dynamics of Fields

In the Standard Model, which is the model that describes almost all known particle-related phenomena, there are only spin-0, spin-$\frac{1}{2}$, and spin-1 particles. Let us now attempt to embed all of those types of particles into fields. In particular, if we accept that relativity is the correct description of space-time, then we will try to embed our particles into fields that transform under the Lorentz group.

The first step to deriving dynamical equations is to construct a "Lagrangian". A Lagrangian is a scalar valued functional of a field and its space and time derivatives: $\mathcal{L}[\phi, \partial_\mu \phi]$. For a classical object, such as a ball, the Lagrangian is a function of the ball's position and velocity, which is roughly what the field and its derivatives correspond to. Note that ∂_μ stands for $\{\partial_t, \partial_x, \partial_y, \partial_z\}$. The field here ϕ is itself a function of space and time. The Lagrangian is defined by its relation to the "action", S:

$$S = \int d^4x \mathcal{L}[\phi, \partial_\mu \phi].$$

A particle's classical path is one that minimizes the action. The action is minimized when the following equation is satisfied:

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = 0. \tag{G.20}$$

Using this equation, we can find the traditional dynamical field equations. In classical physics, the Lagrangian is normally expressed as the difference between an objects kinetic and potential energies: $\mathcal{L} = T - V$. So for a ball with Kinetic energy $T = \frac{1}{2}mv^2$ and potential energy $V = mgy$, the Lagrangian is $\mathcal{L} = \frac{1}{2}mv^2 - mgy$. For motion only in the up-down y direction, Eq. G.20 tells us that $ma = mg$, exactly as expected from Newtonian mechanics.

The Lagrangian is a Lorentz scalar, so must construct our Lagrangians out of combinations of Fields and 4-dimensional derivatives that are themselves scalars. This severely constrains the possible terms we can include: this is a good thing, as it will make our search for the right equations much simpler. Before worrying about interactions between fields or fields that interact with themselves, we will first concern ourselves with non-interacting (potentially massive) fields.

Spin-0 particles have only 1 degree of freedom, so we should try embedding them in Lorentz scalar fields, which have only 1 component. Scalar fields are typically denoted by

$\phi(x)$. Here, x stands for 4-dimensional space time. The derivative of a scalar field, $\partial_\mu\phi$ is a vector, so it cannot be a term of a Lagrangian on its own. Still, we want a field that can change in space and time, so we need terms in our Lagrangian with derivatives. If we multiply the $\partial_\mu\phi$ vector by itself, we get a scalar: $(\partial_\mu\phi)(\partial_\mu\phi)$. This is a valid term! In fact, $\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)(\partial_\mu\phi)$ is the correct Lagrangian for a massless scalar field. The $\frac{1}{2}$ has been added for normalization. To include mass in the Lagrangian, we include the term $-\frac{1}{2}m^2\phi^2$, to get

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)(\partial_\mu\phi) - \frac{1}{2}m^2\phi^2. \tag{G.21}$$

Using Eq. G.20, we find that the dynamical equation for a scalar field is

$$(\Box + m^2)\phi = 0. \tag{G.22}$$

The $\Box$ symbol stands for $\partial_t^2 - \partial_x^2 - \partial_y^2 - \partial_z^2$. If $m = 0$, this is just a massless wave equation. If $m \neq 0$, then plane waves of the form $e^{i\omega t - ik\cdot x}$ still solve this equation, except that now $-\omega^2 + k^2 + m^2 = 0$. If we restore the $\hbar$'s and c's in this equation and note the relation of energy and momentum to frequency and wave number, we recover the Einstein energy equation: $E^2 = p^2c^2 + m^2c^4$ (or $E = mc^2$ in the rest frame).

You might ask why the Langrangian does not have terms that depend on a term that has ϕ alone or on a term with ϕ^3. These are valid Lorentz scalars after all. Terms with ϕ^3 actually imply a self-interacting field. Such terms are found in the Standard Model, though in this section we are only concerned with fields that don't self-interact. And though it is too complicated to get into here, a term linear in ϕ would require a coefficient with dimension of $[\text{mass}]^3$; it turns out that such terms are rendered irrelevant by quantum corrections.

Now that we understand scalars, we can turn our attention to spin-1 particles. Spin-1 particles have three degrees of freedom (or 2 in the massless case), so we can try to embed them in Lorentz vectors, which have four degrees of freedom, as they have four components. We will need our field equations to conspire in such a way to eliminate the fourth degree of freedom.

First, we can try to list valid scalar terms built out of Lorentz vectors, which we will denote as A_μ and derivatives. The simplest possible term is $A_\mu^2 = A_\mu A_\mu$, which will show up in our Lagrangian. A term of the form $\partial_\nu A_\mu$, which is a rank-2 tensor, would be reminiscent of the $\partial_\mu\phi$ term for scalars. We can construct a scalar as $\partial_\nu A_\mu \partial_\nu A_\mu$, which would again be a scalar. One might guess that the Lagrangian would be of the form $\mathcal{L} = -\frac{1}{2}\partial_\nu A_\mu \partial_\nu A_\mu + \frac{1}{2}m^2A_\mu^2$. This is not a bad guess, and would give us dynamical field equations of the form $(\Box + m^2)A_\mu = 0$. Unfortunately, that is really just four separate equations for scalar degrees of freedom, one for each index of A.

The next terms we might try that have derivatives in them are $A_\mu\partial_\nu\partial_\nu A_\mu = A_\mu\Box A_\mu$ and $A_\mu\partial_\mu\partial_\nu A_\nu$. If we choose a Lagrangian of the form

$$\mathcal{L} = \frac{1}{2}A_\mu\Box A_\mu - \frac{1}{2}A_\mu\partial_\mu\partial_\nu A_\nu + \frac{1}{2}m^2A_\mu^2,$$

then the equations of motion derived from Eq. G.20 actually force $\partial_\mu A_\mu = 0$, removing a degree of freedom, exactly as we want. It is common to use the identity $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, after which the Lagrangian has the form:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 + \frac{1}{2}m^2 A_\mu^2. \tag{G.23}$$

In the context of Electromagnetism, the Lorentz vector A_μ is taken as a four-vector that combines the scalar Electric potential, Φ, as the 0^{th} component and the Magnetic vector potential, A, as the spatial vector component. With this identification, we find that:

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix}. \tag{G.24}$$

In the limit that $m \to 0$, the equations of motion for these fields are exactly the Maxwell equations in free space. In fact, Eq. G.23 is the correct Lagrangian for a massless spin-1 particle when m is taken to be 0. Technically, this extension is not a trivial one, as another degree of freedom had to be eliminated, but the discussion of how this worked will be reserved to the next section. We can also take a moment to note that $F_{\mu\nu}$ is a true Lorentz tensor. When we perform Lorentz transformations on it, we will mix together the Electric and Magnetic fields; a "pure" Electric field in one frame becomes a mixture of Electric and Magnetic fields in another frame. The forces which we normally interpret as coming from separate Electric and Magnetic fields are due to a unified Electromagnetic field governed by the dynamics of Eq. G.23.

Lastly, we can look at spinor fields. These are a bit more complicated, as they transform under the $(\frac{1}{2}, 0)$ or $(0, \frac{1}{2})$ representations of the Lorentz group, and are complex-valued functions. As a bit of notation, we will spinors that transform under $(\frac{1}{2}, 0)$ "right-handed", and those that transform under $(0, \frac{1}{2})$ as "left-handed". We will denote spinors as ψ_R or ψ_L accordingly. Handedness is often visualized by imagining that are tops rotating as they fly through space: right handed spinors are spinning clockwise around their direction of motion, while left-handed spinors spin counterclockwise. A spinor dot-product is written as $\psi^\dagger\psi$, which just indicates that taking the complex conjugate of the first spinor is necessary. Lorentz transformations of the spinors can be found via the exponential mapping of the generators found in Eq. G.3. It turns out that $\psi_R^\dagger\psi_R$ and $\psi_L^\dagger\psi_L$ are not scalars, but $\psi_R^\dagger\psi_L$ and $\psi_L^\dagger\psi_R$ are.

If we look for terms with derivatives, we find that terms like $\psi_R^\dagger\Box\psi_L + \psi_L^\dagger\Box\psi_R$, while valid, just give the equations of motion of decoupled scalars, which is not what we are looking for. Motivated by the fact that the four-derivative is in the form of a 4-vector, we should try to construct a 4-vector using spinors. Through a slick bit of notation, we can write a list of 2x2 matrices involving the generators found in Eq. G.3 as

$$\sigma_\mu = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}.$$

A list where the last three matrices carry a factor of -1 is denoted by $\bar{\sigma}_\mu$. After examining their transformation properties, we actually find that $\psi_R^\dagger \sigma_\mu \psi_R$ and $\psi_L^\dagger \bar{\sigma}_\mu \psi_L$ are Lorentz vectors! It is possible to build a scalar Lagrangian of the form

$$\mathcal{L} = i\bar{\psi}_R \sigma_\mu \partial_\mu \psi_R + i\bar{\psi}_L \bar{\sigma}_\mu \partial_\mu \psi_L - m(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R). \tag{G.25}$$

This is the correct form of a Lagrangian for a massive spinor particle. The $\bar{\psi}$ notation here is the same as $\psi^\dagger$. In the limit that $m \to 0$, the right-handed and left-handed spinors are fully decoupled. A massless spinor field is either purely right-handed or purely left-handed, and there is no need for both a right-handed *and* left-handed version of any massless spinor field.

The equations of motion for this Lagrangian are can be written in the form

$$(i\sigma_\mu \partial_\mu - m)\psi_R = 0. \tag{G.26}$$

If this equation is multiplied by a factor of $(i\sigma_\mu \partial_\mu + m)$, we recover the wave equation.

There is one last important thing to note in this section. Earlier I mentioned that the relationship between energy (frequency) and momentum (wave number) satisfy the Einstein energy equation $E^2 = p^2c^2 + m^2c^4$. We should remember from Algebra that E actually has two solutions, one positive and one negative: $E = \pm\sqrt{p^2c^2 + m^2c^4}$. Solutions with negative energy are perfectly valid, but there is no such thing as a negative energy particle, so what do these solutions mean? The particles that correspond to these solutions are called "antiparticles", and they really do exist and are easy to produce in a lab. They satisfy exactly the same dynamical equations as their particle counterparts, so they have exactly the same mass and spin. The only difference is that they have exactly opposite charge. Neutral particles, like the photon, are their own antiparticles, but charged particles have distinguishable antiparticles. The antiparticle of the electron is called the positron, though most other particles' antiparticles are designated with the prefix "anti-" attached to them, like muon and "antimuon". Antiparticles can annihilate particles and vise versa. The concept of annihilation here is related to the quantum harmonic oscillator interpretation of particles referenced at the beginning of Sec. G.2.3. In the quantum harmonic oscillator, the quanta of energy can be created and destroyed by absorbing energy from some external source or by emitting energy. The interaction of a particle and an antiparticle can result in the destruction of both; however, the energy present prior to the annihilation will persist in some other form. For example, an electron and positron can annihilate, creating two photons.

Both ψ_L and ψ_R carry two degrees of freedom each. Separately, they can represent massless fermions. When mass terms such as $m(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R)$ appear, the two fields are combined. The four degrees of freedom in the theory are usually interpreted in this case as two particle degrees of freedom and two antiparticle degrees of freedom, where both the particle and the antiparticle are now massive fermions.

Throughout this section we have seen that we can construct correct Langrangians mainly by respecting the Lorentz group. When all the terms in a theory invariant under the transformations of a certain group, that group is considered a "symmetry" of the theory. Symmetries

are a foundational concept in modern particle physics. As we have seen, they dictate particle dynamics, and we will soon see that they dictate how particles interact.

G.2.5 Coupling Fermions to Vector Bosons

Now that we understand the dynamics of the fields we need to build the Standard Model, we can turn to the question of how they interact. For instance, we know that electrons, which are charged particles, must interact with the Electromagnetic field. We can embed electrons into spinor fields, and the Electromagnetic field is expressed as a vector field. Somehow we must have a term in the Lagrangian that involves both of these fields. The form of this term will be guided by a "local gauge symmetry".

A gauge symmetry occurs when transformation under the action of a Lie group is imposed on fields, but the Lagrangian is invariant after such transformations. The Lorentz group is not a gauge symmetry, as the fields are defined by and exist in 4D Lorentz space. A gauge theory is applied on top of that. For example, we could imagine that a spinor field transforms under the action of the group $U(1)$. That is to say that it is able to acquire a complex phase of the form

$$\psi \to e^{-i\alpha}\psi.$$

The gauge symmetry is considered local if the α is allowed to vary from place to place, or in other words, it is itself a function of space-time: $\alpha(x)$.

What are the consequences of imposing this gauge symmetry? If we refer back to Eq. G.25, we find that the $m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$ term is invariant. However, the term with derivatives is *not* invariant. That term for right-handed fields transforms as

$$i\bar{\psi}_R\sigma_\mu\partial_\mu\psi_R \to i\bar{\psi}_R e^{i\alpha}\sigma_\mu\partial_\mu e^{-i\alpha}\psi_R = i\bar{\psi}_R\sigma_\mu(\partial_\mu - i\partial_\mu\alpha)\psi_R.$$

If our Lagrangian is not invariant, does that mean the whole idea is useless? Not quite! If we have a theory that has a spinor field *and* a massless vector field, we can save ourselves if the vector field transforms as

$$A_\mu \to A_\mu + \frac{1}{q}\partial_\mu\alpha.$$

Actually, the invariance of A_μ to transformations of this form is an important feature known from Classical Electromagnetism. Choosing a gauge is actually how the second degree of freedom was removed for massless spin-1 fields in the previous section. However, the choice of gauge doesn't change the form of the equations, so it can be slipped in without much notice.

If A_μ transforms in this way, then $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is invariant, so $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2$ is invariant. That's good! If we make the prescient choice of constructing a Lagrangian term of the form $i\bar{\psi}_R\sigma_\mu(\partial_\mu + iqA_\mu)\psi_R$, we find that it will transform as

$$i\bar{\psi}_R\sigma_\mu(\partial_\mu + iqA_\mu)\psi_R \to i\bar{\psi}_R e^{i\alpha}\sigma_\mu(\partial_\mu + iqA_\mu + i\frac{q}{q}\partial_\mu\alpha)e^{-i\alpha}\psi_R = i\bar{\psi}_R\sigma_\mu(\partial_\mu + iqA_\mu)\psi_R \quad \text{(G.27)}$$

after carrying out the differentiation. This term is invariant, and we have a term that involves both our spinor *and* our vector field. The strength of the coupling between ψ and A is governed by q, which we can call "charge". In fact, if we identify the term $q\bar{\psi}_R\sigma_\mu\psi_R$ as a current, J_μ, then we can see in the above equation a factor of $J_\mu A_\mu$, which is part of the Lagrangian of Classical Electrodynamics.

As a quick note, the dynamical equations for the spinor field that result from a Lagrangian of the form of Eq. G.27 would be $(i\sigma_\mu\partial_\mu - e\sigma_\mu A_\mu)\psi_R = 0$ and $i\partial_\mu\bar{\psi}_R\sigma_\mu + e\bar{\psi}_R\sigma_\mu A_\mu = 0$. The sign of the coefficient in the term that couples ψ_R and A is opposite of that in the conjugate equation. The typical interpretation due to this is that the ψ_R field creates particles and annihilates antiparticles and that its conjugate, $\bar{\psi}_R$, creates antiparticles and annihilates particles.

It is common to define a "covariant derivative" to simplify how our Lagrangians look:

$$D_\mu = \partial_\mu + iqA_\mu.$$

With this notation, a Lagrangian that is consistent with Lorentz symmetry and with a $U(1)$ gauge symmetry can be written as:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 + i\bar{\psi}_R\sigma_\mu D_\mu\psi_R + i\bar{\psi}_L\bar{\sigma}_\mu D_\mu\psi_L - m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R). \tag{G.28}$$

The form of the interaction between ψ and A was dictated by the gauge symmetry, but remarkably, this is the Lagrangian that correctly encapsulates Electromagnetism. For whatever reason, gauge symmetries are a defining concept of the theories that mathematically represent nature. We should be careful to note that while Electromagnetism is a $U(1)$ gauge theory, not every $U(1)$ gauge theory will be Electromagnetism.

The other gauge groups that appear in the Standard Model are $SU(2)$ and $SU(3)$, both of which were introduced in Section G.1.2. The key difference between these two groups and $U(1)$ is that their generators do not commute: $SU(2)$ and $SU(3)$ are non-Abelian groups. Whereas a $U(1)$ transformation just added a complex phase to a spinor field, $SU(2)$ and $SU(3)$ transformations will require the introduction of doublets or triplets of spinor fields. For example, if we had two spinors, ϕ_1 and ψ_2, we could stack them on top of each other in a single doublet that transforms under elements of $SU(2)$. Note that the Lorentz transformations for these individual spinors are still represented by elements of $SU(2)$, but those tranformations are among the two components of each ψ separately. The gauge $SU(2)$ is totally unrelated and mixes ψ_1 and ψ_2:

$$\Psi = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \to e^{i\alpha_1\sigma_1 + i\alpha_2\sigma_2 + i\alpha_3\sigma_3}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}.$$

We have used the exponential map represention for elements of $SU(2)$ using the generators from Eq. G.3. If we Taylor expand the exponential for infinitesimal values of α, the transformation takes the form

$$\Psi \to \Psi + i\alpha_a\sigma_a\Psi,$$

where Einstein summation notation has been used to indicate a sum of three matrices, where each has a unique coefficient.

To preserve the gauge symmetry, we will once again introduce a covariant derivative, now with three vector bosons:

$$D_\mu \Psi = \partial_\mu \Psi - igA^a_\mu \sigma^a \Psi. \tag{G.29}$$

Unfortunately, if we use same vector boson transformations as before, $A^a_\mu \to A^a_\mu + \frac{1}{g}\partial_\mu \alpha^a$, we would find that our Lagrangian is not invariant. Terms with a factor of $\sigma_a\sigma_b - \sigma_b\sigma_a$ are introduced. Luckily, we know that this is just a commutation, and we know the commutation relations for the $SU(2)$ generators from Eq. G.5. To eliminate these terms, the transformations of the vector bosons must now take the form

$$A^a_\mu \to A^a_\mu + \frac{1}{g}\partial_\mu \alpha^a - f^{abc}\alpha^b A^c_\mu. \tag{G.30}$$

Every non-Abelian group has a unique set of structure constants, labelled here as f^{abc}. For $SU(2)$, $f^{abc} = \epsilon^{ijk}$, as defined below Eq. G.5.

With the introduction of the additional term to the vector boson gauge transformation, the old kinetic term for vector bosons is no longer invariant. If the vector bosons transform under a non-Abelian gauge, then the correct kinetic Lagrangian is

$$\mathcal{L} = -\frac{1}{4}\sum_a (\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu)^2. \tag{G.31}$$

This term is often written as $\mathcal{L} = -\frac{1}{4}(F^a_{\mu\nu})^2$, where Einstein notation is used to sum over μ, ν, and a. Here, $F_{\mu\nu}$ has been redefined with a structure-constant dependent term: $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu$. This can even be taken as the form of $F_{\mu\nu}$ for the Abelian case of $U(1)$, because there the structure constants are just 0, and there is only one boson for the sum over a.

The logic for the constructing a Lagrangian in the case of an $SU(3)$ gauge symmetry is exactly the same as for the $SU(2)$ case, except for the presence of 8 vector bosons now for the 8 generators, and different structure constants.

In the case of a non-Abelian gauge group, we followed a somewhat complicated progression of introducing spinor doublets, introducing a number of vector bosons equal to the number of gauge group generators, observing that a new term must be added to the vector boson gauge transformation, and adjusting the kinetic Lagrangian term for the vector bosons accordingly. Though this process might seem complicated at first glance, it is primarily an example of the guiding power of symmetry. $SU(2)$ and $SU(3)$ gauge groups both appear in the Standard Model, and the associated terms in the full Standard Model Lagrangian use the covariant derivative in Eq. G.29 and kinetic terms like Eq. G.31.

There is one last property of Eq. G.31 which has consequences that are highly relevant to this thesis. If we multiply out the square, we will find terms with factors of $(\partial_\mu A_\nu)A_\mu A_\nu$ and $A_\mu A_\nu A_\mu A_\nu$. The vector bosons can self-interact! This is a stunningly different behavior than for Electrodynamics, where photons cannot self-interact.

G.2.6 Scalar Fields and the Higgs Mechanism

The previous section was about coupling spinor fields to vector fields via gauge symmetry principles. The same principles can be applied to scalar fields, but with a slight twist. Spinor fields are complex so Lie group transformations, which are in general complex, will change the complex phase of the field. Scalar fields are real, and cannot be complexified by themselves. In order to create a complex scalar field, two scalars are actually needed. They can be combined as

$$\Phi = \phi_1 + i\phi_2.$$

The rest of the gauge principle follows the same way as in the spinor case. Now that the scalar field has been complexified, the Langrangian for a scalar coupled to a vector field could be written as

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 + (D_\mu\Phi)^\star(D_\mu\Phi) - m^2\Phi^\star\Phi, \tag{G.32}$$

where the use of the complex conjugate makes the Lagrangian real and invariant.

Scalar fields can be used to do something interesting in particle physics. Having spent the last several sections building up Lagrangians using symmetry principles, we can use scalar fields to break the symmetries. What I mean by this is that certain symmetries can be used to create Lagrangians, but a scalar field can be used to make the theory appear to *not* have some of those symmetries in real life. The principle of breaking symmetries using scalar fields is called the Higgs Mechanism, and it has many important consequences.

As an example of this mechanism, we can imagine a theory that has both a $U(1)$ and an $SU(2)$ gauge symmetry. If this is the case, then there will be four vector bosons in the theory: one for the $U(1)$ gauge and three for the $SU(2)$ gauge. There can be an arbitrary number of spinors, which can be charged under both, one, or neither gauge. "Charged" here means that the spinor field couples to the gauge boson of the field. For example, if a particle has 0 electric charge, that means that its field simply does not acquire a phase under the $U(1)$ transformations that generate Electromagnetism. For the sake of argument, let us assume that this theory has one massless, left-handed spinor that is charged under both the $SU(2)$ and $U(1)$ gauges, and one massless, right-handed field that is only charged under the $U(1)$ field. Let us set the $U(1)$ charge of Ψ_L to $-\frac{1}{2}$ and the charge of ψ_R to -1. In this case, the left-handed field must be a doublet spinor. In total, prior to the introduction of any scalar fields, this theory has four gauge bosons, a left-handed spinor field that can be written as $\Psi_L = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}$, and a right-handed spinor field, which we will call ψ_R. The Lagrangian will be written as

$$\mathcal{L} = -\frac{1}{4}(W_{\mu\nu}^a)^2 - \frac{1}{4}B_{\mu\nu}^2 + i\bar{\Psi}_L(\partial\!\!\!/ - igW\!\!\!/^a\sigma^a + i\frac{1}{2}g'B\!\!\!/)\Psi_L + i\bar{\psi}_R(\partial\!\!\!/ + ig'B\!\!\!/)\psi_R. \tag{G.33}$$

Terms like $W_{\mu\nu}$ stand for $W_{\mu\nu} = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + gf^{abc}W_\mu^b W_\nu^c$. The three W_μ^a fields are those associated with the $SU(2)$ gauge, and the B_μ field is the one associated with the $U(1)$ gauge. The W^1 and W^2 fields are associated with the off-diagonal generators in Eq. G.3,

and are thus associated with transformations that switch ψ_1 and ψ_2. The W^3 generator is associated with the diagonal generator, and thus does not switch the two fields. The sign and coefficient of the coupling to the B_μ field has been influenced by the choice of charge. The slashed notation indicates that a Lorentz vector (here ∂_μ, W^a_μ, and B_μ) has been dotted with the σ_μ matrices, as is seen in Eq. G.28. There are two real constant parameters present in this Lagrangian, g and g', which are field coupling constants. The Lagrangian in Eq. G.33 is a scalar under Lorentz tranformations, respects both gauge symmetries, and even exhibits a new symmetry called "chiral symmetry". Chiral symmetry occurs when left-handed spinors and right-handed spinors can undergo separate transformations while still leaving the full Lagrangian invariant. Here, an $SU(2)$ transformation will affect Ψ_L but not ψ_R, yet the Lagrangian does not change. Mass terms for spinors, such as those found in Eq. G.28, generally violate chiral symmetry. A transformation on a left-handed field alone in a term such as $m\bar{\psi}_R\psi_L$ would change the Lagrangian.

Now let us examine the consequences of adding a scalar field. If the field is charged under $U(1)$, it must be a complex field of the form $H = h_1 + ih_2$. If it is charged under $SU(2)$, then it must be a complex doublet of the form $H = \begin{pmatrix} h_1 + ih_2 \\ h_3 + ih_4 \end{pmatrix}$. Such a field is also allowed to be charged under $U(1)$. Let us allow our scalar field to have charge 1 under $SU(2)$ and $\frac{1}{2}$ under the $U(1)$. If such a field is included in the theory, then Lorentz symmetry and gauge symmetry allow us to have terms that are proportional to $H^\dagger H$, $(H^\dagger H)^2$, and $(D_\mu H)^\dagger (D_\mu H)$. Here D_μ is the covariant derivative: $D_\mu = \partial_\mu - igW^a_\mu \sigma^a - i\frac{1}{2}g'B_\mu$. From this we see that the dynamical scalar field can be coupled to both itself and to the vector bosons as expected. There is also an allowed term that will couple the scalar to the spinors. We can write a term that is proportional to $\bar{\Psi}_L H \psi_R$. The Lorentz invariance of this term is manifest, as it is a combination of spinors terms such as $\bar{\psi}_1\psi_R$, which is a scalar, and explicit scalars, such as h_1. The gauge invariance of this term was accomplished through a bit of numerology. Transformations of $\bar{\Psi}_L$ and H in this term under will cancel out, and $U(1)$ transformations will get a factor of $e^{\frac{i}{2}}e^{\frac{i}{2}}e^{-i} = 1$, thanks to the judiciously chosen $U(1)$ charges of the fields. With the addition of the scalar field, we can now have the Lagrangian:

$$\begin{aligned} \mathcal{L} = & -\frac{1}{4}(W^a_{\mu\nu})^2 - \frac{1}{4}B^2_{\mu\nu} + (D_\mu H)^\dagger (D_\mu H) + \frac{1}{2}m^2 H^\dagger H - \lambda (H^\dagger H)^2 \\ & + i\bar{\Psi}_L(\not{\partial} - ig\not{W}^a\sigma^a + i\frac{1}{2}g'\not{B})\Psi_L + i\bar{\psi}_R(\not{\partial} + ig'\not{B})\psi_R \\ & - y\bar{\Psi}_L H\psi_R + h.c. \end{aligned} \tag{G.34}$$

Three new constant real parameters have been introduced to go along with g and g': m, λ, and y. The $h.c.$ term just stands for "hermitian conjugate", which here is $h.c. = y\bar{\psi}_R H^\dagger \Psi_L$. In this example, we now have two unbroken gauge symmetries, five real constant parameters, four massless vector bosons, two massless spinors, and one complex doublet scalar field. If we were to count field degrees of freedom, we would have $4 \times 2 = 8$ from the vector bosons, $3 \times 2 = 6$ from the spinors, and 4 from the scalars.

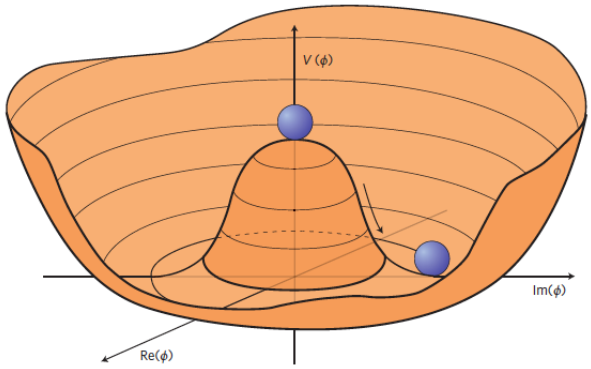


Figure G.1: Form of the Higgs potential when $\lambda > 0$ [277]. The vacuum expectation value of the scalar field is then non-zero.

However, something interesting might happen depending on our (or maybe nature's) choice of λ. If $\lambda > 0$, then the coefficients in the "Higgs potential" term $V(H) = -\frac{1}{2}m^2H^\dagger H + \lambda(H^\dagger H)^2$ have opposite sign. The resulting potential has a form as shown in Fig. G.1. The dynamical equations derived from Lagrangians push the values of the fields and their derivatives to minimize the potential. The same is true of Lagrangians for classical objects, ensuring that objects fall in gravitational fields or want to sit at a spring's equilibrium point. In this case, the potential is minimized when the value of H is non-zero! This non-zero value of the function is referred to as a vacuum expectation value, or *vev*. As this is a scalar field, a *vev* analogous to the temperature in a room having a constant, non-zero value everywhere, say room temperature. The scalar field is dynamical, so it is allowed to deviate from the *vev* from place to place and from time to time in the same way that temperature will vary from place to place in a room, but the value is centered around the *vev* rather than being centered around 0. By finding the minimum of the potential we find that the *vev* for H is $v = \frac{m}{\sqrt{\lambda}}$. The potential in Fig. G.1 is shown for scalar field of the form $\Phi = \phi_1 + i\phi_2$, and the *vev* is shown pointing in some direction in $\{\phi_1, \phi_2\}$ space. For the potential in Eq. G.34, we have the freedom to choose which direction the *vev* points in $\{h_1, h_2, h_3, h_4\}$ space, and the traditional direction is in the h_3 direction.

Having found the magnitude of and selected a direction for the *vev*, we can rewrite the H field as

$$H = e^{2i\frac{\pi^a\sigma^a}{v}} \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} + \frac{h}{\sqrt{2}} \end{pmatrix}. \tag{G.35}$$

The four scalar degrees of freedom are now split between the three scalar π^a fields, which have no *vev*, and the scalar h field, which is in the original h_3 direction and is centered around the *vev* of H. This sounds complicated, but it is really just a recentering and rescaling of the fields, analogous to measuring temperature in a room relative to room temperature. If room temperature is 75°, and you measure values of 76° and 73° in different places in the room,

those values would be 1° and -2° in the recentered units, assuming you kept the degree scale the same. For temperature measurements, that would probably be a little impractical, but it will considerably simplify the analysis of our example field theory. It also turns out that we do not need to explicitly carry the π^a fields around as we rewrite equations, these degrees of freedom can be absorbed by the vector fields as we will see.

With H now in the form $\begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} + \frac{h}{\sqrt{2}} \end{pmatrix}$, let us examine Eq. G.34. If we look at the parts of $(D_\mu H)^\dagger (D_\mu H)$ term have a factor of v rather than h, we find:

$$(D_\mu H)^\dagger (D_\mu H) = g^2 \frac{v^2}{8} \Big[(W^1_\mu)^2 + (W^2_\mu)^2 + \Big(\frac{g'}{g} B_\mu - W^3_\mu \Big)^2 \Big]. \tag{G.36}$$

These are just vector boson mass terms that are proportional to v! The $\psi_1 \leftrightarrow \psi_2$ switching W^1 and W^2 fields have their own terms. The fields W^3 and B do not switch those fields so they do not have conflicting effects in $SU(2)$ space. We can re-express Eq. G.36 in such a way that we remove $W_\mu B_\mu$ terms by mixing the fields as

$$\begin{aligned} Z_\mu &= \cos\theta_w W^3_\mu - \sin\theta_w B_\mu \\ A_\mu &= \sin\theta_w W^3_\mu + \cos\theta_w B_\mu. \end{aligned} \tag{G.37}$$

Here the mixing angle, θ_w is defined by $\tan\theta_w = \frac{g'}{g}$. Having done that, the kinetic terms for the Z_μ and A_μ fields will now be

$$-\frac{1}{4}(F_{\mu\nu})^2 - \frac{1}{4}(Z_{\mu\nu})^2 + \frac{1}{2} m_Z^2 (Z_\mu)^2. \tag{G.38}$$

Here, $m_Z = \frac{1}{2\cos\theta_w} g v$, which is a purely a function of g, g', and v from the original parameter set. From this, we see that Eq. G.36 gives us three massive vector bosons and one massless boson. Massive vector bosons have three degrees of freedom though. These used to all be massless bosons, so where did the new degrees of freedom come from? These degrees of freedom are actually come from the π^a fields in Eq. G.35. It is often said that the newly-massive vector bosons "ate" the degrees of freedom from the scalar fields.

There are also terms in the third line of Eq. G.34 that will now depend on v. We find

$$-y\frac{v}{\sqrt{2}}(\bar{\psi}_2\psi_R - \bar{\psi}_R\psi_2).$$

This is just a fermion mass term where the mass is $m = y\frac{v}{\sqrt{2}}$! The ψ_1 field does not get a mass term in this theory. But still, we have another remarkable phenomenon. In the Lagrangian in Eq. G.34, Ψ_L and ψ_R are two unrelated massless fermion fields, but after the scalar field acquires a *vev*, the fields have been joined into a massive fermion field. If we look at how these fields couple to the A_μ field, we get

$$e\Big\{ \big(-\frac{1}{2} + Y_L \big) \bar{\psi}_2 \not{A} \psi_2 + \big(-\frac{1}{2} + Y_L \big) \bar{\psi}_1 \not{A} \psi_1 + Y_e \bar{\psi}_R \not{A} \psi_R \Big\}, \tag{G.39}$$

Where Y_L and Y_e refer to the charges of the fields Ψ_L and ψ_R under the $U(1)$ gauge, which are $-\frac{1}{2}$ and -1, respectively. Given that, we see that ψ_2 and ψ_R couple to the massless A vector field with the same charge.

In fact, all of the dynamics of the ψ_2, ψ_R, and A fields are exactly those of Electromagnetism with a massive fermion-antifermion pair. With this identification, the A field generates the photon, and the massive fermion field contains the electron-positron degrees of freedom. The massless ψ_1 field does not interact with A, so we can identify it as a neutrino.

Lastly, we can consider the terms that involve h. This is a single, real scalar field. The dynamical terms in the Lagrangian that involve this field are:

$$-\frac{1}{2}h(\Box + m_h^2)h - g\frac{m_h^2}{2vg}h^3 - \frac{g}{8}\frac{m_h^2}{v}h^4. \tag{G.40}$$

These terms represent a scalar particle with mass $m_h = \sqrt{2}m$ (the m there is the same as the one from Eq. G.34). This massive particle couples to itself both cubically and quadratically.

Let's quickly take stock of where we started and where we are now. We started out with a theory with $SU(2)$ and $U(1)$ gauge symmetries. There were four massless vector boson fields, two massless, left-handed fermion fields combined into an $SU(2)$ doublet, and a massless, right-handed fermion field that was only charged under $U(1)$. A complex, $SU(2)$-doublet scalar field with a *vev* was introduced, as were terms that coupled this scalar to the vector bosons and the fermions. This resulted in a theory with three massive vector bosons, one massless vector boson, a massive fermion, a massless fermion, and a massive scalar. The massless vector boson can be identified as the photon, and the two massless fermion fields that couple to it have the same charge, so the resulting massive fermion is consistent with a charged fermion, such as an electron. In fact, the original $SU(2)$ and $U(1)$ gauge symmetries have been broken. It is no longer possible to perform these gauge rotations while leaving the Lagrangian invariant. However, there is a new $U(1)$ symmetry in the theory associated with the massless vector field. This new $U(1)$ combines certain aspects of the old gauges and is of the same form as the $U(1)$ gauge symmetry that generates classical Electromagnetism.

The masses of the the vector bosons and fermions were created because of their coupling to the scalar field. This aspect of the Higgs mechanism is why the associated h field, the Higgs boson, is called the "God particle". Without it, particles would not have mass, at least if you start with a Lagrangian resembling Eq. G.34. Without mass, all the energy in the universe would stream around at the speed of light, and we certainly would not exist.

G.2.7 Feynman Diagrams

Having understood how to express the dynamics of fields, how to couple the different types of particles using symmetries, and how to generate particle masses, we now have all of the theoretical tools needed to construct the Standard Model. Before we address the full Standard Model though, it will be useful to introduce a tool for visualizing processes in particle physics, which we call "Feynman Diagrams".

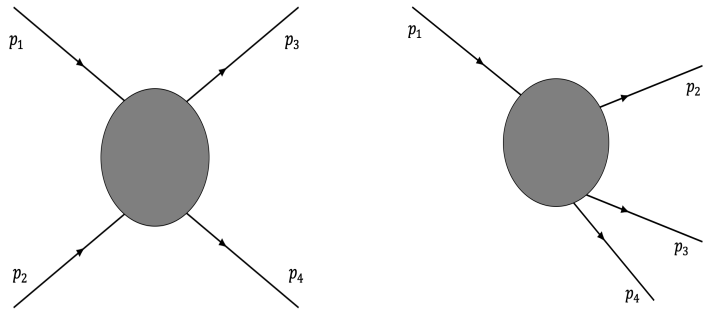


Figure G.2: Right: depiction of $2 \to 2$ scattering. This diagram depicts two particles with momenta p_1 and p_2 approaching each other and then scattering. After the scattering the two particles have new momenta, p_3 and p_3. Left: depiction of particle decay. A single particle with momentum p_1 decays into three particles with momenta p_2, p_3, and p_4. Here time flows from left to right, and distance is depicted by distance in the up-down direction.

The most basic goal of particle physics to predict how particles will behave and interact. In an ideal world, measurable quantities, such as particle decay rates and scattering cross-sections, should be calculable starting from the Lagrangian. For example, given two particle approaching each other with known momenta as seen in the left of Figure G.2, we might want to predict how likely it is that the particles scatter off of each other into directions specified by p_3 and p_4. Similarly, we might want to calculate how long it takes for an unstable particle to decay, such as the one shown in the right of Figure G.2. In the case of the decay-rate calculation, the directions of the final state particles is not particularly important as long as momentum is conserved. In both cases, the shaded blob simply depicts the fact that *something* happens to cause the particles to change momenta or to allow it to decay.

Unfortunately, the differential equations governing particle dynamics are not exactly solvable, so the *something* that happens in the shaded region cannot be expressed as a single simple term. Instead these interactions can be expressed as the sum of an infinite number of processes. The sum must have a finite result, as there can't be a transition probability greater than 1 in the real world. Sums of infinite numbers can be finite when each successive number added to the sum is smaller than the last. For example $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + ... = 2$. Solving a differential equation by summing up different contributions to the solution is known as perturbation theory. When we are lucky, it is easy to determine which processes are the most relevant to the sum and which are less relevant. In general, we just have to be careful to not to miss any part of the sum.

Terms in the Lagrangian actually tell us the form of the different components of the sum. For example, if we consider the theory in Eq. G.34, there is a massive electron, a massless photon, and a massive vector boson that couples to electron currents, which we can denote as Z. If we were to consider the $2 \to 2$ scattering of two electrons off of each other, then the two electrons can exchange three different bosons to transfer momentum to each other.

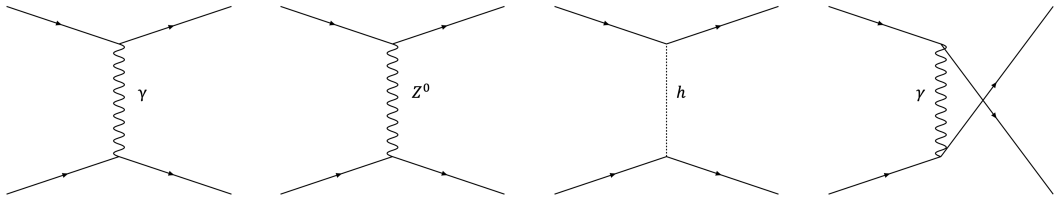


Figure G.3: Feynman diagrams of the leading order processes contributing to $2 \to 2$ electron-electron scattering based on the Lagrangian in Eq. G.34. The three left diagrams demonstrate the different bosons that can be exchanged between the two electrons, allowing them to exert forces on each other. There are two additional diagrams like the right-most one for the two other mediator possibilities. The two Fermion lines after the photon vertices do not interact when the lines cross.

These possibilities are shown in Figure G.3. Because electrons are identical particles, there are 6 diagrams resulting from the three mediator possibilities; it is impossible to tell which of the initial electrons becomes the one that would be labelled as p_3 in Figure G.2, and which would be labelled as p_4. In Feynman diagrams, fermions are drawn as straigt solid lines with an arrow. An arrow that points from left to right indicates a particles, and an arrow that points from right to left indicates an antiparticle. Vector bosons are indicated by wiggling lines, and scalar bosons are depicted by straight dashed lines. Only certain combinations of particles are allowed to interact at vertices. If there are Fermions, there must be an even number of fermion lines touching the vertex, and there must be an equal number of arrows on the Fermion lines pointing into the vertex as away from it with respect to the time axis (left to right). Odd numbers of bosons are allowed to interact at a single vertex. Interactions that are depicted by three bosons and four bosons interacting at a single vertex are particularly important for this thesis. Momentum must be conserved at every vertex, and energy present in the initial state must be conserved in the final state. Particles that are not in the initial or final state of the diagram are called "virtual particles", and they do not not have to satisfy the equation $E^2 = p^2 + m^2$. This temporary violation of energy occurs only for a brief instant in time, and is allowed due to the uncertainty relationship between time and energy: $\Delta t \Delta E \geq \frac{\hbar}{2}$. When this occurs, the particle is referred to as "off-shell". Virtual particles do not really exist, they are just a tool for doing calculations using perturbation theory. Particle decay is mediated by virtual particles. For example, a muon can radiate a highly off-shell W boson, resulting in an electron and two neutrinos. In quantum machanics, anything that can happen will happen, but once the decay occurs, it is irreversible.

Each vertex in a diagram represents field interactions. The allowed field interactions and their strengths are expressed by terms in the Lagrangian. For example, after the scalar field acquires its *vev* in the Lagrangian, an electron current (when both of the fermion lines pointing into and out of a vertex are electrons) can interact with the photon (denoted by γ), the massive boson in Eq. G.38 (denoted by Z^0), and the massive scalar boson (h). The

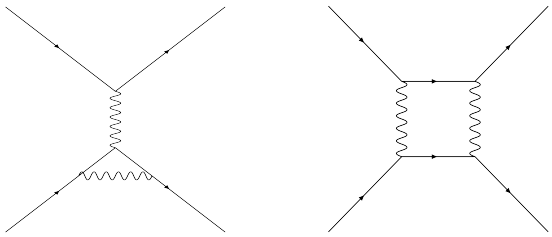


Figure G.4: Example Feynman diagrams of processes involving loops that contribute to $2 \to 2$ electron-electron scattering based on the Lagrangian in Eq. G.34. There are many more diagrams for this process that involve a single loop, and many more than involve more than one loop.

strength of the interaction of the electron with the photon can be read off from Eq. G.39: every vertex that involves two electrons and one photon brings a factor of the electorn charge, e. In natural units, the electron charge is a dimensionless value that is about 0.3. As the left-most diagram in Figure G.3 has two such vertices, the whole diagram carries a factor of $0.3 \cdot 0.3 = 0.09$. The diagram involving the Z^0 boson will also involve two factors of e (and a few additional factors), but it will be further suppressed relative to the photon diagram because the Z boson will likely be very off-shell, depending on the energy of the electrons. The coupling of the electron to the h boson is proportional to its mass, which is very small, so the diagram involing the h boson will involve two factors of m_e. So of these diagrams, the one with an intermediary photon is the most important, and the others providing subtle modifications. However, all of these diagrams are considered to be tree diagrams because they do not involve loops.

Two example diagrams involving loops are shown in Figure G.4. There are many, many more possible diagrams that involve one loop and even more that involve multiple loops. All diagrams that involve loops with photon-electron vertices will bring in extra factors of e, leading them to contribute to the sum less strongly than the photon diagram in Figure G.3. The presence of loops brings with it the true quantum heart of particle physics. Whereas in leading order diagrams, the momentum of intermediate particles is constrained to conserve the momenta of incoming and outgoing particles, when loops are present, all possible intermediate momenta must be summed over. This has important consequences which will be discussed later.

When the coupling of the fields involve small factors such as e or m_e, it is relatively easy to determine which diagrams are the most important to the sum simply by counting the numbers of vertices. Tree diagrams involve the fewest vertices, so they contribute the most strongly to the perturbation expansion. They are called leading order processes. Loop diagrams provide corrections to the leading order ones.

Unfortunately, it is possible that the fields in a theory are not coupled by a small number. This makes it impossible to order diagrams by significance. When this occurs, and it does

occur in practice, performing accurate calculations becomes intractable using perturbation theory techniques.

G.3 Conclusions

Hopefully these lecture notes have provided a somewhat-accessible introduction to particle physics. Some material that would normally be included in these notes has been incorporated into Chapter 2, such as sections on the Standard Model, Electroweak theory, and cross-sections. However, these topics are relevant to the structure of the thesis in general.