

A MEASUREMENT OF BOOSTED DIBOSONS WITH GAUSSIAN PROCESS
BACKGROUND MODELING AT THE ATLAS DETECTOR

Riley Xu

A DISSERTATION

in

Physics and Astronomy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2024

Supervisor of Dissertation

Elliot Lipeles

Professor, Physics and Astronomy

Graduate Group Chairperson

Ravi K. Sheth

Professor, Physics and Astronomy

Dissertation Committee

Evelyn Thomson, Professor, Physics and Astronomy

Dylan Rankin, Assistant Professor, Physics and Astronomy

James Aguirre, Associate Professor, Physics and Astronomy

Justin Khoury, Professor, Physics and Astronomy

Elliot Lipeles, Professor, Physics and Astronomy



A MEASUREMENT OF BOOSTED DIBOSONS WITH GAUSSIAN PROCESS
BACKGROUND MODELING AT THE ATLAS DETECTOR

COPYRIGHT

2024

Riley Xu

All rights reserved.

Acknowledgements

Firstly I would like to thank Elliot Lipeles for being an amazing advisor. I of course learned an incredible amount from him, but I especially appreciate his thoughtfulness and attentiveness. Together with the other Penn ATLAS faculty, Evelyn Thomson, Joe Kroll, Brig Williams, and Dylan Rankin, they have created a uniquely collaborative and supportive group. The open and friendly nature of the group has made my time at Penn immensely enjoyable, and I cannot overstate how much I treasure it. I also thank the other students and postdocs of the group who have helped foster this amazing environment. In particular I thank Elliot's senior graduate students Christian Herwig and Ben Rosser for being amazing mentors and helping me get acclimated.

In addition I would like to thank the Penn engineering team and especially Adrian Nikolica for their mentorship on my initial firmware projects. Having never really worked with hardware or firmware before, I would have been completely lost without their meticulous guidance.

I want to also thank the members of the HTTSim team, especially Jahred Adelman and Francesca Pastore. It was a pleasure working with them on tearing up the old FTKSim and creating a new and improved framework, despite the numerous bugs and headaches we encountered. Thanks also to the members who joined after HTTSim was migrated to EF Tracking, especially Stephanie Majewski for expertly leading the new project.

I also extend my gratitude to the many members of the VV Semileptonic analysis, particularly Rob Les and Dominik Duda for their work organizing and spearheading the analysis. Such a massive project could only be brought together due to the team's dedicated efforts and collaboration. I am extremely grateful to have conducted my first analysis alongside such motivating and helpful colleagues.

I need to also acknowledge the many mentors and peers that helped me arrive at Penn in the first place. Thanks to Gordon Rule and Farah Abdelaziz for their help on my first ever research project, PCR mutagenesis and protein synthesis. Thanks to Minki Jeong and Henrik Rønnow for their mentorship during my internship at EPFL, and to Ashutosh Kotwal and Sourav Sen for their guidance during my internship with Duke. And finally thanks to Reinhard Schumacher and Kunal Ghosh for their assistance during my application process.

Of course, this journey could not have been possible without the caring support of my friends and family. My sincere gratitude goes to the many friends I have met over the course of my PhD, including Luis Felipe Gutierrez Zagazeta, Tanay Bhandarkar, Ethan Torres, Yvonne Zagzag, and Charlotte Slaughter. From

crafting paper plants to cocktail nights, I truly appreciate the time we have spent together, and it has made my time in Philly all the more memorable. I also thank again the members of the Penn ATLAS group for our eccentric lunches and mutually assured procrastination. I am forever indebted to my parents Ningli and Jiyang and my sister Tanya for their continuous support throughout my life. I want to especially thank my best friend Balto for keeping me company through the Covid lockdown. Finally, my deepest gratitude goes to Abbey Li, for her unwavering encouragement and compassion, without which this journey would not have been the same.

ABSTRACT

A MEASUREMENT OF BOOSTED DIBOSONS WITH GAUSSIAN PROCESS BACKGROUND MODELING AT THE ATLAS DETECTOR

Riley Xu

Elliot Lipeles

This dissertation presents a measurement of WW , WZ , and ZZ production in the semileptonic final state at the ATLAS detector. The measurement was performed using 139 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ proton-proton collisions collected at the Large Hadron Collider between 2015 and 2018. An innovative data-driven approach to model the V+jets background was investigated, using a Bayesian method called Gaussian process regression. The method, more commonly found in other disciplines and yet to see much adoption in high-energy physics, was established to be much more robust and precise than traditional approaches using functional fits. As the analysis is still blinded as of this publication, expected results are presented here instead. Differential fiducial cross sections are obtained in several variables, and a total sensitivity of around 10σ to the Standard Model process is expected. Expected limits on the Effective Field Theory coefficient c_W are determined at approximately $|c_W/\Lambda^2| < 0.42 \text{ TeV}^{-2}$ at the linear order and $|c_W/\Lambda^2| < 0.03 \text{ TeV}^{-2}$ at the quadratic order. In addition to the diboson measurement, firmware, software, and algorithm developments on tracking in the trigger for the HL-LHC upgrade are presented. The highly increased pileup of the HL-LHC environment requires a much smarter trigger system to reject background events, and tracking enables more informed decisions. Two algorithms, a pattern matching method and a Hough transform approach, were implemented in a simulation framework and optimized for the HL-LHC environment.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	v
Table of Contents	vi
List of Tables	ix
List of Illustrations	x
Preface	xiii
 Part I: Introduction	 1
1 Introduction	1
2 Theoretical Framework	3
2.1 The Standard Model	3
2.1.1 Particle Content	3
2.1.2 Forces and Interactions	5
2.1.3 Beyond the Standard Model	6
2.2 Lagrangians	6
2.2.1 Basics	6
2.2.2 Symmetries	8
2.2.3 The Standard Model Lagrangian	9
2.3 Effective Field Theory	10
2.3.1 SMEFT Formulation	12
2.3.2 Linear and Quadratic Terms	13
3 Experimental Setup	15
3.1 The Large Hadron Collider	15
3.1.1 HL-LHC	17
3.1.2 Run Nomenclature	18
3.2 The ATLAS Detector	19
3.2.1 Coordinate System	20
3.2.2 Inner Detector	22
3.2.3 Calorimeters	24
3.2.4 Muon Spectrometer	27
3.2.5 Magnet System	29
3.2.6 Trigger and Data Acquisition	30
3.2.7 Object Reconstruction	31
3.2.8 Simulation	35

Part II: Tracking in the Trigger	36
4 Hardware-based Tracking for Trigger	36
4.1 ATLAS Phase-II Upgrades	36
4.1.1 TDAQ	36
4.1.2 Inner Tracker	39
4.2 Functional Description	39
4.2.1 Pattern Matching	39
4.2.2 Track Fitting and Second Stage Extension	41
4.2.3 Hardware Organization	41
4.3 TP Firmware Developments	43
4.3.1 SoC	43
4.3.2 IPMC	45
4.4 Simulation and Pattern Matching Optimizations	47
4.4.1 Framework Architecture	47
4.4.2 Pattern Matching Optimizations	52
5 Hough Transforms	55
5.1 Hough Transform Algorithm	55
5.1.1 1D Hough	57
5.1.2 Slicing	58
5.2 Optimization Studies	59
5.2.1 Binning and Smearing	59
5.2.2 Magnetic Field Correction	60
Part III: Diboson Production in the Semileptonic Final State	65
6 Analysis Overview	65
6.1 Motivations	65
6.2 Object Selection	67
6.3 Event Selection and Classification	68
7 Fitting the V+jets Background	71
7.1 Functional Fits	71
7.1.1 Fits using the Three-variable Tagger	73
7.1.2 Fits using the ANN Tagger	77
7.2 Gaussian Process Regression	80
7.2.1 Bayesian Regression	80
7.2.2 Functions as an Infinite Vector Space	81
7.2.3 Definition of a Gaussian Process	82
7.2.4 Gaussian Process Regression	82
7.2.5 Selecting a Kernel	84
7.2.6 Numerical Approximation of the Marginal Posterior	85
7.3 Toy Examples for Gaussian Process	86
7.3.1 The Probability Density Function	87
7.3.2 The Posterior	92
7.3.3 Marginal Likelihood	93
7.4 Studies on Using Gaussian Process	96
7.4.1 Kernels, Binning, and ANN Warping	97
7.4.2 Marginal Likelihood Scans	97
7.4.3 Two-dimensional Fits	101

7.4.4	Other Channels and Subtleties	103
8	Unfolding and EFT	104
8.1	Profile Likelihood Unfolding	104
8.2	Fiducial Region Studies	105
8.2.1	Particle versus Parton	106
8.2.2	Benchmark Definition	106
8.2.3	Efficiency Discrepancy	107
8.2.4	Optimized Binning	112
8.3	Interference Resurrection	112
9	Fit Procedure and Results	115
9.1	Fit Procedure	115
9.1.1	Integrating the Gaussian Process	116
9.1.2	Likelihood Construction	117
9.2	Asimov Results	120
	Part IV: Conclusion	127
10	Conclusion	127
	Appendices	128
	Appendix A SMEFT Operators	128
	Appendix B Resonance Event Selections	130
	Appendix C V+jets Fits	132
C.1	Fits to Monte Carlo	132
C.1.1	Functional Fits	133
C.1.2	Gaussian Process	139
C.2	Fits to Data	146
C.2.1	0-lepton Channel	146
C.2.2	1-lepton Channel	148
C.2.3	2-lepton Channel	150
	Appendix D Glossary	152
	Bibliography	157

LIST OF TABLES

4.1	Model numbers of the SoC module tested for the TP board	44
5.1	Summary of Hough transform performance with various configurations	61
5.2	Effect of the magnetic field correction on efficiency and dataflow	62
6.1	Recent ATLAS and CMS measurements in diboson channels	67
6.2	W and Z branching ratios for each decay mode	67
6.3	Primary selections for the electron and muon definitions	68
6.4	Primary jet selections	68
6.5	Event selection cuts in the unfolding signal region channels	69
7.1	η scores of four test functions using the RBF kernel	91
8.1	Fiducial region event selection cuts	107
8.2	Energy scaling in different helicity modes for $q_L \bar{q}_R \rightarrow W^+ W^-$ in the SM and SMEFT	113
9.1	Measured SM and EFT signal strengths from fits on Asimov data	121
A.1	Bosonic dimension-6 SMEFT operators in the Warsaw basis	128
A.2	Four fermion dimension-6 SMEFT operators in the Warsaw basis	129
B.1	Cuts in the 0-lepton channel of the resonance side of the VV+VH Semileptonic analysis	130
B.2	Cuts in the 1-lepton channel of the resonance side of the VV+VH Semileptonic analysis	131
B.3	Cuts in the 2-lepton channel of the resonance side of the VV+VH Semileptonic analysis	131

LIST OF ILLUSTRATIONS

2.1	The particle content of the SM	4
2.2	Electron scattering Feynman diagrams	6
2.3	SM Feynman diagrams	7
2.4	Summary of ATLAS supersymmetry mass exclusions	11
2.5	Schematic comparison between resonance searches and EFT analyses	12
3.1	Schematic diagram of the CERN accelerator complex	16
3.2	Pileup at ATLAS in Run-2 and Run-3	17
3.3	Total luminosity delivered to ATLAS between 2011 and 2023	18
3.4	Proposed HL-LHC timeline	19
3.5	General cut-away view of the ATLAS detector	20
3.6	Illustration of the path of different particles through the ATLAS detector	21
3.7	Illustration of the ATLAS inner detector	23
3.8	Layout of the ATLAS inner detector	23
3.9	Cutaway illustration of the ATLAS calorimeters	25
3.10	Layout of the ATLAS calorimeters	25
3.11	Layout of the ATLAS forward calorimeters	27
3.12	Cutaway illustration of the ATLAS muon spectrometer	28
3.13	Layout of the ATLAS muon spectrometer	28
3.14	Illustration of the ATLAS magnet system	30
3.15	Magnetic field map of the ATLAS solenoid	30
3.16	Diagram of the ATLAS TDAQ system	31
3.17	Flowchart of the ATLAS tracking algorithm	33
4.1	Diagram of the Phase-II TDAQ baseline system	37
4.2	Diagram of the Event Filter system	38
4.3	Potential layout of the ITk	40
4.4	Images of an HTT demonstrator board	42
4.5	Diagram of an HTT unit	43
4.6	Flowchart of the TP board components	44
4.7	Waveforms of a loopback test of the TP SoC's gigabit transceivers	46
4.8	Block diagram of the HTTSim architecture	48
4.9	Flowchart of the HTTSim algorithms	49
4.10	HTT pattern matching performance with various DC bit configurations	53
4.11	HTT pattern matching performance with various wildcards and matching criteria	54
5.1	Diagram of an idealized track in the Hough transform	56
5.2	Schematic example of using a Hough transform to find tracks	57
5.3	Schematic example of the 1D-Hough transform	58
5.4	Example of key-layer z -slicing for the Hough transform	59
5.5	Differences between the actual and expected ϕ_h in two regions	63
5.6	Empirical fit to $\Delta\phi$ as a function of r_h	64
5.7	Violin plots of $\Delta\phi_h$ before and after the magnetic field correction	64

6.1	Feynman diagram of the decay signature for the VV+VH Semileptonic analysis	65
6.2	A subset of SMEFT operators organized by the processes that they primarily affect	66
6.3	Nominal MC yields in the eight regions used in unfolding	70
7.1	MC simulations of W+jets and diboson events in the 1-lepton HP region	72
7.2	Three-variable tagger cuts	73
7.3	$m(J)$ response curves of the three-variable and ANN taggers	74
7.4	MC fits in the three-variable tagger HP region in two bins of $p_T(J)$	75
7.5	Empirical derivation of a two-dimensional functional form for fitting W+jets	76
7.6	2D functional fit to W+jets MC	76
7.7	A scan of the lower and upper $m(J)$ cuts for the ANN tagger	78
7.8	MC fits in the ANN HP region in bins of $p_T(J)$, using third order Chebyshev polynomials	79
7.9	MC fits in the ANN HP region in bins of $m(VV)$, with multiple functions	79
7.10	Samples drawn from a Gaussian process using a RBF kernel	88
7.11	Test functions for evaluating a Gaussian process' PDF	90
7.12	Posterior of a toy Gaussian process using an RBF kernel	92
7.13	Effect of hyperparameters on the posterior of a toy Gaussian process	94
7.14	Posteriors of Gaussian processes with isotropic kernels	95
7.15	Marginal likelihood schematic example and contours	96
7.16	Gaussian process fit to W+jets MC with $p_T(J) \in (370, 410)$ GeV	98
7.17	Yields of Gaussian process fits with different kernels and binning, in bins of $p_T(J)$	99
7.18	Yields of Gaussian process fits with different $p_T(J)$ cuts, in bins of $m(VV)$	99
7.19	Gaussian process marginal likelihood contours and marginal posterior yield distributions	100
7.20	Comparison of different estimates of the Gaussian process marginal posterior	101
7.21	Two-dimensional Gaussian process fit against $m(VV)$ and $m(J)$	102
7.22	Summary of Gaussian process fits to V+jets MC in the 0- and 2-lepton channels	103
8.1	Fiducial and detector-level kinematic distributions of the VV system	108
8.2	Efficiency and fiducial accuracy in the 1-lepton channel for diboson and c_W	109
8.3	Cut-flow plot of the efficiency for diboson and c_W binned in truth $p_T(VV)$	109
8.4	The $m(VV)$ distribution in bins of $p_T(VV)$ for the SM and c_W processes	110
8.5	Efficiency and fiducial accuracy as functions of $p_T(VV)/m(VV)$	111
8.6	ΔR between the two bosons in bins of $p_T(VV)$ for diboson and c_W	111
8.7	Migration matrices with optimized binning in all three lepton channels	113
8.8	Distributions of $ \varphi_1 $ in two bins of $p_T(W)$ with $c_W = 1$	114
9.1	MC distributions of $m(J)$ in the signal region and MCR	116
9.2	Gaussian process regression signal contamination tests in all three lepton channels	118
9.3	Pre- and post-fit distributions of the SM diboson fit in all lepton channels	122
9.4	PLU post-fit distributions in all lepton channels	123
9.5	Measured fiducial distributions in all lepton channels from Asimov tests	124
9.6	Additional fit plots for $m_T(VV)$ in the 1-lepton channel	125
9.7	Validation test of the PLU uncertainties in the 1-lepton channel	126
C.1	MC fits in the three-variable tagger HP region in bins of $p_T(J)$	133
C.2	A scan of the lower and upper $m(J)$ cuts for the ANN tagger in bins of $p_T(J)$	134
C.3	The highest significance mass window as a function of $p_T(J)$	135
C.4	A scan of the lower and upper $m(J)$ cuts for the ANN tagger at the 80% working point	135
C.5	MC fits in the ANN HP region in bins of $p_T(J)$, using a single exponential function	136
C.6	MC fits in the ANN HP region in bins of $p_T(J)$, using two exponential functions	137
C.7	MC fits in the ANN HP region in bins of $p_T(J)$, using third order Chebyshev polynomials	138
C.8	Gaussian process fits to W+jets MC using the RBF kernel and 1 GeV bins	139

C.9	Gaussian process fits to W +jets MC using the RBF kernel and 3–5 GeV bins	140
C.10	Gaussian process fits to W +jets MC using the Matern 5/2 kernel and 1 GeV bins	141
C.11	Gaussian process fits to W +jets MC using the Matern 5/2 kernel and 3–5 GeV bins	142
C.12	Two-dimensional Gaussian process regression fits against $m(VV)$ projected onto $m(J)$	143
C.13	Gaussian process fits to V +jets MC in the 0-lepton channel	144
C.14	Gaussian process fits to V +jets MC in the 2-lepton channel	145
C.15	Gaussian process fits to data in the 0-lepton channel, binned in $m_T(VV)$	146
C.16	Gaussian process fits to data in the 0-lepton channel, binned in $p_T(J)$	147
C.17	Gaussian process fits to data in the 1-lepton channel, binned in $m(VV)$	148
C.18	Gaussian process fits to data in the 1-lepton channel, binned in $p_T(J)$	149
C.19	Gaussian process fits to data in the 2-lepton channel, binned in $m(VV)$	150
C.20	Gaussian process fits to data in the 2-lepton channel, binned in $p_T(J)$	151

PREFACE

This dissertation collects a record of my work as a graduate student in the ATLAS group at the University of Pennsylvania. I joined the group in the summer of 2018 straight after having finished my undergraduate studies at Carnegie Mellon University. At this point the Large Hadron Collider (LHC) was wrapping up its second round of data taking and poised to enter an upgrade and commissioning phase, aptly named Long Shutdown 2 (LS2). The Penn group however was looking forward to LS3, then scheduled to take place starting 2024. During this phase many new components in ATLAS were to be installed, including the Hardware-based Tracking for Trigger (HTT), which as the name suggests, uses custom hardware to provide tracking information to the trigger system.

My first projects revolved around various parts of the HTT system. My first summer I started by optimizing algorithms used in HTT by employing the simulation framework of Fast TracKer (FTK), a similar system that was planned to be installed in LS2 but was ultimately canceled. This proved to be a daunting task as the framework consisted of dozens of files, each containing thousands of lines of sparsely-commented C++. Thankfully I had the guidance of senior graduate student Christian Herwig who helped me get the spherical cow rolling. Once the fall semester started, I shifted my focus more on my coursework and lab teaching assistant position, but resumed research in the summer of 2019.

At this point, I began my qualification task, an ATLAS procedural step that would allow me to join the author list. My project was to write firmware for monitoring and control on HTT printed circuit boards. I had never worked with firmware before, so I spent a good chunk of time familiarizing myself with the ecosystem, especially Verilog and Xilinx's Vivado program. As a learning project, I implemented an unfulfilled aspiration of former student Elodie Resseguie: a one-dimensional version of the game Pong, played using the LEDs and buttons on a Xilinx test board. A strip of LEDs represented the table, and a single lit LED represented the location of the ball. Pressing a button when the "ball" was at the end of table would bounce it back; failing to do so would count as a goal for the opponent and the LEDs would helpfully blink the updated score. Although not particularly practical for physics, this project helped me build an increased confidence in my firmware abilities, and I was able to efficiently complete my qualification task.

In parallel with my firmware work, I began to be deeply involved with the HTT simulation project. We could not of course rely on the FTK simulation forever, so we needed to architect our own simulation framework, HTTSim. I became one of the main developers of the new framework, taking sole responsibility over several segments of the project. Learning from my previous experience, I placed special emphasis on

properly documenting every file, class, and function such that a future graduate student could easily jump into the project without feeling overwhelmed. The team was also very careful in keeping the new framework as modular as possible; components could be modified and swapped out without needing to alter downstream or upstream code. Once the simulation was in a running state, I used it to help optimize the algorithms again, continuing my work from my first summer.

But then around March 2020, the COVID-19 pandemic reached the United States and everything quickly grinded to a halt. The world entered a global lockdown, with almost all non-essential establishments closing down. Penn suspended all in-person activities, classes were converted abruptly to remote learning, and students were asked to vacate. I ended up moving back home, finishing my coursework in the quiet solitude of quarantine. The pandemic also dramatically affected the LHC schedule; many projects were falling behind, causing LS2 to ultimately be extended over a year. Notably, one LS3 upgrade project, a new pixel detector, was no longer able to support a planned readout mode necessary for the HTT to function. This, among other reasons, led to the cancellation of the HTT project in October 2020.

This was certainly a disappointing outcome, having spent much of my first years working on HTT firmware and software. However, a replacement to HTT was needed, as the trigger system still needed tracking information. ATLAS therefore launched a few task forces to investigate potential options. I joined one of these teams, testing an algorithm called the Hough transform. One of my first major tasks was to retool HTTSim to simulate the new algorithm instead of the custom HTT hardware. Fortunately, thanks to our prior efforts to make HTTSim modular, the components unique to the HTT hardware were more or less isolated into a single module with a well-defined interface. So I implemented the Hough transform as a separate module following the same interface, which then enabled the two modules to be swapped seamlessly. This allowed me to complete the transition within a few weeks. I spent the next few months working with the team on optimizing the Hough transform algorithm, culminating in a task force report submitted June 2021.

After the task force wrapped up, I joined an analysis group called VV+VH Semileptonic, which made up the second half of my PhD. The analysis seeks to investigate diboson production (hence the VV, for two vector bosons, and VH, for a vector boson and a Higgs boson) in the semileptonic decay channel, where one of the bosons decays to leptons and the other to hadrons. The analysis is a bit of an unusual conglomerate; it combines a search, which looks for specific new particles, with a measurement, which looks to accurately measure a known quantity. Traditionally in ATLAS these would form two separate analyses, but because the two sides share common setup, including basic event selection and input files, the analysis was merged. I joined initially as the only student working on the measurement side, but other students got involved later.

My work in the analysis can be roughly broken down into two major tasks, which I worked on in parallel. First, we wanted to do the measurement under a theoretical framework called effective field theory (EFT). Roughly speaking, the framework enables our measurement of diboson production to be back-interpreted by

theorists. Whenever they come up with a new theory, they can use our EFT measurement to determine if that theory is actually feasible. So, while the search side of the analysis tests specific existing theories, the EFT side provides a broad and general constraint for theories past and future. For this analysis, I helped implement the procedures needed, validate their performance, and optimize our precision. I also helped mentor an undergraduate student in exploring the use of machine learning to further increase our sensitivity.

My other task, and more novel contribution, was to develop a data-driven method to estimate our largest background, V +jets. Typically analyses use computer simulations to estimate the contribution from background processes, but for certain processes like V +jets, the simulation is rather inaccurate. In these cases we can use a data-driven method to improve the accuracy of our measurement. Such methods have been used before to model V +jets, generally relying on functional fits. However this method has certain problems; most notably it has unknown systematic errors, which have to be estimated using ad-hoc procedures. I therefore investigated an innovative procedure called Gaussian process regression which offers many advantages over functional fits. As Gaussian process regression has not seen much exposure in the high energy physics community, I spent significant effort testing and validating its performance in our use case. It was very rewarding to single-handedly create a full workflow from start to finish and get a new technique up and running.

By around March 2024, the measurement side of the analysis was more or less complete. I had also landed a job offer, so I began writing this dissertation and drawing my PhD to a close. Unfortunately, at this point the analysis is somewhat stuck in bureaucratic limbo. ATLAS requires several administrative steps to review analyses. These steps tend to take a tremendous amount of time, as they involve bringing in many reviewers external to the analysis. One of the most important steps that we are waiting on is called unblinding approval. Up until this approval, the analysis is blinded, which means the analyzers are not allowed to look at results using real data. This helps prevent bias from being introduced by the analyzers, but unfortunately means that I was not able to present real results in this dissertation. Instead, expected results using simulated data are presented instead. These demonstrate the remarkable precision of our analysis, but I eagerly await to see our unblinded results.

The past six years have been an incredibly enriching experience, being able to work alongside many of the smartest people in the world at one of the greatest experiments ever built. Despite being just one small part in a project that is older than I am, I will be eternally proud to have contributed to this landmark experiment. The memories and lessons of this journey are things I will forever carry with me as I look to the next stage in my life.

Riley Xu

Philadelphia, May 2024

CHAPTER 1

Introduction

With the discovery of the Higgs boson in 2012, the Standard Model (SM) of particle physics was nominally completed. However, the SM still leaves many questions unanswered, such as dark matter and gravity. The past decade has therefore seen great effort looking for physics beyond the Standard Model (BSM). Much progress has been made from experiments at the LHC, carving out exclusion limits that highly constrain once favored models such as supersymmetry. But with a decade of null results, it is prudent to focus more heavily on alternative strategies.

Most LHC searches have a large caveat: they rely on BSM particles being light enough to be produced by the LHC. After all, how does one measure a particle that cannot be produced? But thanks to the nature of quantum mechanics, heavy particles can still enter Feynman diagrams and subtly affect LHC processes. These manifest as minute shape deviations from the SM prediction. Experimentally, this means one can infer the presence (or absence) of new heavy particles by doing precision tests of the SM.

Such measurements are usually organized under a framework called the Standard Model Effective Field Theory (SMEFT). The framework quantifies the shape deviations from heavy BSM physics by using a set of effective operators, and provides a coherent method to compare and combine multiple measurements from different channels. One of its major appeals is that it is very model-independent. SMEFT's only major assumption is that BSM physics is out of reach of the LHC; it does not rely on assuming any new particles nor how they decay. Precision tests of the SM can thus be reframed as precision measurements of SMEFT operators.

The effects of BSM contributions tend to grow with energy, so the best way to measure SMEFT operators is by probing rare, high-energy processes at the LHC. This thesis presents in part III one such analysis at the ATLAS detector: the VV+VH Semileptonic analysis. This analysis targets highly boosted diboson (WW , WZ , and ZZ) and VH (WH and ZH) production in the semileptonic decay channel, where one of the bosons decays to leptons and the other to quarks. It is one of the few SMEFT measurements at ATLAS to use hadronic vector boson decays, as most measurements focus on fully-leptonic channels. The semileptonic channel benefits from the higher branching ratio of hadronic decays, which greatly increases the data rate and consequently the energy reach. However, it suffers from a much larger background contamination, which must be accurately modeled. An overview of the analysis is provided in chapter 6, and the background

modeling is discussed in chapter 7, with an emphasis on the use of a novel method called Gaussian process. Details of the precision measurement procedure are described in chapter 8, and finally expected results from the analysis are shown in chapter 9.

A key limitation of the VV+VH Semileptonic analysis, and many other analyses, is the amount of data available. Most analyses could be dramatically improved by collecting more data, but this is costly and time consuming. Therefore the LHC collaboration is planning an ambitious upgrade project, the High-Luminosity LHC (HL-LHC), which will drastically increase the rate of collisions, making data taking much more efficient. However, many of the detectors on the ring, including ATLAS, are not ready to deal with this increased rate. Many components in ATLAS will need to be upgraded. Part II of this dissertation describes one such upgrade project, tracking in the ATLAS trigger system. Chapter 4 outlines a baseline implementation using custom hardware, and details firmware developments, software developments, and algorithm optimizations. Unfortunately, the custom hardware solution was ultimately canceled due to upstream requirement changes. An implementation of trigger tracking was still sought however, and chapter 5 describes testing an alternative method using Hough transforms.

This introductory part continues with an overview of the SM and SMEFT in chapter 2, and a description of the LHC and ATLAS in chapter 3.

CHAPTER 2

Theoretical Framework

This chapter gives a brief overview of the Standard Model (SM) of particle physics and one of its extensions, the Standard Model Effective Field Theory (SMEFT). The SM forms the theoretical backdrop for the majority of the experiments at the LHC, and SMEFT in particular is the motivation behind the analysis in part III. The descriptions given here will focus on only what is needed to understand the context of parts II and III. For a more complete description, see references [1–3].

2.1 The Standard Model

The SM is the leading theory that describes the elementary particles that make up the universe and how they interact. It is a quantum field theory (QFT), a framework that largely arose to combine the two pillars of early 20th century physics: quantum mechanics, which describes the discrete nature of subatomic phenomena, and special relativity, which governs spacetime symmetries and mass-energy equivalence. The mathematical formalism behind QFT is incredibly complex, with a plethora of important subtleties that tend to obfuscate the underlying physics. Fortunately the SM can be understood heuristically by focusing on the properties of each particle, and understanding how they interact through the use of Feynman diagrams.

2.1.1 Particle Content

A summary of the particles in the SM is shown in figure 2.1. These can be organized into two main classes, fermions and bosons, corresponding to the left and right halves of the figure respectively.

Fermions are the particles that make up matter. For example, the proton and neutron consist of up quarks and down quarks, meanwhile atoms are composed of protons, neutrons, and electrons. Fermions can be divided into three generations, corresponding to each of the left columns in the figure. The three generations have essentially the same properties across the rows, except the latter generations are heavier. Therefore the majority of the universe consists of the lightest, first generation. All fermions also have antiparticles, not shown in the figure, which have the exact same properties but opposite charge.

The fermions can also be subdivided into quarks and leptons. Quarks participate in the strong force, and one of the consequences is that they rapidly group up into objects called hadrons. The proton and neutron

Standard Model of Elementary Particles

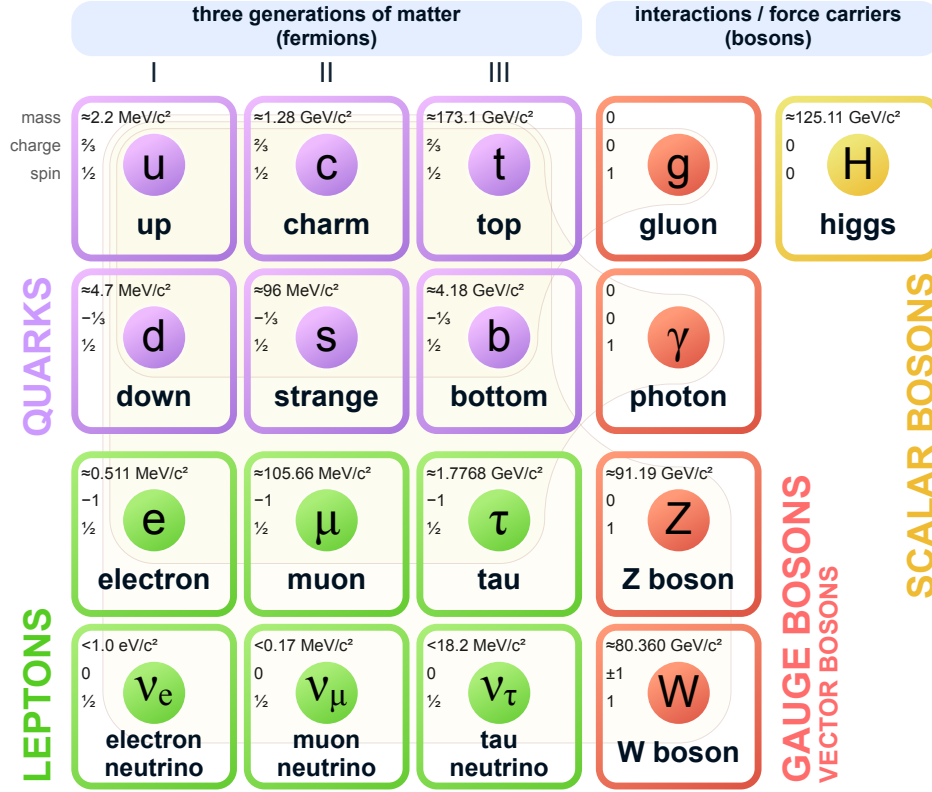


Figure 2.1: The particle content of the SM, from [4].

are examples of hadrons, consisting of three quarks each. Meanwhile leptons are not affected by the strong force. In particle collider experiments, this means that leptons generally leave a very clean signature in the detector, while quarks leave a messy one. This is discussed further in chapter 3.

Bosons control the interactions between fermions. The gauge bosons are the force carriers. Most familiarly, the photon is a carrier for the electromagnetic force. That is why light, which is an electromagnetic wave, can be interpreted as a stream of photons. However, the photon also appears in other instances of the electromagnetic force. For example, when two electrons bounce off each other because like charges repel, they are actually exchanging a photon. Similarly the gluon is a carrier for the strong force, and the W and Z bosons are carriers for the weak force. These particles are often called vector bosons because in the mathematical formalism, they are represented as three-dimensional vectors. Note that there are actually eight types of gluons, though experiments generally cannot distinguish them. Meanwhile there are two W bosons, W^+ and W^- , which are antiparticles of each other.

The Higgs boson is unique. It is responsible for generating the masses of all massive elementary particles. Therefore all particles other than the gluons and neutrinos interact directly with it. The Higgs boson has been the focus of much study since it was the most recently confirmed particle in the Standard Model, having been discovered in 2012 [5]. However it will not be a major subject of this dissertation, which focuses instead on the W and Z bosons.

2.1.2 Forces and Interactions

The SM describes the effects of three of the four fundamental forces: electromagnetic, weak, and strong. The strong force is also often referred to as quantum chromodynamics (QCD). Notably the SM omits gravity, which among other reasons is why it is considered incomplete. The easiest way to understand the effects of the three forces is by examining Feynman diagrams. These are pictorial representations of particle interactions, first introduced by Richard Feynman [6, 7].

An example diagram is shown in figure 2.2a, which represents electron-electron scattering. The diagram should be interpreted with time progressing from left to right. Two electrons, represented by the left solid lines, move towards each other and exchange a photon, represented by the squiggly line. Then the two electrons both scatter away, represented by two right solid lines. Another process, electron-positron annihilation, is shown in figure 2.2b. Here an electron and positron, represented by the left two solid lines, meet and annihilate, producing a photon. Then the photon converts back into an electron positron pair, represented by the right two solid lines. The arrows in this diagram can be misleading; they represent the particle flow, so normal particles have right-facing arrows and antiparticles have left-facing arrows. They do not represent the direction of motion, which flows from left to right for all particles.

Notice that both of these diagrams are constructed out of the same ingredient: a three-point vertex where two solid lines and one squiggly line meet, as in figure 2.3a. In this example the vertex represents the action of the electromagnetic force, which couples charged fermions with the photon. These diagrams also illuminate why the photon is called the force carrier for the electromagnetic force. In general, any SM process can be constructed in similar manner by combining interaction vertices. A list of all SM interaction vertices is shown in figure 2.3. Note that all of these diagrams can be rotated and mirrored, as long as charge and particle/antiparticle number is conserved.

Every Feynman diagram corresponds to a mathematical term that can be used to calculate the probability, or cross section, that the process occurs. The details of such calculations can be found in a number of QFT textbooks, such as references [1, 2]. In practice, such calculations are extremely difficult to do analytically, and physicists often rely on Monte Carlo (MC) simulations instead. These simulations generate events of a given process through repeated random sampling to arrive at a numerical estimate of cross sections. Commonly used frameworks include MadGraph [8] and Sherpa [9].

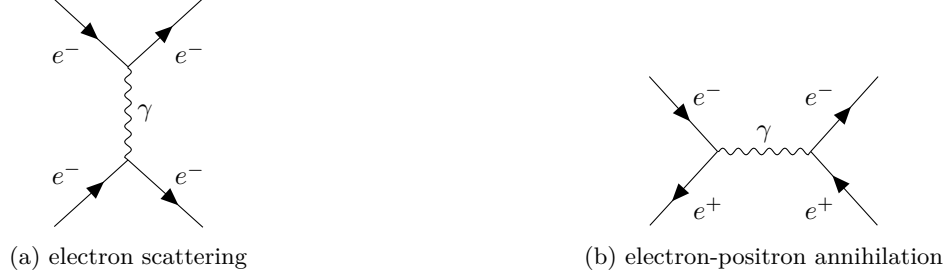


Figure 2.2: Example Feynman diagrams of two SM processes. See text for how to read these diagrams.

2.1.3 Beyond the Standard Model

The SM is an incredibly well-tested theory, making accurate predictions for thousands of measurements [10]. However, it is incomplete. It does not explain a variety of phenomena such as gravity, neutrino masses or dark matter. A few discrepancies between the SM and experiments also exist, such as a measurement of the W boson's mass by the CDF Collaboration [11] or the measurement of the muon magnetic moment at Fermilab [12]. Thus, the current drive of high-energy physics is to search for physics beyond the Standard Model (BSM).

In this context, the SM is still incredibly important as it serves as the benchmark against which almost all high-energy physics experiments are tested. More formally speaking, the SM is the null hypothesis in these experiments; if the result of the experiment agrees with the SM prediction, one declares that no new physics has been discovered. Meanwhile deviations from the SM prediction are the coveted goldmines. These are signs of BSM physics, which help point the way for the next great development in particle physics.

2.2 Lagrangians

While this chapter focuses on explaining the physics heuristically without delving too much into the mathematical details, this section will briefly describe the Lagrangian, which will be necessary to understand the subsequent section.

2.2.1 Basics

A Lagrangian is often introduced as the kinetic energy K of a system minus the potential energy V of a system,

$$L = K - V. \quad (2.1)$$

More importantly is the Lagrangian density $\mathcal{L}$, related via

$$L = \int d^3\mathbf{x} \mathcal{L}(x, \varphi_i, \partial_\mu \varphi_i). \quad (2.2)$$

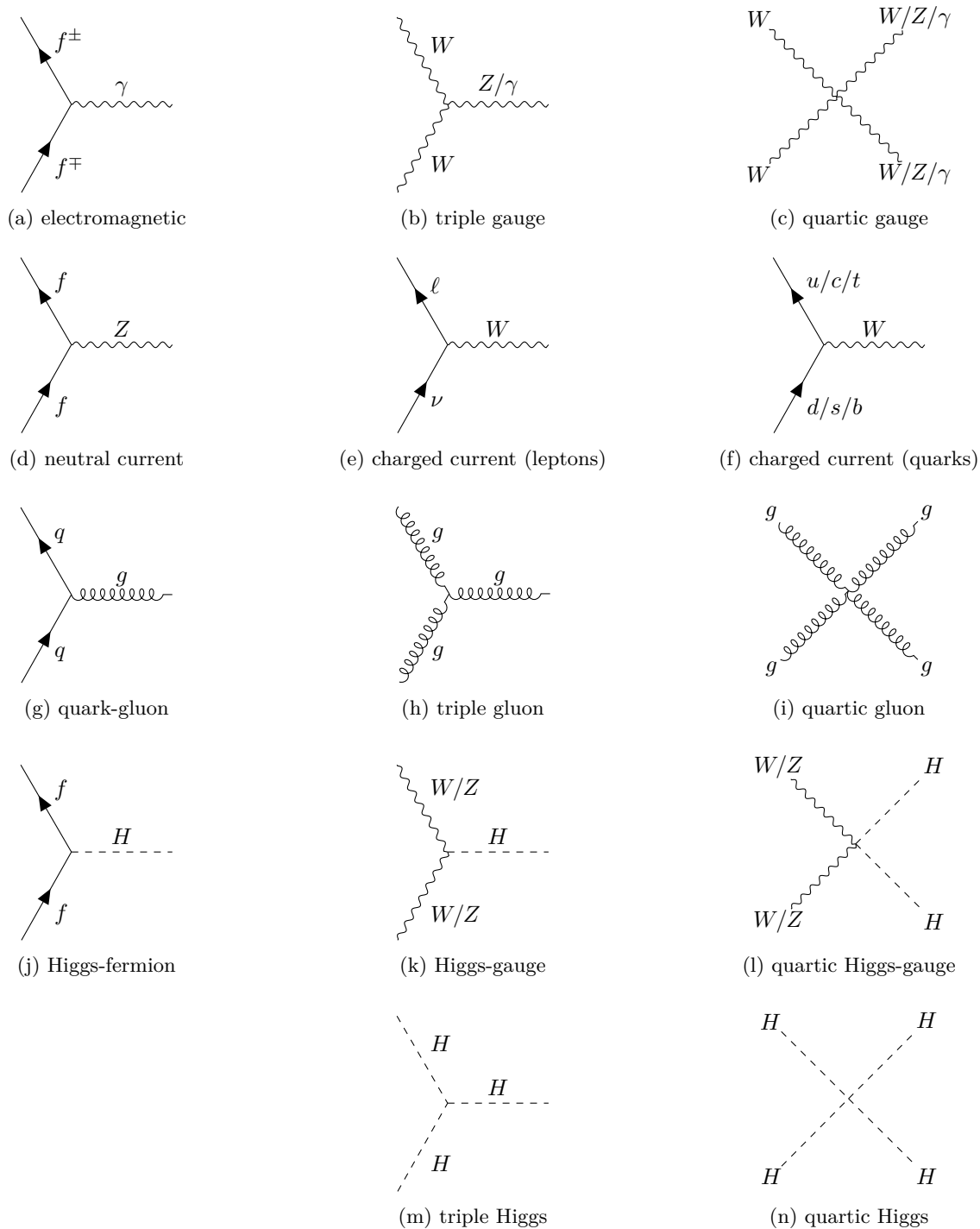


Figure 2.3: Feynman diagrams of all SM interactions.

Here $\mathcal{L}$ is written as a function of spacetime x and fields φ_i , along with first derivatives of the fields. The fields are also functions of spacetime x , and correspond to the particles in the system.¹ Note that the Lagrangian has units of energy, so the Lagrangian density has units of $[E] \cdot [x]^{-3} = [E]^4$ in natural units where $\hbar = c = 1$. Somewhat confusingly, the Lagrangian density is often referred to as simply the Lagrangian; this dissertation will also drop the “density” term from here forward.

Instead of focusing on equation (2.1) as the defining property of a Lagrangian, it is often helpful to think of a Lagrangian as a “master expression” from which properties of the system can be derived. In particular, the equations of motion of any field (i.e. particle) can be derived using the Euler-Lagrange equations,

$$\frac{\partial \mathcal{L}}{\partial \varphi_i} = \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \varphi_i)} \right). \quad (2.3)$$

From a certain perspective, any function of spacetime and fields can be interpreted as a Lagrangian, and the above equation can be applied to determine corresponding equations of motion. With this thought pattern, instead of starting with a system with known properties and trying to back-solve the Lagrangian from equation (2.3), one can instead write down an educated guess for the Lagrangian and hope equation (2.3) comes out with the correct answer.

In practice the equations of motion do not need to be used directly. Recall that the quantum mechanical probability, or the cross section σ , that a process occurs is given by

$$\sigma \propto |\mathcal{M}|^2 \equiv |\langle f | S | i \rangle|^2. \quad (2.4)$$

Here f represents the final state, i represents the initial state, and $\mathcal{M}$ is called a matrix element. S is the evolution operator which encodes all the details of the Lagrangian model. Thankfully, the path integral formalism, which is the mechanism behind Feynman diagrams, handles much of the details regarding S . Using Feynman diagrams, the matrix element can almost be directly read off each diagram. Again, the details can be found in QFT textbooks such as references [1, 2].

2.2.2 Symmetries

Note that the Lagrangian for a given system is not unique and does not necessarily have to correspond to kinetic and potential energies as in equation (2.1). Rather, any Lagrangians that produce the same equations of motion in equation (2.3) represent identical systems. For example, taking a Lagrangian and adding a constant term does not affect equation (2.3), so the two Lagrangians represent the same system.

Transformations that leave the equations of motion unchanged are deemed symmetries of the system. Often these symmetries reflect fundamental symmetries of the universe. For example, in particle physics a Lagrangian should be invariant under translations $x \rightarrow x + \delta x$, because the laws of physics in Philadelphia

¹Formally, a particle is a localized excitation of a field. For example a photon is an excitation of the electromagnetic fields, and an electron is an excitation of the electron field.

should be the same as those in Geneva. From the theory of special relativity, the same argument can be applied to rotations and boosts, which together are called the Lorentz group. Lagrangians in particle physics are thus required to be Lorentz invariant, meaning their equations of motions are unchanged under Lorentz group operations.

Other transformations can act on the fields themselves instead of the spacetime coordinates x . For example, consider a Lagrangian with a complex field ψ ,

$$\mathcal{L} = \partial^\mu \psi^\dagger \partial_\mu \psi - m^2 \psi^\dagger \psi. \quad (2.5)$$

If we transform $\psi \rightarrow e^{i\alpha} \psi$ where α is some real constant, notice that the Lagrangian is unchanged and hence the equations of motion are unchanged too. Since transformations of the form $e^{i\alpha}$ form a mathematical group called $U(1)$, we state that this Lagrangian has a global $U(1)$ symmetry. These symmetries do not reflect fundamental assumptions about relativity, and are not requirements per se, but have important consequences on the physics of the system.

In the above example, the term global refers to how the factor $e^{i\alpha}$ is applied uniformly, independent of x . In contrast, a local transformation would depend on x , for example $\psi \rightarrow e^{i\alpha(x)} \psi$ where $\alpha(x)$ is now a function of x . One can readily see that the Lagrangian in equation (2.5) is not invariant under this local transformation because the derivatives will grab onto the $\alpha(x)$ terms. However, the Lagrangian can be made invariant by introducing a vector field A^μ , referred to as a gauge field. If the gauge field transforms as $A_\mu \rightarrow A_\mu - \frac{1}{q} \partial_\mu \alpha(x)$ and the derivative is replaced by $D_\mu = \partial_\mu + iqA_\mu$, we note $D_\mu \psi \rightarrow D_\mu (e^{i\alpha(x)} \psi)$ under the $U(1)$ transformation. The Lagrangian

$$\mathcal{L} = (D^\mu \psi)^\dagger (D_\mu \psi) - m^2 \psi^\dagger \psi \quad (2.6)$$

now has a local $U(1)$ symmetry.

The above procedure may seem a bit artificial, but it turns out such gauge theories accurately describe particle physics. We can interpret this example as a particle and its antiparticle, represented by the fields ψ and $\psi^\dagger$ respectively, with mass m given by the quadratic term $m^2 \psi^\dagger \psi$. The fields are charged, in that they interact through their coupling with a force carrier, represented by the gauge field A_μ . This is highly reminiscent of electromagnetism, where ψ is like an electron and A_μ is like a photon, and indeed the Lagrangian in equation (2.6) is a crucial precursor to quantum electrodynamics.

2.2.3 The Standard Model Lagrangian

The SM is similarly written using a gauged Lagrangian. But instead of just one scalar field with a $U(1)$ local symmetry, the SM has 12 fermion fields and a $SU(3) \times SU(2) \times U(1)$ local symmetry corresponding to the four types of gauge bosons. The SM also has additional complications to properly include the fermion and gauge boson masses, which are solved via the Higgs field and boson. Writing down the complete SM

Lagrangian is not a particularly enlightening ordeal, especially in its fully expanded form after accounting for symmetry breaking caused the Higgs mechanism, but can found in references [13, 14]. Instead, equation (2.7) presents a highly summarized version as it appears on CERN t-shirts and coffee mugs [15]:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + (\psi_i y_{ij} \psi_j \phi + \text{h.c.}) + |D_\mu \phi|^2 - V(\phi). \quad (2.7)$$

This expression can be broken down by each term, which roughly correspond to the Feynman rules in figure 2.3:

- $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$: self-interactions between the gauge bosons.
- $i\bar{\psi}\not{D}\psi$: interactions between fermions and the gauge bosons.
- $\psi_i y_{ij} \psi_j \phi + \text{h.c.}$: Yukawa couplings between fermions and the Higgs boson. This term gives the fermions their masses. The *h.c.* stands for hermitian conjugate of the prior term.
- $|D_\mu \phi|^2$: couplings between the W/Z bosons and the Higgs boson, also how they gain their mass.
- $-V(\phi)$: Higgs self-potential and self-interactions.

2.3 Effective Field Theory

As mentioned in section 2.1.3, the SM is incomplete. Many theorists thus attempt to extend the SM to cover these shortcomings, such as introducing new particles that could act as dark matter. Such models are called BSM theories. These typically take a top-down approach: they postulate new fundamental particles with particular characteristics, leading to specific and concrete testable predictions. Experiments then target those predictions to verify if they are accurate, a so called *direct search*. Of course, such models still contain many free parameters, commonly including the mass scale of the new particles. This means that generally, the models can not be conclusively ruled out. Instead, null results from experiments produce constraints on the free parameters, such as a lower-bound constraint on the mass of the new particles.

For example, a very popular BSM theory is supersymmetry, which postulates a wide array of new particles that mirror the currently known ones [16]. ATLAS has conducted dozens of searches for supersymmetry over the past decade. Each search typically produces a lower limit on the mass of a supersymmetric particle. Results from these experiments are summarized in figure 2.4. Notably, many of the supersymmetric particles already have limits around 1 TeV, which is several factors larger than the heaviest SM particle, the top quark at 173 GeV. Similar exclusion limits also emerge from searches targeting a variety of other BSM models [17], which suggest an interesting conclusion: it seems highly probable that there exists a large separation in energy scales between SM physics and BSM physics.

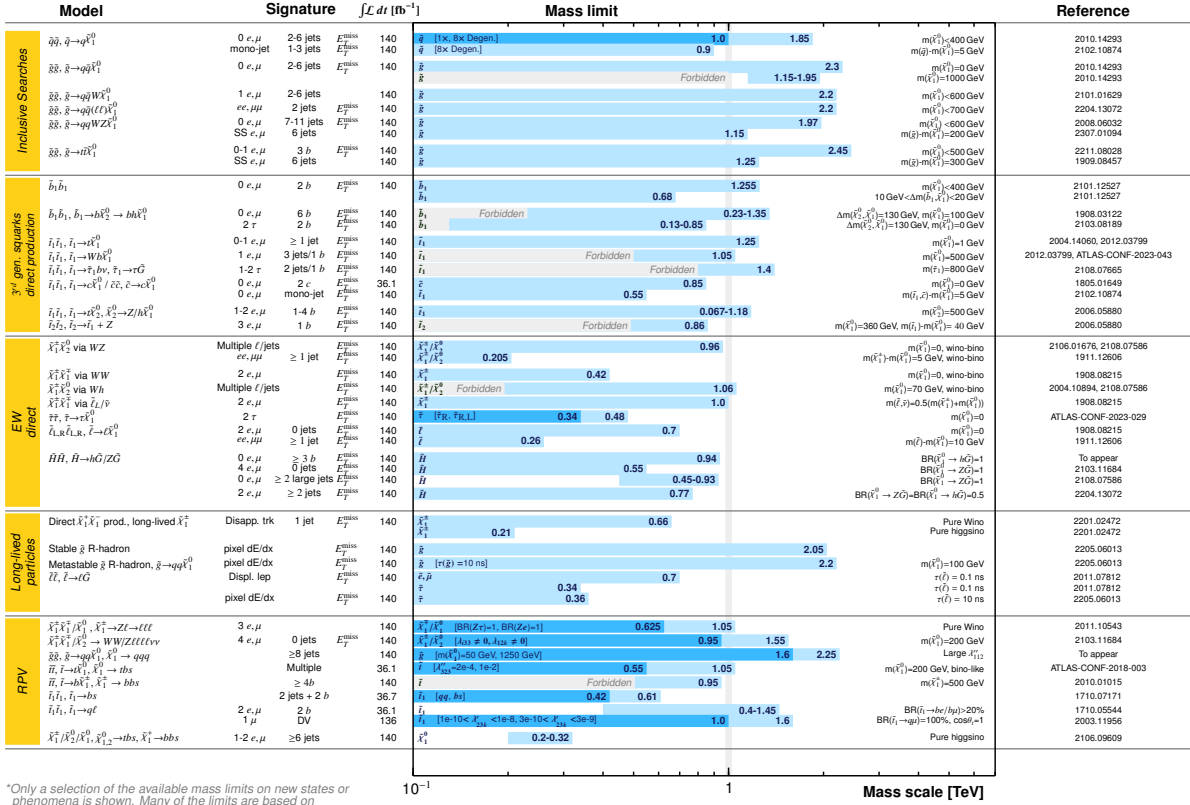


Figure 2.4: A summary of ATLAS exclusion limits on various supersymmetric particle masses, from [18].

In such cases, we can turn to a powerful method called effective field theory (EFT). EFTs can be roughly thought of as the field theory equivalent of a Taylor series expansion. In a Taylor series, an arbitrary² function f can be approximated using a series polynomial terms, where adding more terms increases the accuracy of the expansion. But as long as $|x| \ll 1$ the function can be accurately represented by only the first few terms (assuming an expansion around $x = 0$). Similarly, using EFT we can approximate arbitrary² BSM theories using a series of Lagrangian terms. As long as the energy E of relevant experiments is much less than the energy scale Λ of the BSM theory, the EFT will be accurate using only the first few terms of the series.

The advantage of this method is that we can probe unknown physics at the higher energy scale without worrying about specific BSM models. In contrast to the top-down methodology of direct searches, EFT measurements take a bottom-up approach: make generic measurements that are model-independent, then match those results up to specific BSM models to see if they agree. Whereas direct searches target a specific BSM model, EFT analyses are applicable to a wide range of models. EFT analyses can also be sensitive to

²Subject to certain constraints.

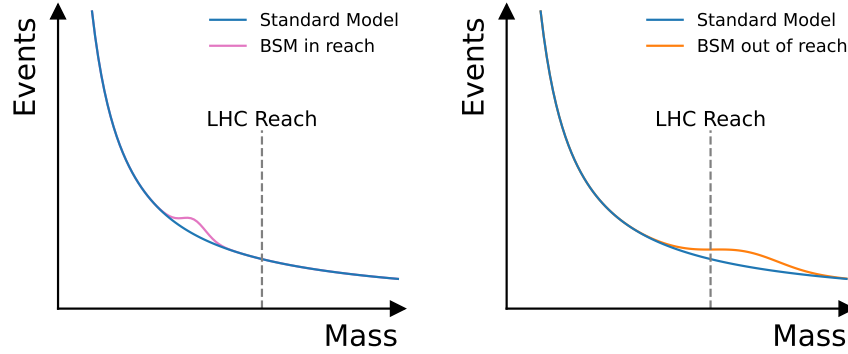


Figure 2.5: On the left, the BSM effect is within the experiment’s energy reach, and a resonance search would be optimal. On the right, the BSM effect is out of reach and a resonance search would likely fail. However the tail end of the resonance can still be measured by an EFT analysis.

physics just out of reach of direct searches. For example, direct searches often look for resonances, as shown in the left diagram of figure 2.5. But if the resonance falls above the reach of experiment, as in the right side of figure 2.5, a resonance search would likely be unable to measure it. But thanks to quantum width effects, the tail end of the resonance would still be measurable by an EFT analysis.

Such approximation methods that take advantage of separations of scales are quite common in physics. One example is the multipole expansion in electrostatics: the electric potential of a charge configuration on length scales a can be approximated at large distances $r \gg a$ by a series of spherical harmonic terms. Measuring the multipole terms gives hints about the charge distribution without needing specific knowledge of a . As another example, EFT was famously used by Fermi in his theory of beta decays [19], which accurately modeled the weak force without any knowledge of W or Z bosons, thanks to the large separation in scale between the energies of the experiments at that time and the weak scale.

2.3.1 SMEFT Formulation

Applying the EFT formalism to the SM produces the Standard Model Effective Field Theory (SMEFT). This extends the Lagrangian formalism of the SM by treating the SM as the lowest order term in an energy power series, and adding terms of increasing energy dimension:

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda} \sum_i c_i \mathcal{O}_i^{(5)} + \frac{1}{\Lambda^2} \sum_i c_i \mathcal{O}_i^{(6)} + \dots \quad (2.8)$$

Here $\mathcal{L}_{\text{SM}}$ refers to the Lagrangian in equation (2.7), $\mathcal{O}^{(n)}$ are operators of dimension n , c_i are the corresponding dimensionless Wilson coefficients, and Λ is the scale of BSM physics. Experiments measure the dimension-ful combination of c_i/Λ^{n-4} . Note that an explicit value for Λ is not needed, just as in the multipole expansion the value of a is not needed and in Fermi’s beta decay theory the masses of the weak bosons are not needed. Rather, the underlying assumption is that the general energy of experiments E is much

less than Λ . When solving for cross sections the operator terms will bring down factors of E , such that the power counting will look like

$$\sigma \propto 1 + E/\Lambda + (E/\Lambda)^2 + \dots, \quad (2.9)$$

which emphasizes that the SMEFT expansion is in powers of $E/\Lambda \ll 1$.

The operators are written using only the SM fields. Again, the power of EFT is that the effect of new particles is already included via the power series expansion, without needing to explicitly create a new field. In general, all operators that respect the symmetries of the SM, namely Lorentz invariance and the $SU(3) \times SU(2) \times U(1)$ gauge symmetry, are included. Because the operators in the SM Lagrangian have dimension $[E]^4$, the first term in the expansion consists of $[E]^5$ operators, which come along with a factor of $1/\Lambda$ to keep the units consistent. It turns out that there is only a single dimension-5 operator, which violates lepton number conservation. Indeed every odd dimension operator will violate lepton or baryon number (the number of quarks minus the number of antiquarks). These conservation laws are accidental features of the SM, and in principle do not have to be respected. However, dedicated experiments have much better sensitivity to lepton and baryon number violation, so most LHC searches will ignore the odd dimension operators.

Therefore, the first order terms of relevance are the dimension-6 operators, followed by the dimension-8 ones. Writing down a complete basis of these operators is a non-trivial task even at the dimension-6 order, as many operators which appear distinct at the Lagrangian level turn out to be redundant in their contribution to the equations of motion. It turns out there are 59 dimension-6 operators when assuming flavor symmetry and 2499 operators with full flavor structure [20]. These counts ignore baryon number violating terms. Meanwhile at the dimension-8 order, there are 993 and 44,807 operators in the flavor-agnostic and flavor-aware cases, respectively [21].

2.3.2 Linear and Quadratic Terms

The power counting described above concerned operators at the Lagrangian level, which propagates equivalently to matrix elements. But recall that cross sections, the physical quantities of interest, are proportional to the matrix element squared, as in equation (2.4). When applied to the SMEFT Lagrangian in equation (2.8), one finds

$$\sigma \propto |\mathcal{M}_{\text{SMEFT}}|^2 \quad (2.10)$$

$$\propto \left| \mathcal{M}_{\text{SM}} + \frac{1}{\Lambda^2} \mathcal{M}_{\text{EFT}}^{(6)} + \frac{1}{\Lambda^4} \mathcal{M}_{\text{EFT}}^{(8)} + \dots \right|^2 \quad (2.11)$$

$$\propto \underbrace{|\mathcal{M}_{\text{SM}}|^2}_{\text{SM}} + \underbrace{\frac{1}{\Lambda^2} \mathcal{M}_{\text{SM}}^* \mathcal{M}_{\text{EFT}}^{(6)}}_{\mathcal{O}(6) \text{ interference}} + \underbrace{\frac{1}{\Lambda^4} \left| \mathcal{M}_{\text{EFT}}^{(6)} \right|^2}_{\mathcal{O}(6) \text{ pure EFT}} + \underbrace{\frac{1}{\Lambda^4} \mathcal{M}_{\text{SM}}^* \mathcal{M}_{\text{EFT}}^{(8)}}_{\mathcal{O}(8) \text{ interference}} + \dots \quad (2.12)$$

Here $\mathcal{M}_{\text{SM}}$ corresponds to the matrix element derived using only terms from $\mathcal{L}_{\text{SM}}$, and $\mathcal{M}_{\text{EFT}}^{(n)}$ corresponds to matrix element terms derived using SMEFT operators of dimension n . Notably, at the cross section level, the leading order term in $1/\Lambda$ is actually an interference effect between the SM and dimension-6 terms. Meanwhile the pure dimension-6 contribution is a second-order term, the same order as the SM and dimension-8 interference effect.

This produces a slight problem for experiments as interference effects tend to be hard to measure. The diboson process in particular has highly suppressed interferences due to helicity selection differences between the SM and the dimension-6 terms [22]. Meanwhile, because the pure EFT contribution is the same order as the dimension-8 interference, formally they should be considered at the same time. But given the plethora of dimension-8 terms and the complexity they introduce, it is generally impractical to fully consider the dimension-8 terms. Thus a common practice is that experiments will report two measurements: one assuming only the dimension-6 interference and ignoring the higher order terms (the *linear* fit), and another using both the dimension-6 interference and pure EFT contribution, ignoring the dimension-8 terms (the *linear+quadratic* fit). ATLAS' most recent SMEFT combination paper uses this method, for example [23].

CHAPTER 3

Experimental Setup

This chapter introduces the experimental context for the work in this dissertation. Section 3.1 describes the Large Hadron Collider (LHC), which produces the high energy events that are needed to study fundamental physics. Then section 3.2 covers the ATLAS detector, which is used to actually observe the events produced by the LHC.

3.1 The Large Hadron Collider

The LHC [24] is the world's most powerful particle collider. Managed by CERN and completed in 2008, it lies in the tunnel formerly used by the Large Electron-Positron Collider (LEP) [25], a 27 kilometer ring between 50 and 175 meters underground, located near Geneva and spanning the border between France and Switzerland. It accelerates two beams of protons up to 6.8 TeV each, which travel in opposite directions and therefore produce head-on collisions of 13.6 TeV in energy. The collisions occur at four points along the ring, at which are stationed the major detectors: ATLAS, CMS, ALICE, and LHCb. ATLAS and CMS are both general-purpose detectors, and serve as cross-checks of each other. Meanwhile ALICE and LHCb are more specialized, with ALICE focusing on ion collisions and LHCb on bottom quark physics. The high energy collisions produced by the LHC, about 7 times larger than the previous record holder, the Tevatron at 2 TeV [26], allow the detector experiments to probe otherwise inaccessible physics at the finest scales.

The LHC is fed by a series of smaller accelerators, as seen in figure 3.1. The protons originate as negatively ionized hydrogen, which are then accelerated up to 160 MeV by Linear Accelerator 4. Then the Proton Synchrotron Booster strips the electrons from the hydrogen ions, leaving behind just the protons. The protons are accelerated to 2 GeV, when they then enter the Proton Synchrotron and are energized further to 26 GeV. Lastly the Super Proton Synchrotron accelerates the protons up to 450 GeV, where they are finally ready to enter the LHC ring. Over the course of 20 minutes, the protons pass through 16 radiofrequency cavities over 10 million times, which accelerate them to their final energy of 6.8 TeV [27, 28]. While proton-proton collisions are the main focus of the LHC, note that the facility does collide ions as well.

Note that the protons in the LHC do not actually form a continuous stream, but rather are grouped in bunches each containing around 10^{11} protons at the start of the fill cycle. The bunching is produced in the Proton Synchrotron accelerator stage. There are around 2800 bunches per beam, each separated by 25

The CERN accelerator complex *Complexe des accélérateurs du CERN*

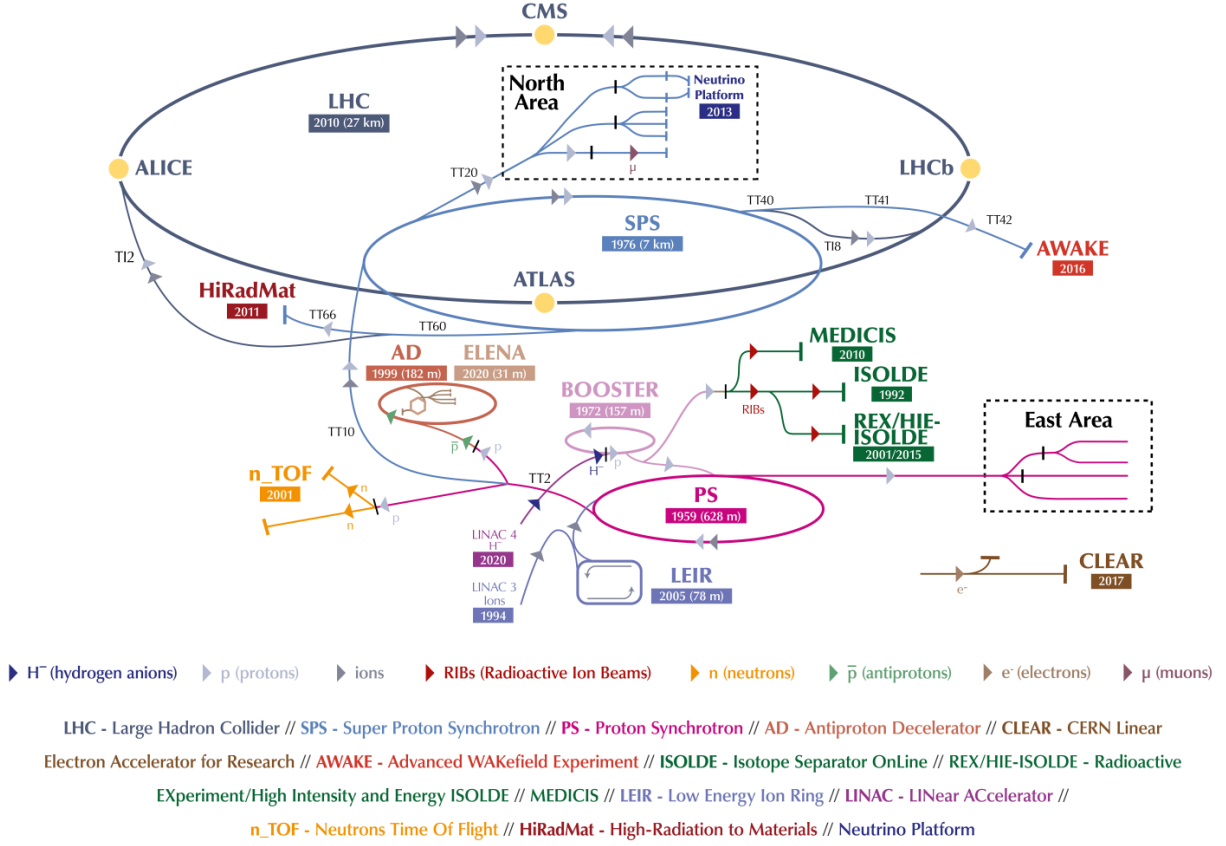


Figure 3.1: Schematic diagram of the CERN accelerator complex, including the LHC and several feeder accelerators [29].

nanoseconds [30]. Collisions occur when focusing magnets force a bunch from each beam to cross over each other, called a bunch crossing. Thus the events recorded by detectors also occur at a 25ns frequency, with one event for every bunch crossing. Note that only a tiny fraction of the protons in each bunch actually collide during a bunch crossing. This is referred to as pileup $\langle \mu \rangle$, or the average number of non-elastic proton-proton interactions per bunch crossing. Distributions of pileup are shown in figure 3.2 at ATLAS for the Run-2 data taking period (2015-2018) and the first two years of Run-3 data taking (2022-2023). Higher pileup means more collisions are happening, which corresponds to more data. But it also means a messier detector signature, as the decay products of the dozens of collisions all overlap.

Highly related to pileup is the concept of luminosity, which is used as a measure of how much data was

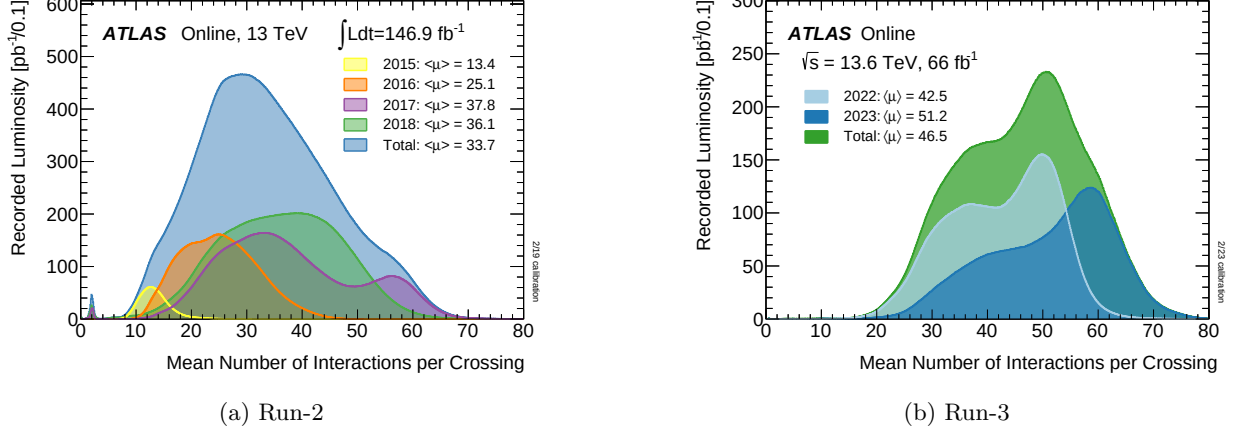


Figure 3.2: Measured pileup at ATLAS in (a) Run-2 (2015-2018) [31] and (b) the first two years of Run-3 (2022-2023) [32].

taken. The instantaneous luminosity is defined as

$$\sigma_{\text{inel}} \cdot L_{\text{inst}} = \langle \mu \rangle n_b f_r. \quad (3.1)$$

Here n_b is the number of colliding bunches per revolution, f_r is the bunch revolution frequency in the LHC (nominally 11246 Hz), and σ_{inel} is the cross section for inelastic collisions (nominally 80 mb) [31]. The total luminosity (also called the integrated luminosity) is the instantaneous luminosity integrated over time,

$$L = \int dt L_{\text{inst}}. \quad (3.2)$$

Where cross sections represent the probability of a process occurring, the luminosity corresponds to the total number of events measured, such that their product gives the number of times the process occurred:

$$N_{\text{process}} = \sigma_{\text{process}} L. \quad (3.3)$$

Since cross sections have units of area, a byproduct of classical scattering experiments where probabilities actually corresponded to geometric cross sections, the total and instantaneous luminosities have units of fb^{-1} and $\text{fb}^{-1}\text{s}^{-1}$ respectively. A summary of the total luminosity delivered by the LHC to ATLAS is shown in figure 3.3.

3.1.1 HL-LHC

Higher luminosity, which means more events of interest, directly correlates with the statistical precision of analyses. Therefore increasing the luminosity of the LHC can dramatically improve the sensitivity to rare BSM physics. Given equation (3.1), one way to increase luminosity is to increase the pileup $\langle \mu \rangle$. To this effect, an upgrade to the LHC program called the High-Luminosity LHC (HL-LHC) was ratified, with studies

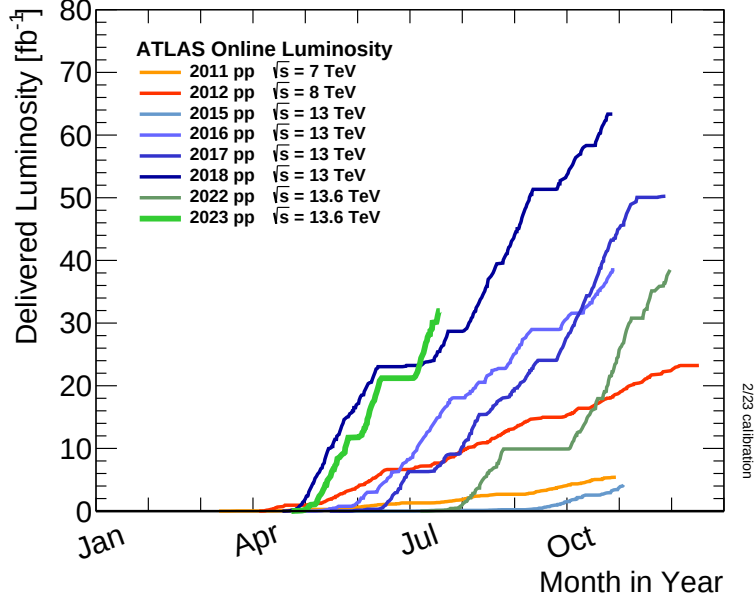


Figure 3.3: Total luminosity delivered to ATLAS between 2011 and 2023 [32].

beginning around 2010 and formal CERN approval in 2016 [33]. A proposed timeline of the project is shown in figure 3.4. The HL-LHC will increase the current pileup from around 50 to around 140, or in terms of peak instantaneous luminosity, from $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ to around $5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. This will provide around 250 fb^{-1} of integrated luminosity per year, a tremendous increase when compared to the approximately 350 fb^{-1} that will be collected by the end of the baseline LHC lifetime. Given a proposed runtime of about 12 years between 2029 and 2040, the HL-LHC will accumulate 3000 fb^{-1} , an order of magnitude more data than the LHC. An ultimate configuration, which may be possible if safety margins were excessive, could push the pileup to around 200. However, the increased luminosity will pose a tremendous technical challenge to the detectors, which were not originally designed for such high pileup. This is discussed further in the next section and forms the basis for the work in part II.

3.1.2 Run Nomenclature

The major periods of the LHC schedule, as shown in figure 3.4, are commonly referenced by name. These are:

- Run-1: Collisions at 7 TeV in 2011 and 8 TeV in 2012, totalling 26 fb^{-1} of data recorded by ATLAS.
- Long Shutdown 1 (LS1): Idle period between 2013 and 2014. Upgrades to ATLAS during this period are known as Phase-0, most notably the addition of the insertable B-layer (IBL).
- Run-2: Collisions at 13 TeV between 2015 and 2018, totalling 147 fb^{-1} of data recorded by ATLAS.



Figure 3.4: Proposed HL-LHC timeline, from [34].

- Long Shutdown 2 (LS2): Idle period between 2019 and 2022. Upgrades to ATLAS during this period are known as Phase-I, most notably the addition of the New Small Wheels.
- Run-3: Collisions at 13.6 TeV starting in 2022 and planned to end 2025.
- Long Shutdown 3 (LS3): Idle period planned from 2026 to 2029. Upgrades to ATLAS during this period are known as Phase-II. Here the HL-LHC will be installed, and major upgrades to ATLAS will correspondingly occur.
- Run-4: Planned to start around 2029 using the upgraded luminosity from the HL-LHC.

3.2 The ATLAS Detector

ATLAS is one the two general-purpose detectors on the LHC ring that analyzes the collisions produced by the LHC [35, 36]. The term general-purpose reflects how ATLAS was designed to analyze a variety of phenomena, and so the detector consists of several layered components each developed to help identify different collision products and their properties. A cut-away diagram of the detector is shown in figure 3.5, while an illustration of a transverse slice is shown in figure 3.6, which highlights the path of different particles through the subsystems.

The detector is cylindrical in shape, with the axis of symmetry corresponding to the beamline of the LHC, and centered around the nominal collision point, also called the interaction point (IP). It is 25 meters

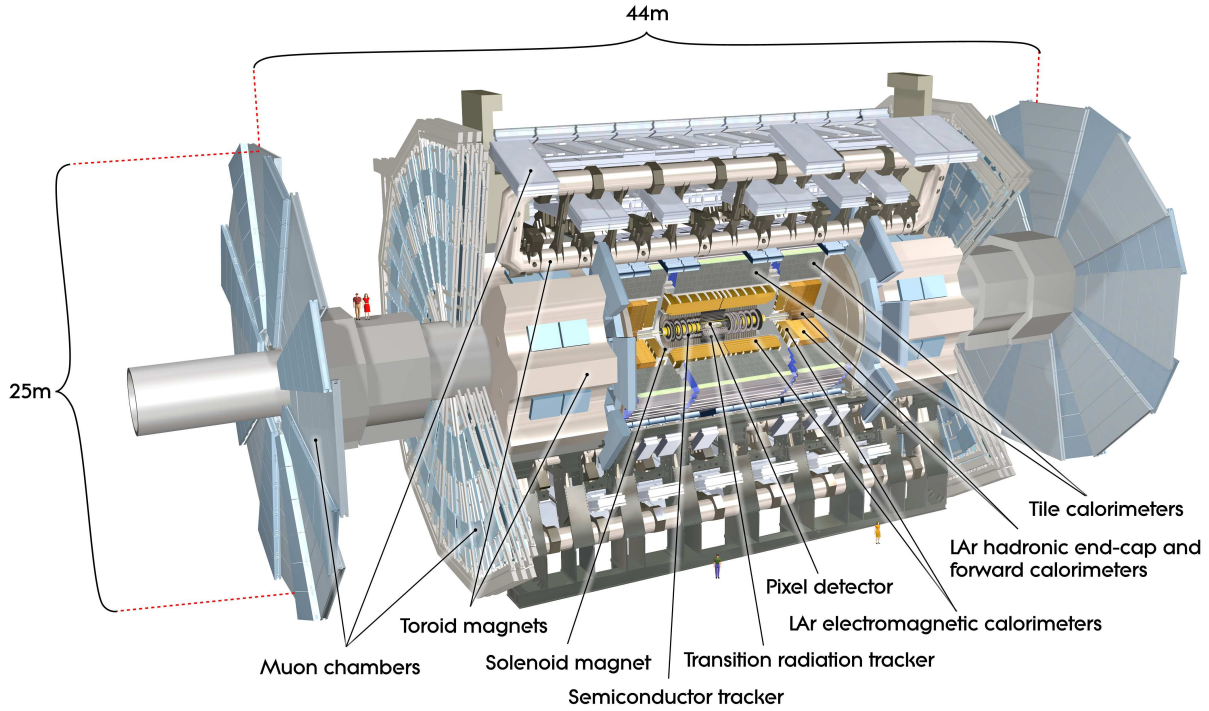


Figure 3.5: General cut-away view of the ATLAS detector [35].

wide in diameter and 44 meters long, weighing a total of around 7,000 metric tons. The detector consists of several stacked layers surrounding the IP. Near the center, the layers form concentric cylinders, termed the *barrel region*. Meanwhile near the two ends of the detector, the layers form vertical discs facing the IP, centered on and perpendicular to the beamline. These are referred to as the *end-caps*.

ATLAS is divided into several subsystems based on their role, with each containing several layers. From inside-out, these are:

1. Inner detector: measures the position of charged particles as they traverse the subsystem.
2. Calorimeters: captures and measures the energy of the majority of particles.
3. Muon spectrometer: measures the energy and trajectory of muons.

There are also two magnet systems that force charged particles to bend, which enables measuring their charge and momentum. Finally, a sophisticated Trigger and Data Acquisition (TDAQ) system manages readout from the sensor systems. These will be described further in the corresponding subsections.

3.2.1 Coordinate System

The cylindrical structure of the detector engenders the use of a cylindrical coordinate system used throughout ATLAS. The origin is located at the IP. The x -axis points towards the center of the LHC ring, the y axis

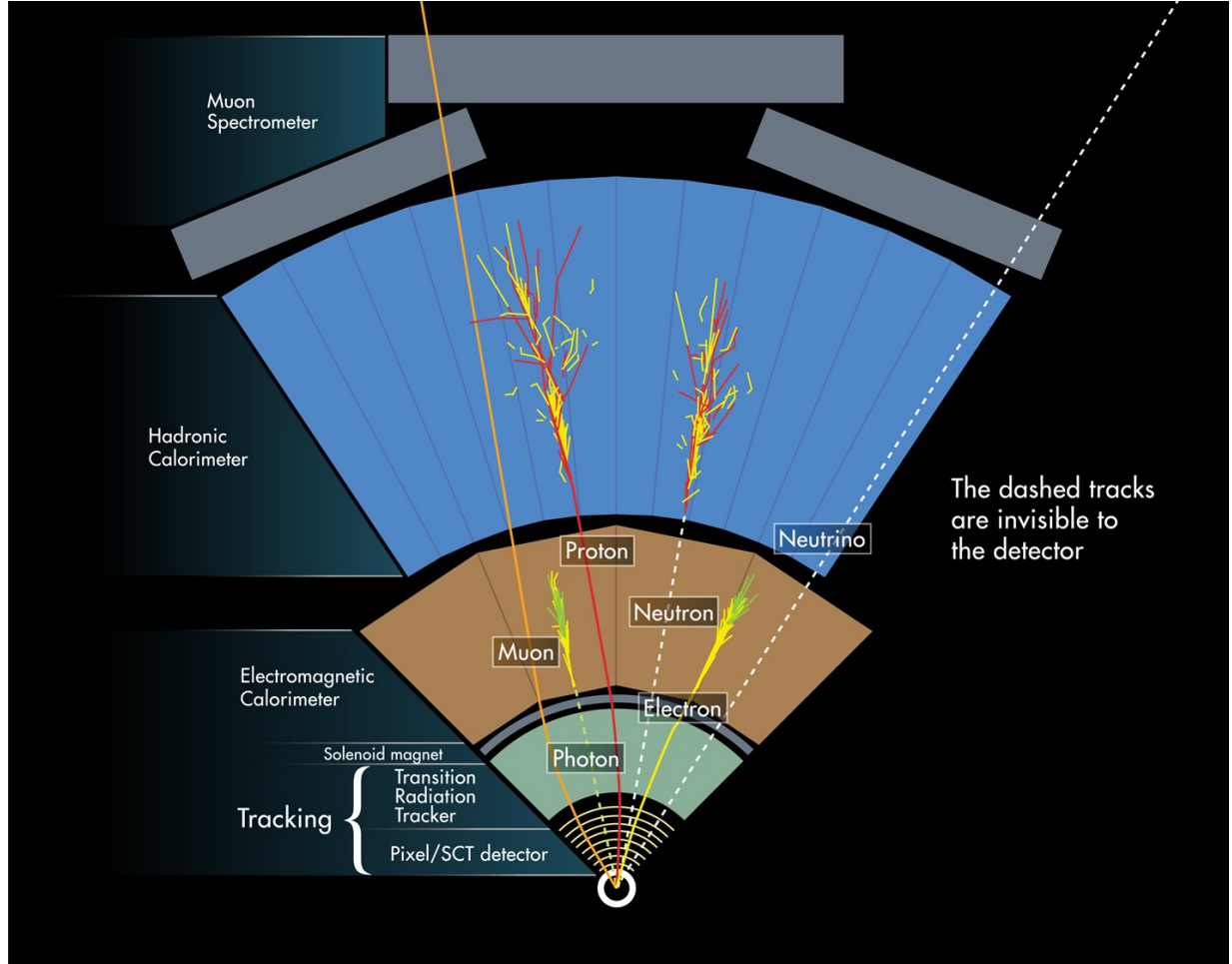


Figure 3.6: Illustration of the path of different particles through the ATLAS detector in the transverse plane. This figure is not to scale [37].

is oriented upwards, and the z axis is oriented along the beamline, with the positive z direction forming a right-handed system (in the counter-clockwise direction of the LHC ring when viewed from above). The cylindrical coordinates are defined in the usual way as $r = \sqrt{x^2 + y^2}$ and $\varphi = \arctan(y, x)$. The x - y plane is termed the *transverse* plane.

In analyses though, one often switches to a spherical coordinate system. However instead of using the polar angle $\theta = \arccos z/r$, experimentalists often use the rapidity or pseudorapidity instead. These are defined as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad \eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (3.4)$$

respectively. They are preferred because, even though the protons at the LHC collide head-on with equal energy, the net z momentum of the collision products is not zero. This is due to the proton being a composite particle, which means the individual partons actually engaging in the collision carry an unfixed fraction of the

z momentum. Rapidity ameliorates this subtlety because differences in rapidity are Lorentz invariant along boosts in the z direction. However rapidity requires knowing both the total energy and the z momentum of a particle, whereas pseudorapidity only requires the polar angle. Pseudorapidity converges to rapidity in the limit that the particle is highly relativistic, or $m \ll E$. An angular separation between two objects is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

Tracks, or the trajectories of charged particles in the detector, are generally referenced by their parameters at the point closest to the beamline. The z and r of the point of closest approach are denoted z_0 and d_0 respectively. Meanwhile the direction of the track is measured by η and ϕ , which are the direction of the tangent at the point of closest approach. Together with the transverse momentum p_T , these give an almost complete description of a particle trajectory.

3.2.2 Inner Detector

The inner detector precisely measures the trajectories, or tracks, of charged particles as they fly through the system [38]. The tracks give the origin and momentum of each charged particle produced in an event. An illustration of the inner detector is shown in figure 3.7, and a layout is shown in figure 3.8. It is the innermost component of the ATLAS detector, occupying a cylindrical volume up to $r = 115\text{cm}$ and $z = \pm 351.2\text{cm}$, covering a pseudorapidity range up to $\eta = \pm 2.5$. The inner detector is composed of three subsystems, which get larger and less granular the further from the IP, representing the cost trade-off between coverage and precision. Ordered inside-out they are a pixel detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT).

3.2.2.1 Pixel Detector

The pixel detector is the innermost system in all of ATLAS, and has the highest granularity out of all the inner detector components. It is composed of four barrel layers and three end-caps on each side. The innermost barrel layer, the IBL, was a new addition added in 2014 [41]. The largest barrel layer has a radius $r = 122.5\text{mm}$, while the furthest end-cap is located at $z = \pm 650\text{mm}$. The pixel detector contains around 92.3 million individual silicon sensors or *pixels*, similar in concept to the pixels in digital cameras. Each pixel is typically $50\mu\text{m}$ in pitch and $400\mu\text{m}$ in length, though the IBL pixels are smaller at $250\mu\text{m}$ in length. A fraction of the pixels are longer to accommodate readout electronics. The detector registers hits by measuring the charge deposited in the silicon when a charged particle travels through the sensor.

3.2.2.2 Semiconductor Tracker

The SCT sits outside the pixel detectors, with radius $275\text{mm} < r < 560\text{mm}$ and end-caps up to $z = \pm 2735\text{mm}$. It is composed of four barrel layers and nine end-caps on each side. The SCT uses a similar

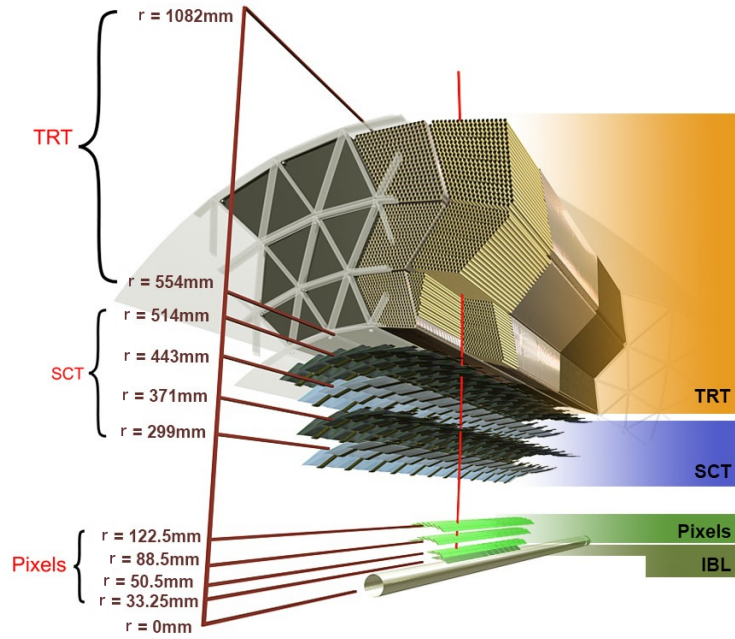


Figure 3.7: Illustration of a portion of the barrel of the ATLAS inner detector [39].

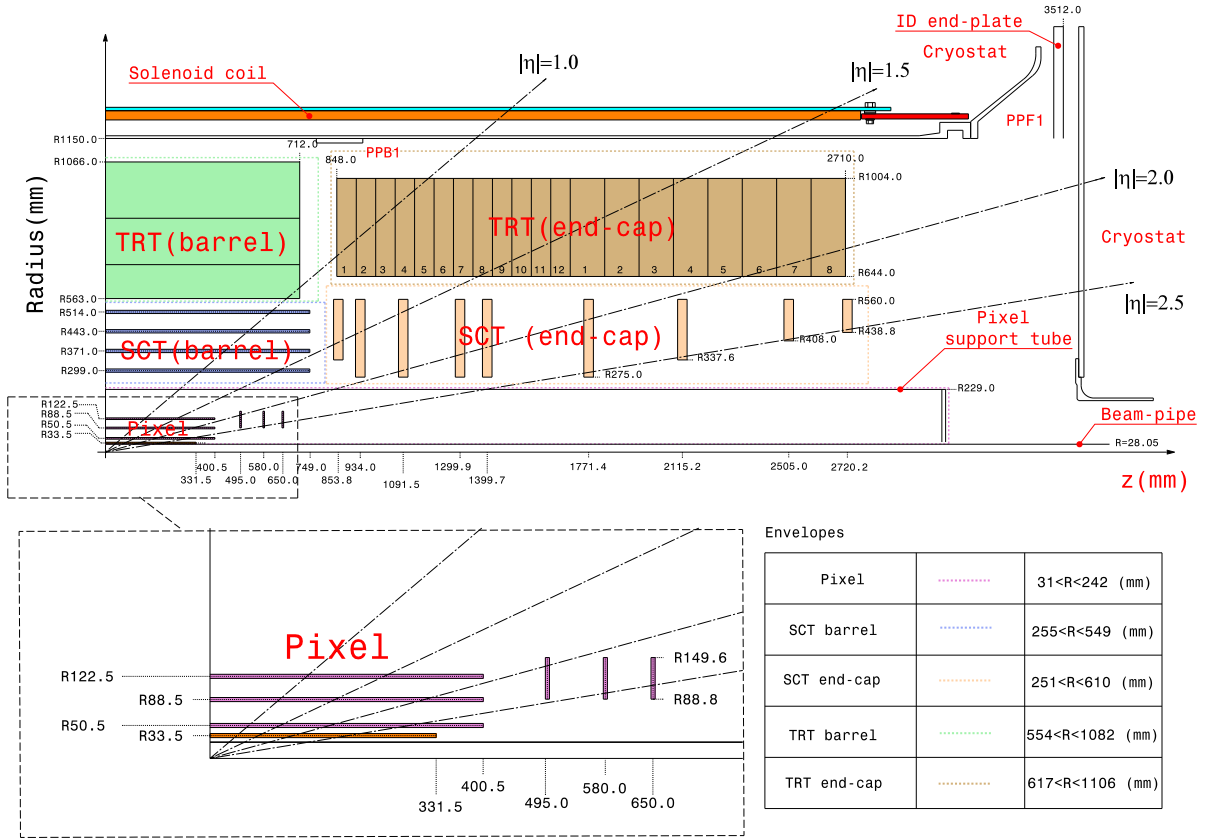


Figure 3.8: Layout of the ATLAS inner detector, viewed as a cross-section in the r - z plane [40].

technology as the pixel detector, relying on silicon sensors that become ionized as a charged particle traverses them. However the SCT sensors are much larger, typically 126mm in length and $80\mu\text{m}$ in pitch. These are often referred to as *strips*, and the SCT as the strip detector. Strips are aligned parallel to the beamline in the barrel and perpendicular to the beamline in the end-caps. Each layer actually consists of two sub-layers, where the strips are placed back-to-back but offset by a small stereo angle of 40 mrad. The angle offset allows the experiment to determine a precise longitudinal location of a particle despite each strip being exceptionally long, since the overlapping region of back-to-back strips is now much more localized. The SCT contains a total of 6.3 million channels.

3.2.2.3 Transition Radiation Tracker

The TRT is the outermost subsystem in the inner detector, with radius $563\text{mm} < r < 1066\text{mm}$ and end-caps up to $z = \pm 2710\text{mm}$. The TRT uses a much different technology than the silicon detectors, relying instead on 4mm diameter straws containing xenon or argon gas mixtures, running the full length of barrel modules (144cm) or radius of end-cap modules (37cm). A wire runs down the center of each straw, and a 1530V voltage is applied between the wire and casing. When a charged particle passes through the TRT, it can interact directly with the gas, or indirectly by emitting transition radiation as it traverses polypropylene fibers surrounding the straws. These both ionize the gas, producing a current through the center wire. The transition radiation has a stronger signal than the charged particle itself, and is also proportional to the particle's Lorentz factor, which can aid in electron identification. The TRT contains a total of 350 thousand readout channels.

3.2.3 Calorimeters

Whereas the inner detector focuses on measuring the trajectories of charged particles, which give their positions and momentums, the calorimeters focus on measuring the energies of both charged and neutral particles. To do so, they generally stop a particle completely, inducing particle showers that use up all the original particle's energy. These systems contain alternating layers of absorbing mediums, which induce the showers, and active mediums, which are used to measure the resultant showers. An illustration of the calorimeters is shown in figure 3.9, and a cross-sectional layout is shown in figure 3.10. ATLAS employs two types of calorimeters, one designed to measure electromagnetic energies and another for hadronic particles. As the calorimeters effectively destroy the particles they measure, they sit outside the tracking layers in the inner detector, which are conversely designed to minimally affect a particle's energy. The barrel starts at $r = 1.4\text{m}$ and the whole system extends to around $r = 4.25\text{m}$. The end-caps extend from around $3.5\text{m} < |z| < 6.1\text{m}$, providing pseudorapidity coverage of $|\eta| < 4.9$.

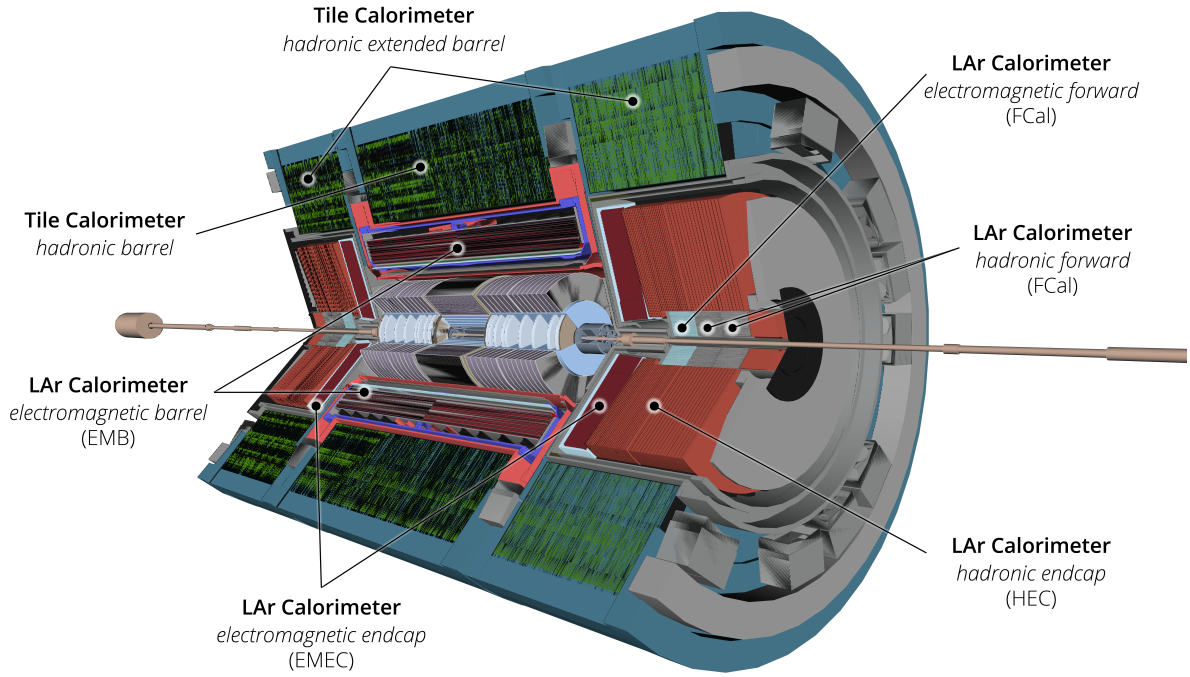


Figure 3.9: Cutaway illustration of the ATLAS calorimeters [36].

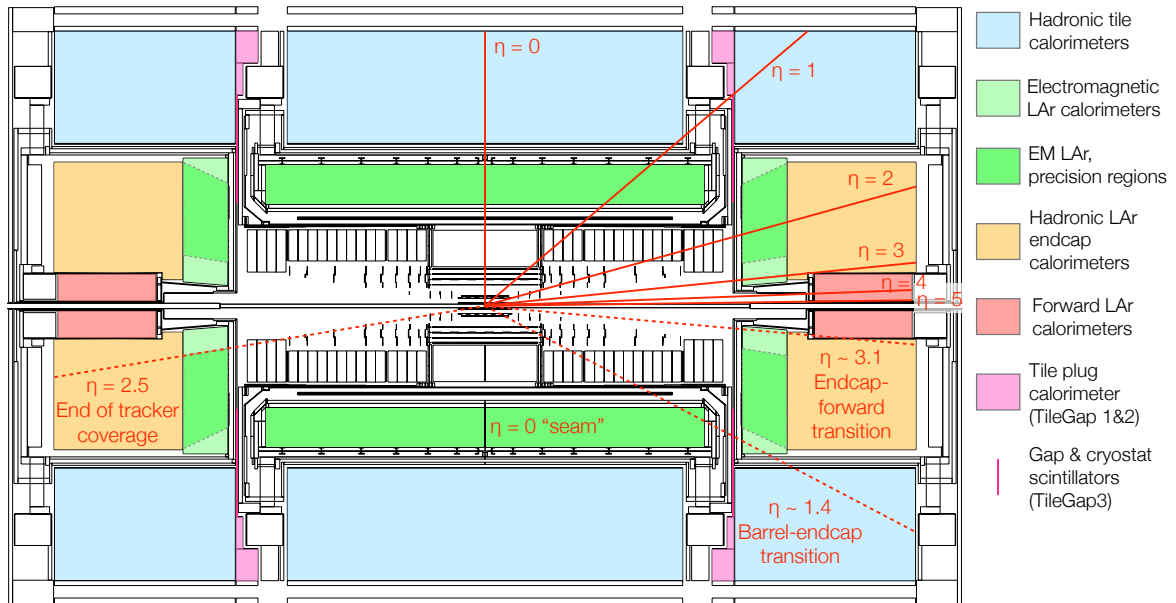


Figure 3.10: Layout of the ATLAS calorimeters, viewed as a cross-section in the r - z plane [42].

3.2.3.1 Electromagnetic Calorimeter

The electromagnetic calorimeter is the first layer, comprising $1.4\text{m} < r < 2\text{m}$ in the barrel and a single end-cap on each side at around $3.7\text{m} < |z| < 4.3\text{m}$, giving a pseudorapidity coverage of $|\eta| < 3.2$. Lead plates between 1.13mm and 2.2mm thick are used as the absorbing material, folded several times in an accordion structure that prevents gaps in φ . Liquid argon (LAr) fills the gaps and is used as the active material. Charged particles, primarily electrons, interact with the lead nuclei via bremsstrahlung to produce photons, while photons themselves will pair produce e^+e^- . This creates a shower of particles that ionize the LAr and produce an electronic signal which corresponds with the total energy deposited. The electromagnetic calorimeter consists of three layers with cells of varying granularity, which gives additional position measurements. The total thickness of the electromagnetic calorimeter is over 22 radiation lengths (X_0), which means a typical electron will only retain $e^{-22} \approx 3 \times 10^{-10}$ of its original energy by the time it leaves the calorimeter, meaning very few electrons leak past it. The electromagnetic calorimeter provides a total of 170 thousand readout channels.

3.2.3.2 Hadronic Calorimeter

The hadronic calorimeter sits outside the electromagnetic calorimeter, and consists of two distinct components: a tile calorimeter in the barrel and a LAr calorimeter in the end-cap. The tile calorimeter uses steel as the absorbing medium and scintillating tiles as the active medium, stacked in periodic layers at approximately a 4.7:1 volume ratio. It covers $2.28\text{m} < r < 4.25\text{m}$, with disjoint barrel extensions pushing out to around $z = \pm 6\text{m}$ and providing a pseudorapidity coverage of $|\eta| < 1.7$. It consists of three sensor layers that contribute to 9.7 interaction lengths (λ) at $\eta = 0$, where 1λ is defined as the mean distance a pion travels before interacting. Like in the electromagnetic calorimeter, when strongly interacting particles collide with the steel, they produce a shower of other particles that activate the scintillators, which are read through fiber optics and photomultiplier tubes to determine the energy deposited. The tile calorimeter has a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, except in the last layer which has $\Delta\eta = 0.2$, totaling 9,852 readout channels.

In the end-caps, a copper-LAr system is used. Each end-cap consists of two wheels with two layers each, extending from $0.475\text{m} < r < 2.03\text{m}$, except near the innermost side where the inner radius shrinks to 0.372m. They cover the range of around $4.3\text{m} < z < 6.1\text{m}$, which gives a pseudorapidity of $1.5 < |\eta| < 3.2$. The copper absorbers are 25mm and 50mm thick plates in the inner and outer wheels respectively, interleaved with 8.5mm of LAr as the active medium. This provides a total of 10 interaction lengths. These end-caps have a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ or 0.2×0.2 at $|\eta| > 2.5$, totalling 5,632 readout channels.

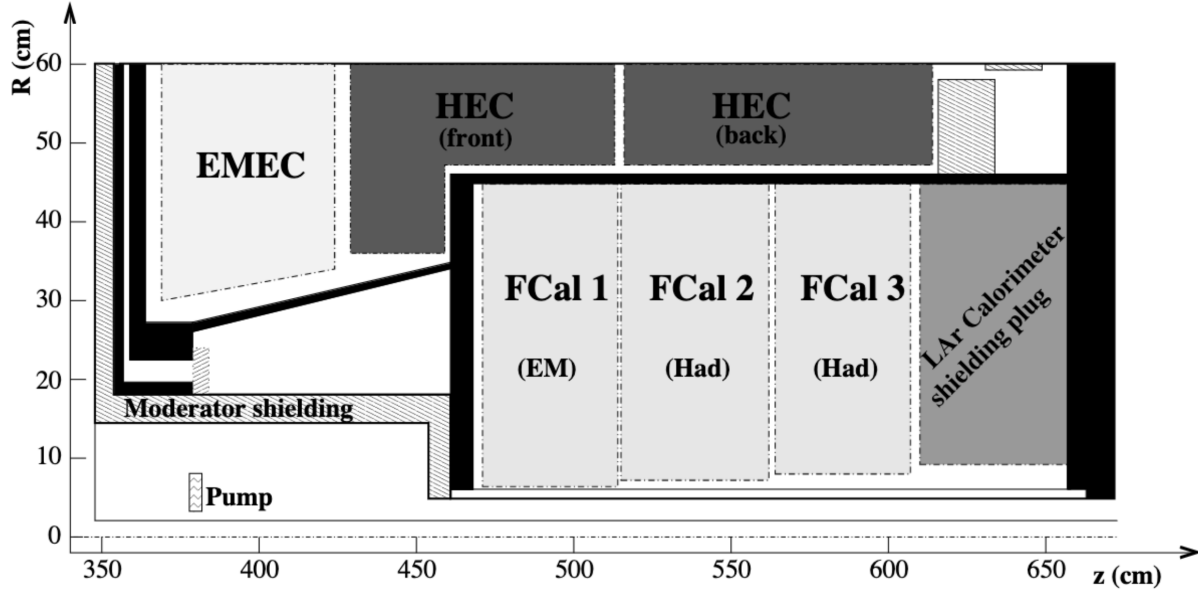


Figure 3.11: Layout of the ATLAS forward calorimeters, viewed as a cross-section in the r - z plane. Note that the vertical scale has been enlarged [35].

3.2.3.3 Forward Calorimeters

A special section is devoted for a forward calorimeter system that covers the highest pseudorapidities, $3.1 < |\eta| < 4.9$, corresponding to $72\text{mm} < r < 449\text{mm}$ and $4.7\text{m} < z < 6.05\text{m}$ [43]. A layout diagram is shown in figure 3.11. For each end-cap, there are three modules: an inner electromagnetic module using copper-LAr, and two hadronic modules using tungsten-LAr. These modules have a unique structure, consisting of numerous rods of the absorbing metal aligned parallel to the beamline, with each rod contained in a copper tube, separated by a thin (0.27–0.51mm) envelope of LAr. The tubes are surrounded by more of the absorbing material, either stacked copper plates or slugs of tungsten, forming a hexagonal array of cells. These provide a typical granularity of $\Delta x \approx 30\text{mm} - 54\text{mm}$ and $\Delta y \approx 26\text{mm} - 47\text{mm}$, totalling 3,524 readout channels. The electromagnetic module has a radiation length of $X_0 = 27.6$, while the whole system has an interaction length of around $\lambda = 10$.

3.2.4 Muon Spectrometer

The calorimeter systems capture and stop the majority of particles, but there are two key exceptions: muons and neutrinos. Neutrinos are very weakly interacting, and they mostly leave the detector untouched and unmeasured. Muons meanwhile do interact with the detector, but have enough mass that they punch through the calorimeters, only losing around 3 GeV on the way. Thus ATLAS uses a dedicated muon spectrometer to measure muon tracks and momentums. It is the outermost system in the detector, at

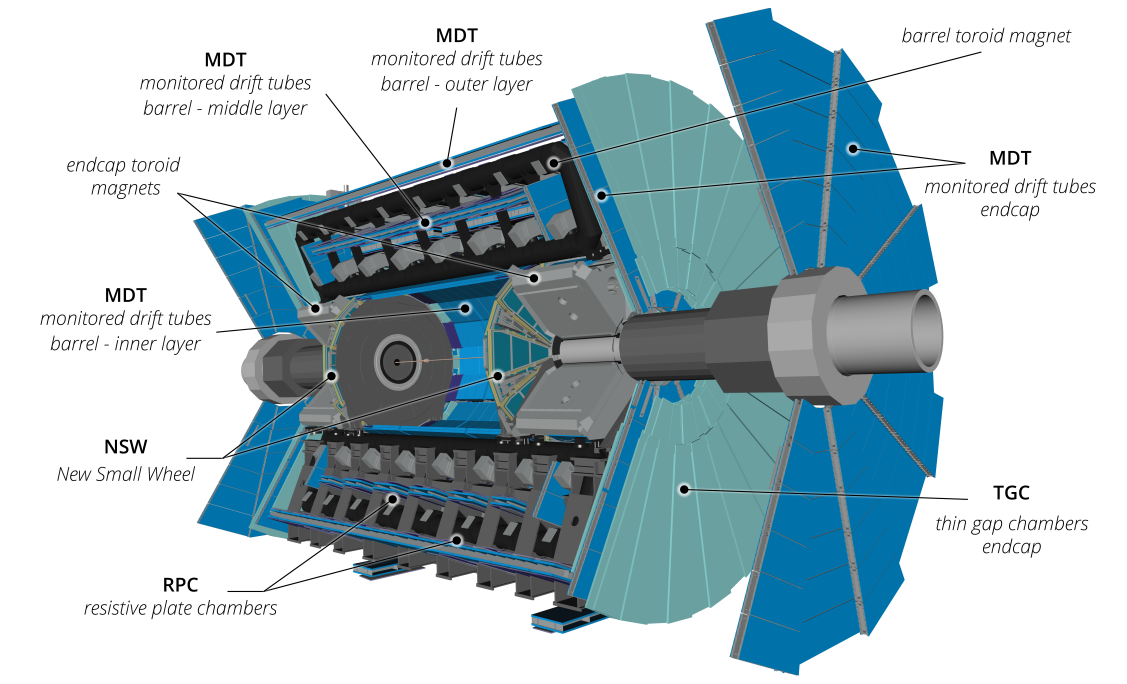


Figure 3.12: Cutaway illustration of the ATLAS muon spectrometer in Run-3 [36].

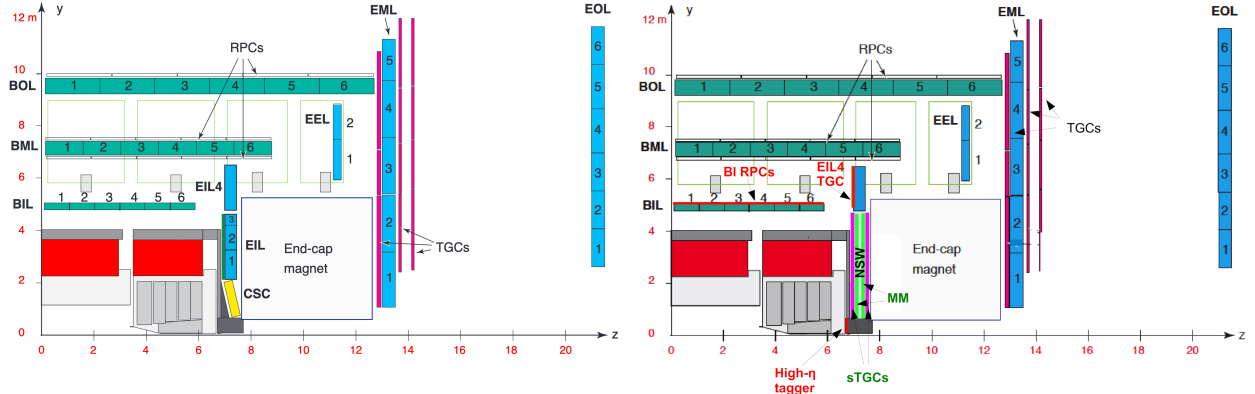


Figure 3.13: Layout of the ATLAS muon spectrometer large sector, viewed as a quarter cross-section in the r - z plane, in Run-2 (left) and Run-3 (right). The large red blocks are the tile calorimeters. Images from [36, 44].

a range of approximately $5\text{m} < r < 10\text{m}$ in the barrel and $7\text{m} < z < 21\text{m}$ in the end-caps, covering a pseudorapidity range of $|\eta| < 2.7$.

An illustration of the muon spectrometer is shown in figure 3.12 and a layout diagram is shown in figure 3.13. There are three layers in the barrel and four in each end-cap. Each layer is actually a staggered bi-layer, consisting of 8 large sectors at the cardinal and ordinal directions, and 8 small sectors filling the gaps. The primary sensors of the muon spectrometer are Monitored Drift Tubes (MDTs) [44]. These are similar in function to the TRT; each MDT is 3cm diameter aluminum straw filled with argon, separated

by a 3,080V difference from a tungsten-rhenium wire in the center. Only the innermost module of the innermost end-cap layer, which receives a much higher flux of particles, uses a different technology, Cathode Strip Chambers (CSCs). These are proportional chambers with arrays of anode wires and cathode strips separated by 1,900V. The chambers are filled with argon, and conceptually operate very similarly to the MDTs but with strips instead of straws. The MDTs and CSCs produce 367 thousand and 31 thousand readout channels respectively. In Run-3, the innermost end-cap layer was replaced by the New Small Wheel (NSW) [45], mostly motivated by trigger constraints. But as the analysis in part III uses Run-2 data, a discussion of this upgrade will be omitted.

Aside from the main MDT tracking chambers, the muon spectrometer also utilizes several trigger chambers placed alongside the MDTs. These are optimized to provide info to the trigger system (see section 3.2.6), but also help measure muon tracks in the orthogonal direction from the MDT straws. In the barrel Resistive Plate Chambers (RPCs) are used, while in the end-caps Thin Gap Chambers (TGCs) are used. RPCs use parallel plastic laminate plates separated by 2mm of primarily gaseous $C_2H_2F_4$, held at a 9,800V potential difference. Each chamber varies in length (in z direction) between 320mm and 1200mm, and in width (ϕ direction) from 1180mm to 3680mm [44]. Meanwhile the TGCs are again proportional chambers using graphite cathodes held at 2,900V and a gas mixture of CO_2 and $n - C_5H_{12}$. They are so named because the anode-cathode gap of 1.4mm is smaller than the anode-anode wire gap of 1.8mm, which help decrease the time jitter in the signal [46]. The RPCs and TGCs contribute a total of 383 thousand and 320 thousand readout channels respectively.

3.2.5 Magnet System

ATLAS uses two superconducting magnet systems to bend charged particle tracks. The bending radius is roughly proportional to qp/B , so by immersing the detector in a magnetic field, one can determine both the momentum and charge of particles. An illustration of the magnet systems is shown in figure 3.14. The first system is a solenoid magnet placed outside the inner detector, providing a nominal 2T field in the z direction across the entire inner detector. It uses a Al-Cu-NbTi coil, passing 7.73kA of current through 1,154 turns. A detailed measurement of the actual magnetic field is shown in figure 3.15. The second system is a toroid magnet, consisting of one barrel and two end-cap systems with eight coils each. These are placed outside the calorimeters and interleaved with the muon spectrometer, providing bending power to the muon spectrometer. They use a similar Al-Cu-NbTi coil, but with a current of 20.5kA through 120 or 116 turns in the barrel and end-caps respectively. These provide a field in the bore between 0.2T and 3.5T, leading to between 1Tm and 7.5Tm integrated across an infinite-momentum muon trajectory.

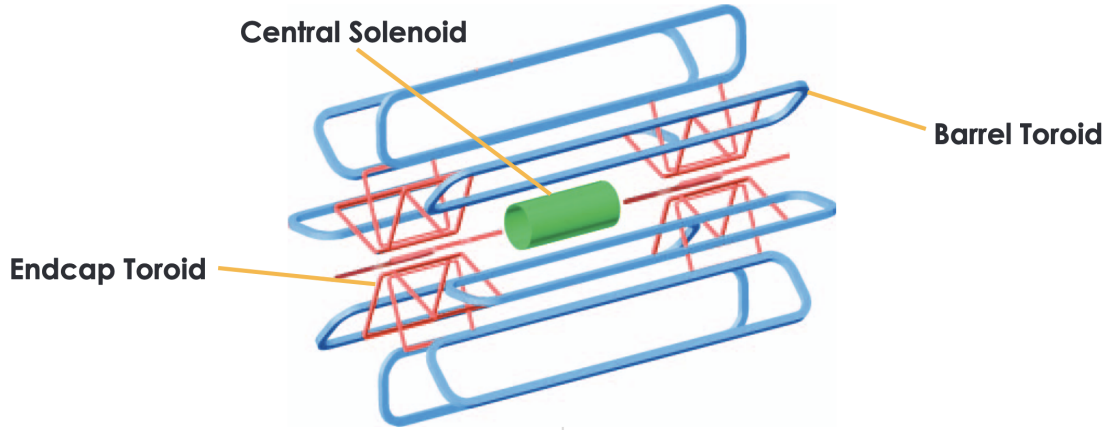


Figure 3.14: Illustration of the ATLAS magnet system in isolation [47].

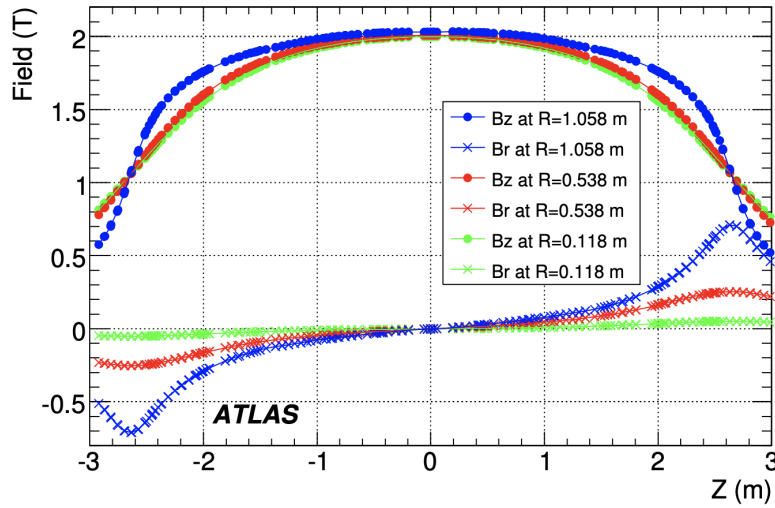


Figure 3.15: Magnetic field map of the ATLAS solenoid [48].

3.2.6 Trigger and Data Acquisition

The over 100 million combined readout channels from the detector components outlined above lead to an average event size of 3 MB at peak luminosity. Given the nominal LHC collision rate of 40 MHz, this would lead to a data rate of 120 TB/s. Such a rate is completely unmanageable, in terms of both providing enough bandwidth and the cost of data storage. Therefore ATLAS uses a *trigger* system to do early filtering of the event data before it is even read off the detector. The system is named such because it triggers the detector modules to readout their data, but only when it encounters an event worth saving. As the majority of events at the LHC are soft collisions, the trigger system typically selects the rarer hard events, which are also generally more interesting to physics.

Since the trigger is closely coupled with the readout systems, these components are managed under a

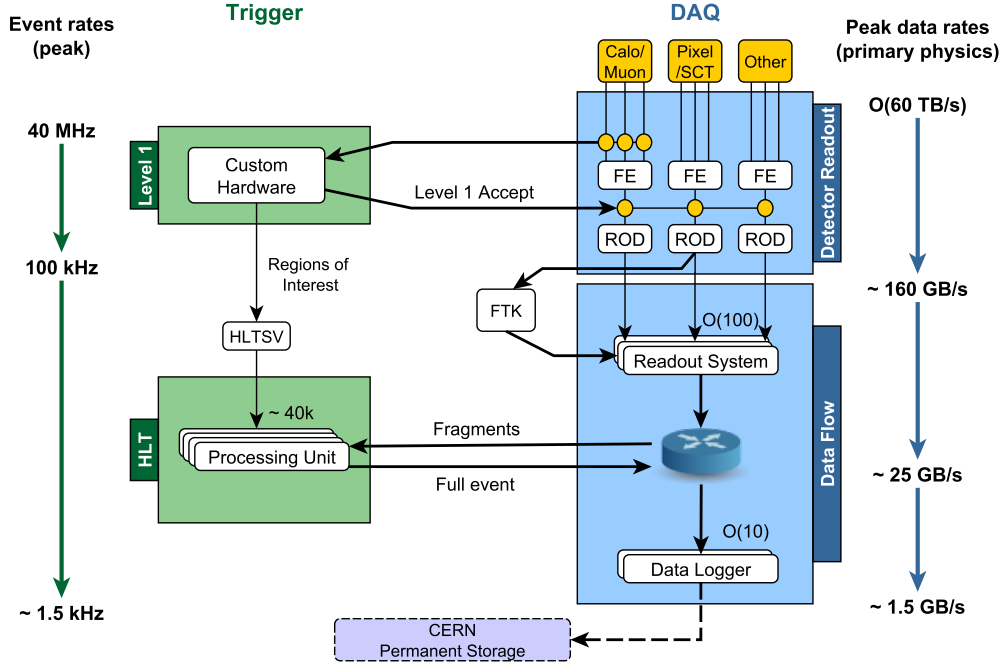


Figure 3.16: Diagram of the ATLAS TDAQ system in Run-2 [49]. Note the FTK system was ultimately not implemented.

single Trigger and Data Acquisition (TDAQ) group. A flowchart of the major components as used in Run-2 is shown in figure 3.16. The trigger mechanism is broken down into two steps. The first is called the Level-1 (L1) trigger, which selects on average one out of every 400 events, taking the original 40 MHz event rate down to 100 kHz. The L1 trigger uses only information from the calorimeters and muon spectrometer, and is implemented in hardware and field-programmable gate arrays (FPGAs) which allow it to process in event in less than $2.5 \mu\text{s}$. When an event is accepted by the L1 trigger, the entire detector is readout into a temporary buffer which is passed to the second step, the High-Level Trigger (HLT). This second stage filters the event rate down to about 1.5 kHz, which can now be saved to permanent storage. The HLT uses a CPU farm of around 40,000 processors to check a variety of criteria for selecting events, with an average latency of 400 ms.

3.2.7 Object Reconstruction

Once events are saved to permanent storage, they can finally be analyzed for physics. However, one should remember that the detector is not measuring particles directly, but only the hits they leave behind, consisting of typically a position vector and sometimes a voltage or current measurement that corresponds to energy.

Mapping the individual hits back to particles and their trajectories is called *reconstruction*. This is a highly non-trivial task, as due to pileup there can be up to 60,000 hits in the inner detector alone [50], representing a complex combinatorial problem. These are further muddled by the close proximity of the pileup collisions, but also by many sources of noise such as beam-induced backgrounds.

The reconstruction workflow can be summarized by a few key steps. First, because a particle will typically activate multiple neighboring cells in each detector layer, nearby hits are clustered together for better accuracy and to reduce redundancy. Then, tracks, the presumed trajectories of charged particles, are constructed from the precision hits in the inner detector, giving the particles' momentums and collision vertices. Finally, the tracks are matched to calorimeter energy deposits to determine their energies and particle types; unmatched calorimeter deposits are identified as neutral particles. Inner detector tracks are also referenced with the muon spectrometer hits to identify and more precisely measure muons. A summary of the main particles and how they interact with the detector is shown in figure 3.6.

3.2.7.1 Tracking

A schematic flowchart of the ATLAS tracking algorithm used in Run-2 is shown in figure 3.17 [50]. It consists of a two-stage approach. In the first stage, triplets of pixel and SCT hits are used to create track seeds. These seeds are iteratively expanded with hits from other layers using a combinatorial Kalman Filter [51] to form track candidates. The candidates are passed through an overlap removal and χ^2 fit [52] to limit fakes, and then finally extended to the TRT.

This first main pass is optimized for prompt particles, those produced in a p - p collision. A second back-tracking pass is run to find missed particles, especially those deriving from a secondary vertex. This pass uses only hits not already accounted for in the first pass, and focuses only on regions of interests as determined by electromagnetic calorimeter readings that are matched with TRT segments. The track reconstruction then proceeds similar to the first pass. Tertiary passes are conducted to cover special cases, such as muons with $|\eta| > 2.5$. An iterative procedure is finally used to fit tracks to interaction vertices.

3.2.7.2 Electrons and Photons

Electrons and photons are generally captured by the electromagnetic calorimeter, which is their defining signature [53]. Hits from both the electromagnetic and hadronic calorimeters are first combined into topological clusters [54]. Clusters where the electromagnetic energy is greater than 400 MeV and also accounts for more than half the total energy are selected as seeds for electron and photon reconstruction. Electron candidates are chosen when a seed cluster can be matched to a track, given metrics that depend on the position, momentum, and energy of the track and cluster.

Photons, which do not leave tracks in the inner detector, are then identified from the unmatched clusters.

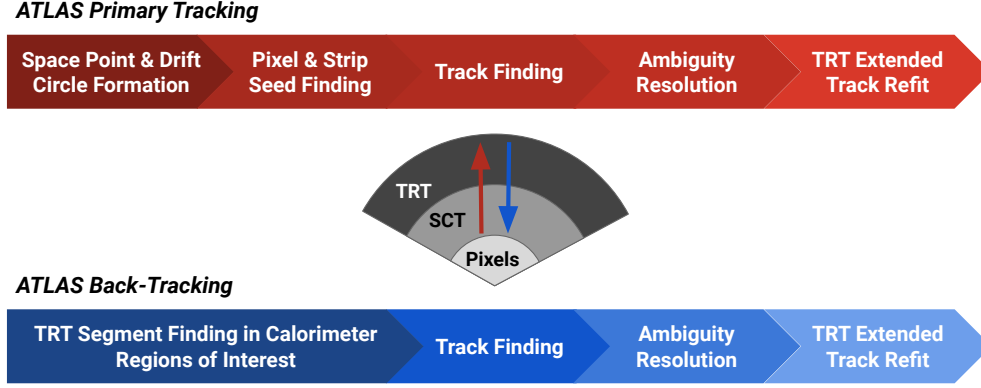


Figure 3.17: Flowchart of the ATLAS tracking algorithm in Run-2 [50].

However, a special procedure is first done to account for converted photons, since around 30% of photons will convert, such as pair producing, before reaching the calorimeters. This entails finding atypical tracks that could be explained by a photon conversion, then matching the seed clusters to these conversion tracks. Any remaining seed clusters are identified as unconverted photons.

A variety of extra filters can then be applied to separate prompt particles from background objects, such as misidentified objects or non-prompt decays from b - and c -hadrons. These filters use discriminating variables including the track properties and the shape of the calorimeter showers. By tuning the threshold of these cuts, one can obtain a series of working points that trade-off efficiency with background rejection. ATLAS uses three working points called Loose, Medium, and Tight, in increasing order of background rejection [53, 55].

3.2.7.3 Muons

Muons are readily identified because they leave tracks in the muon spectrometer, while the majority of other particles will be stopped completely by the calorimeter systems. First, tracks in the muon spectrometer are reconstructed by first creating segments in each layer from the MDT and CSC layers. Then the segments are combined based on a loose assumption of the IP and a parabolic trajectory. Measurements from the RPC and TGC layers are then added to incorporate the cross-axis coordinates, and then finally several iterative χ^2 fits are performed to arrive at a final track candidate.

ATLAS then uses five different reconstruction strategies, to combine information from the inner detector, calorimeters, and the muon spectrometer:

1. Combined: muon spectrometer tracks are matched with inner detector tracks.
2. Inside-out combined: inner detector tracks are extrapolated to muon spectrometer hits.

3. Muon-spectrometer extrapolated: muon spectrometer tracks that cannot be combined with an inner detector track and are extrapolated to the beamline.
4. Segment-tagged: inner detector tracks are extrapolated to at least one muon spectrometer segment, but the track parameters are taken only from the inner detector hits.
5. Calorimeter-tagged: inner detector tracks are extrapolated to calorimeter deposits consistent with a muon, but again the track parameters are taken only from the inner detector hits.

As with electrons and photons, a set of filters are then used to reject backgrounds and mis-identified tracks. These use a variety of variables that include the number of hits, the χ^2 of the fit, and the reconstruction strategy as listed above that was used. This again produces three primary working points called Loose, Medium, and Tight, in increasing order of background rejection. In addition, two special High- p_T and Low- p_T working points are defined for extreme p_T regions [56].

3.2.7.4 Jets

The next class of particles are the strongly interacting ones, namely quarks and gluons. The strong force is strong enough that quarks and gluons cannot live in isolation, a phenomenon called color confinement. Instead, they repeatedly pair produce new quarks, ultimately producing a shower of colorless hadrons. Due to conservation of momentum, these hadrons are aligned in a cone around the direction of the originating quark or gluon. Such cones of hadrons are called jets, and are the key signature of quarks and gluons in the detector.

Reconstructing jets requires some method to group nearby hadronic activity together. ATLAS uses the anti- k_t algorithm [57], which iteratively groups nearby measurements together based on a distance metric derived from a flexible choice of ΔR . ATLAS typically uses two settings, $\Delta R = 0.4$ and $\Delta R = 1.0$, termed small-R and large-R jets respectively. The $\Delta R = 1.0$ jets are also often called *fatjets*, and are primarily used to reconstruct boosted objects that decay into quarks, like W and Z bosons. At high p_T , the individual quark jets will start to overlap and it would be difficult to resolve them independently, so they are instead reconstructed using a single fatjet. Substructure techniques, which look at the individual constituents that make up the jet, are later used to differentiate boosted object decays from other background sources.

The inputs to the anti- k_t algorithm are typically the topological calorimeter clusters introduced above, as hadrons are generally captured by the hadronic calorimeters. However the calorimeter clusters generally have poor spatial resolution, through great energy resolution. But since charged hadrons will also leave tracks in the inner detector, which have superior spatial resolution, one can combine the inner detector tracks with the calorimeter clusters to produce a much more accurate measurement. ATLAS does this via an algorithm called PFlow [58], and the PFlow objects replace the raw calorimeter clusters as inputs to the anti- k_t algorithm.

For fatjets, ATLAS uses an even more advanced algorithm called Unified Flow Object (UFO) [59]. The PFlow algorithm has degraded performance at the highest p_T 's because tracks with high p_T bend less in the solenoid field, so their momentum resolution is worse. However precise momentum determinations of the jet constituents is particularly necessary for fatjets because of the use of substructure algorithms. UFO remedies this issue by using PFlow as a base, but at higher jet p_T 's, switches to an alternative track-cluster combination algorithm that ignores the track momentums.

3.2.7.5 Missing Transverse Momentum

The final class of particles are neutrinos. These are very weakly interacting, and generally escape the detector without any traces. However, the experiment can infer their presence through momentum conservation. Because the protons collide head-on in the z axis, the total transverse momentum should be zero. Therefore the missing transverse momentum is defined as the inverse sum of all other detected particles. Note that z component of a neutrino cannot be inferred because, as mentioned in section 3.2.1, the initial z momentum is not zero. Also, multiple neutrinos cannot be resolved, only their sum. The magnitude of the missing transverse momentum is often referred to as MET and denoted via the symbol E_T^{miss} [60].

3.2.8 Simulation

A sophisticated simulation framework of ATLAS is also maintained to predict the detector's behavior [61]. MC simulated physics events are used as inputs. The ATLAS simulation then handles two main roles: simulating the interaction of the decay particles with the detector sensors and inactive material, and then the behavior of the front-end readout electronics and other firmware systems. The former relies heavily on the GEANT4 simulation toolkit [62]. The output of the simulation can be saved in the exact same format as real detector data, and passed through the same reconstruction and analysis pipelines. These MC simulations are extensively used for a variety of purposes, including optimizing algorithms, measuring performance, and estimating the SM or BSM yield of specific processes.

CHAPTER 4

Hardware-based Tracking for Trigger

Hardware-based Tracking for Trigger (HTT) was a planned TDAQ subsystem for the Phase-II upgrade of the ATLAS detector in preparation for the HL-LHC [63]. Given the much larger pileup of the HL-LHC, smarter triggers are needed to keep the readout data rate low while maintaining a similar level of acceptance. To this effect, HTT was intended to provide tracking information to a high-level trigger algorithm, the Event Filter, so that the latter could make more informed decisions. Strict latency requirements meant that HTT needed to rely on custom hardware components to do the tracking, hence the name. Unfortunately, a variety of upstream problems caused the project to be ultimately canceled in 2020. Nonetheless, developments on HTT are presented here to document the efforts undertaken and findings obtained.

This chapter first gives an overview of relevant parts of the ATLAS Phase-II upgrade in section 4.1, including the role of HTT in the new TDAQ system. Section 4.2 then gives a functional description of HTT. Finally section 4.3 discusses firmware developments for monitoring and control on the main HTT host board, while section 4.4 covers software developments and algorithm optimizations.

4.1 ATLAS Phase-II Upgrades

The ATLAS detector will undergo many extensive upgrades in preparation for the HL-LHC and its significantly higher pileup [64]. These are referred to as Phase-II upgrades, and will largely be commissioned in LS3, currently planned between 2026 and 2028 as shown in figure 3.4. A variety of detector components will be affected, but most prominently and of relevance to HTT are upgrades to the TDAQ system and a complete replacement of the inner detector.

4.1.1 TDAQ

A diagram of the baseline Phase-II TDAQ system is shown in figure 4.1. The L1 trigger from the Phase-I system, as described in section 3.2.6, is now called Level-0 (L0). The core components are largely the same, still using muon spectrometer and calorimeter measurements, though with several improvements such as the addition of the NSW. The output rate of the L0 trigger is relaxed to 1 MHz for a bandwidth of 5.2 TB/s. Events passing the L0 trigger are then passed to the Event Filter, which replaces the role of the HLT. Here

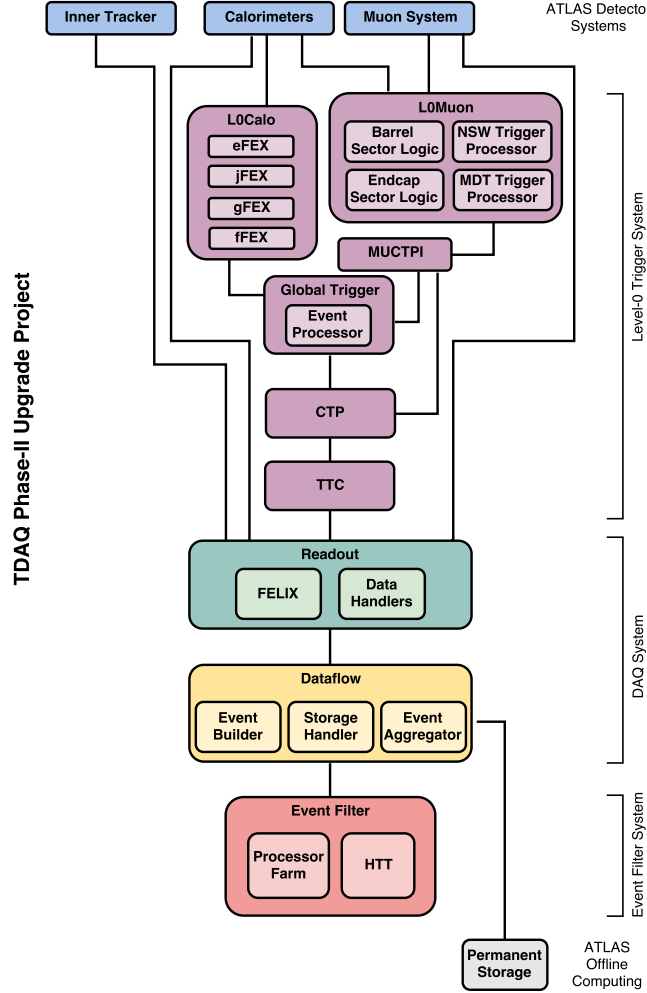


Figure 4.1: Diagram of the Phase-II TDAQ baseline system [63].

the processor farm is supplemented by the HTT as a coprocessor which provides accelerated track finding for tracks with $p_T > 1$ GeV. The Event Filter reduces the data rate to 10 kHz, for a bandwidth of around 60 GB/s saved to permanent storage.

The Event Filter uses the HTT in two distinct fashions, as seen in figure 4.2. First, the HTT takes the initial 1 MHz rate but only analyzes a portion of the detector data, around 10%, as informed by the results from the L0 trigger. This enables a fast initial rejection of background events to decrease the rate down to 400 kHz. This step is called the regional HTT (rHTT) because the tracking is only done in regions of interest. Then, the software processor farm will then takeover but occasionally, in around a quarter of events, call on the HTT for tracking information when necessary. This step is referred to as global HTT (gHTT), as the tracking is now done across the entire detector, and operates at 100 kHz.

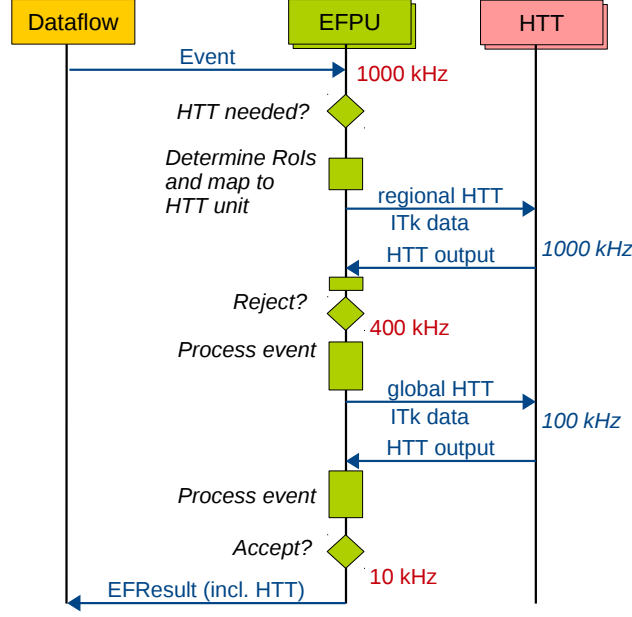


Figure 4.2: Diagram of the Event Filter system in the baseline scenario [63]. The EFP block represents an Event Filter processing unit.

4.1.1.1 Evolved Scenario

While the baseline system described above will be the initial configuration for Run-4, an evolution scenario was designed to mitigate potential risk factors that could overwhelm the readout system. The two primary risks are uncertainties in hadronic trigger rates and occupancy in the inner pixel layers. An underestimate of the former would lead to a surplus of events passing each trigger stage, while an underestimate the latter would lead to a larger event size than expected. These would cause the TDAQ system to exceed the various allotted bandwidth budgets, requiring undesirable stricter trigger cuts.

The evolution scenario addresses this risk by adding a new L1 trigger. The L0 trigger will now accept events at 4 MHz, but instead of having the entire detector data passed to the Event Filter, only regions of interest, up to 10% of the detector, will be passed to the L1 trigger. The L1 trigger will then accept events at a rate of 600 kHz, a reduction factor of around 7. To accomplish this accurately, the rHTT will be incorporated into the L1 trigger to provide tracking information. Note that the innermost pixel layers, which have the highest occupancy, will not be passed as part of the regions of interest. They will therefore only be readout at the lower 600 kHz rate, which greatly lessens the bandwidth compared to the baseline 1 MHz. Events passing the L1 trigger are then passed to the Event Filter, which now skips the rHTT step but still uses the gHTT.

Because the L1 trigger acts before events are readout to intermediate storage, it has a very strict latency requirement of $30\mu\text{s}$ per event. If a decision takes too long, an event's data will be overridden by subsequent events. This corresponds to a $10\mu\text{s}$ latency requirement for the rHTT processing step. This restriction was

a major motivation for using custom hardware for providing accelerated tracking to the Event Filter in the baseline scenario. Although the Event Filter has no latency requirement, other technologies could not be easily evolved to a L1 trigger if necessary.

4.1.2 Inner Tracker

During LS3 the entire inner detector will be replaced by a new Inner Tracker (ITk) system [65, 66]. This massive overhaul is motivated by several reasons, including higher occupancy at the HL-LHC and the need for faster readouts, as well as radiation damage accumulated by the current systems. Whereas the current inner detector uses three systems, pixels, strips, and the TRT, the ITk will be entirely silicon, with only pixels and strips.

A layout diagram is shown in figure 4.3. There are now five pixel layers in the barrel, spanning up to $r = 271$ mm. Notably the sensors are planned to be inclined near the edges of each layer, which minimizes both the amount of silicon needed and the amount of material that a particle travels through at the larger η ranges. The pixel end-caps are also no longer discrete disks, but rather a series of multiple rings. The rings are aligned in radius in four layers that match the outer four pixel barrel layers, but their z positions vary between the layers.

The strip sensors are less drastically changed from the current SCT. There will be four barrel layers and six end-cap disc layers. Note that like in the SCT, the strips are actually double-sided, offset slightly by a stereo angle to enable an approximate second coordinate measurement. In the barrel this angle is 52 mrad, while in the end-cap it is 40 mrad. In the context of HTT, each side of the strip sensors are considered a layer, so the barrel is considered to have eight layers.

4.2 Functional Description

The goal of the HTT is to take hits from the ITk and find tracks from them, that is a list of track parameters and which hits correspond to the track. It does this in a two-stage approach. The first stage does a rapid pattern matching and track fitting in eight outer ITk layers to find track candidates, while the second stage extends the track candidates to other layers to further cut down backgrounds and refine the track parameters. The second stage is only used in gHTT.

4.2.1 Pattern Matching

The first stage, which handles the bulk of the signal isolation, heavily relies on a custom application-specific integrated circuit (ASIC) called the Associative Memory (AM) chip [67] to do the computation within the latency requirements mentioned above. The AM chip functions by storing a pattern bank of around 500 thousand patterns, where patterns are sequences of eight 18-bit words. In HTT, these correspond to

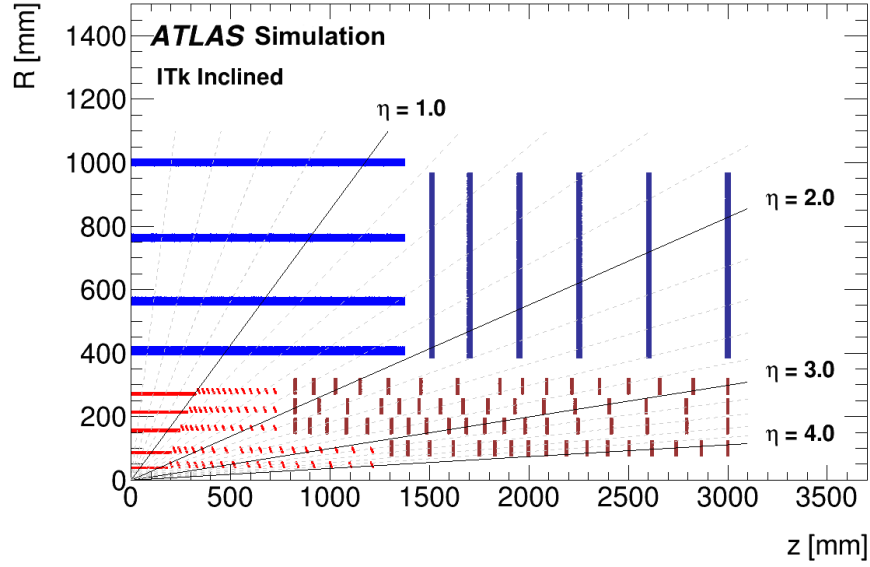


Figure 4.3: Potential layout of the ITk [65]. Pixels are shown in red and strips are shown in blue.

encoded hit locations in eight respective layers of the detector. Patterns are trained using simulations of particle tracks, such that each pattern corresponds to a track trajectory. Then given an event, for each input hit the AM chip does a rapid bit comparison with the pattern words in the same layer, across every pattern in its bank. After iterating through every input hit, the AM chip finds the patterns where all of its eight words are matched. These matched patterns are called *roads* and are the output of the AM chip. This process can be likened to a game of bingo, where the patterns are one-dimensional bingo cards and detector hits are the rolled numbers, though winners are chosen only after all hits have been processed. The AM chip is optimized to perform these bit comparisons rapidly in parallel across all of its stored patterns, generally within $2.5 \mu\text{s}$ per event.

Since the AM chip is hardwired to matching eight-word patterns, exactly eight detector layers must be used in the first stage. In the barrel, these were nominally chosen to be the last pixel layer, the inner side of the first strip layer, and both sides of the remaining three strip layers. More complex combinations are needed in higher η regimes. The encoding of detector hits into 18-bit integer words has more flexibility. Nearby strips or pixels are grouped together into *superstrips*, with each superstrip given a unique superstrip identifier (SSID). A hit is then encoded by the SSID it is contained within. The sizing of the superstrips trades-off precision with the number of patterns needed to fully account for all track trajectories. In the barrel these are nominally 40 strips and 33×402 pixels wide.

Even with superstrips greatly reducing the combinatorics of viable patterns, the size of the AM pattern banks are still a limiting factor. This limits the efficiency of the HTT track finding, meaning that not all tracks will be properly identified. Fortunately the AM chip has a few functions that enable greater efficiency

at the cost of increased fake rates. First, the AM chip accepts a variable threshold on the number of matched words needed to produce a road; instead of needing all eight words of a pattern to be matched, one can alternatively require only seven or six. The AM chip also allows the encoding of “don’t care” (DC) bits in the patterns; these are also referred to as ternary bits. These bits are ignored in the matching, or alternatively match both 0’s and 1’s. Effectively, this combines superstrips together, creating variable-sized superstrips on a per-pattern basis. The SSIDs are therefore Gray coded so that neighboring superstrips only differ by a single bit [68]. The AM chip also enables the use of *wildcards*, which are words that tell the pattern matching to completely ignore that layer, similar to a free space in the bingo analogy. These are reserved for patterns derived from track trajectories that do not actually go through all eight layers due to gaps in detector modules.

4.2.2 Track Fitting and Second Stage Extension

After the AM chip creates the roads, they are then passed to an FPGA which performs a track fit. Note that a road can have multiple hits per layer, so a track fit is conducted on every combination of one hit per layer. The fit is done using a simplified linear interpolation to keep the latency small,

$$p_i = \sum_{j=1}^N C_{ij} x_j + q_i \quad (4.1)$$

$$\chi^2 = \sum_{i=1}^{N-5} \left(\sum_{j=1}^N A_{ij} x_j + k_i \right)^2. \quad (4.2)$$

Here p_i are the five track parameters (p_T , η , ϕ , d_0 , and z_0), x_j are the hit coordinates, and C_{ij} , A_{ij} , q_i , and k_i are constants that are derived from simulation. χ^2 measures the goodness of the fit, and can be used to reject fakes. A different set of constants is derived for every viable combination of one detector module per each of the eight layers; these are called *sectors*. Note that when accepting patterns using a 6/8 or 7/8 threshold, or those with wildcards, separate constants are needed for each configuration of missing layers.

Tracks obtained from the first stage fit are passed through fake rejection filters and duplicate removal steps. Each passing track can then be optionally extended to the other ITk layers in a second stage fit. First hits that could correspond to the given track are collected using a sector-based lookup, then a similar linear fit is done for every combination of the new hits. The combination with the best χ^2 is then chosen. New constants need to be derived for every extended sector as well.

4.2.3 Hardware Organization

HTT utilizes a hardware and protocol standard called Advanced Telecommunications Computing Architecture (ATCA) [69]. The standard lays out specifications for high-reliability communications equipment, which is also pertinent for ATLAS as an enormous amount of data is being passed between different systems,



Figure 4.4: Top: an HTT demonstrator blade. Two TFM demonstrators are present, one with a heat sink installed, the other empty. The back of the blade is on the right, containing the connectors that insert into the shelf. The main processing FPGA is seen installed near the center, while the connectors for the management SoC are seen just above it. The IPMC is inserted into the long vertical black connector to the left of the SoC slot. Bottom: the blade inserted in to an ATCA crate.

and any downtime leads to lost event data. Using an industry standard means that the experiment can take advantage of well-tested components readily available in the general market.

The main component of an ATCA system are printed circuit board (PCB) *blades* inserted into *shelves* or *crates*, and each shelf can be installed into standard server racks. Pictures of a demonstrator HTT blade and an ATCA shelf are shown in figure 4.4. In HTT the blades are called Tracking Processors (TPs), which each have space for two mezzanine cards. Two types of mezzanine cards are used, the Pattern Recognition Mezzanine (PRM) and the Track-Fitting Mezzanine (TFM). When the TP is installed with two PRMs or two TFMs, it is known as an Associative Memory Tracking Processor (AMTP) or Second-Stage Tracking Processor (SSTP) respectively. As the names suggest, the AMTP with PRMs handles the first stage pattern matching while the SSTP with TFMs handles the second stage track extension. The TPs are organized into 96 independent units, each looking at an individual $\eta \times \phi$ slice of the detector. As seen in figure 4.5, each unit uses six AMTPs and one SSTP, though note the SSTPs are located in separate ATCA shelves. The units interface with the rest of the TDAQ system through dedicated HTT InterFace (HTTIF) cards.

The TP board itself acts like a motherboard, while the mezzanines implement the core algorithms.

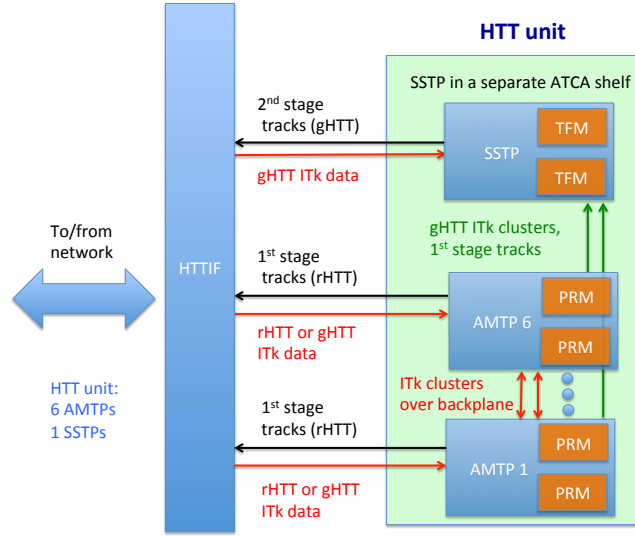


Figure 4.5: Diagram of an HTT unit [63].

Each TP uses a FPGA as the main processors, with responsibilities including I/O, data handling, and interfacing with the ATCA shelf. The TP also incorporates debug functionalities and data monitoring via a separate system on a chip (SoC). Meanwhile each PRM contains a single FPGA that manages 12 AM chips, corresponding to about 4.6 million patterns. The PRM FPGA handles I/O with the AM chips, as well as converting hit clusters to SSIDs and the first stage track fit. Finally the TFM uses two FPGAs that share the tasks of extrapolating hits from first stage tracks and fitting the extensions.

4.3 TP Firmware Developments

Each HTT TP board has two monitoring and control components, as shown in figure 4.6. The major component is a SoC that combines a processor and an FPGA, and is discussed in section 4.3.1. The second component is a shelf interface controller that handles communication between the TP and the ATCA shelf, and is discussed in section 4.3.2.

4.3.1 SoC

The SoC chosen for development testing was Xilinx's Zynq UltraScale+ MPSoC [70], which is hosted on an Enclustra Mercury XU5 module [71]. Model numbers are listed in table 4.1. The SoC combines a CPU with an FPGA into the same chip, referred to as the processing system (PS) and programmable logic (PL) layers respectively. By having both systems, users can take advantage of the respective advantages depending on the use case.

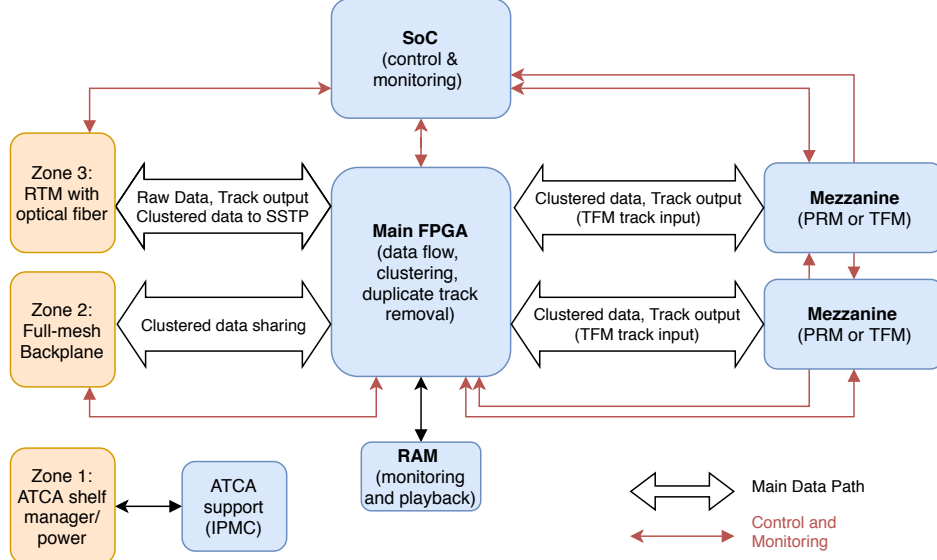


Figure 4.6: Flowchart of the TP board components, figure by Elliot Lipeles.

4.3.1.1 PS Developments: Booting Linux

For the TP board, the PS side was planned to primarily be used for power-up procedures and as a high-level debug interface. Therefore a principal task was to install a Linux system onto the SoC, which would enable a simple and familiar environment. Procedures could then easily be written as C programs and operated with basic shell processes, and the chip could be interfaced with via common network protocols like Secure Shell Protocol (SSH). Xilinx's PetaLinux [72] suite was chosen as the simplest solution, providing a ready-to-install Linux distribution with a relatively few command line instructions.

The Linux OS must be installed in non-volatile memory to survive power cycles. Fortunately the Mercury XU5 module comes with a 16 GB eMMC. However the eMMC cannot be loaded from off-module, and also needs to be configured first. Therefore Linux was first booted onto a temporary QSPI flash card also on the module. The eMMC can then be accessed using the flashed Linux instance and typical command-line utilities. In particular, fdisk [73] was used to create a bootable 128 MB FAT32 partition onto which the Linux distribution was installed. The remaining partition was formatted with ext4 for compatability with

Component	Number
Mercury XU5	ME-XU5-5EV-2I-D12E
Zynq UltraScale+ MPSoC	XCZU5EV-2SFVC784I
Cypress/Spansion QSPI (64MB)	S25FL512S
Kingston eMMC (16GB)	EMMC16G-W525-X01U

Table 4.1: Model numbers of the SoC module tested for the TP board

Linux.

With a Linux OS available on the SoC, C programs can be easily written for execution on the PS layer. Preliminary tests included setting up a power up handshake protocol with the TFM. This was slightly complicated by the TFM connection pins being on the PL side, so an AXI module was created to synchronize the PS and PL sides. Such a connection is also useful for monitoring and manipulating PL state from a Linux shell, such as for debugging purposes.

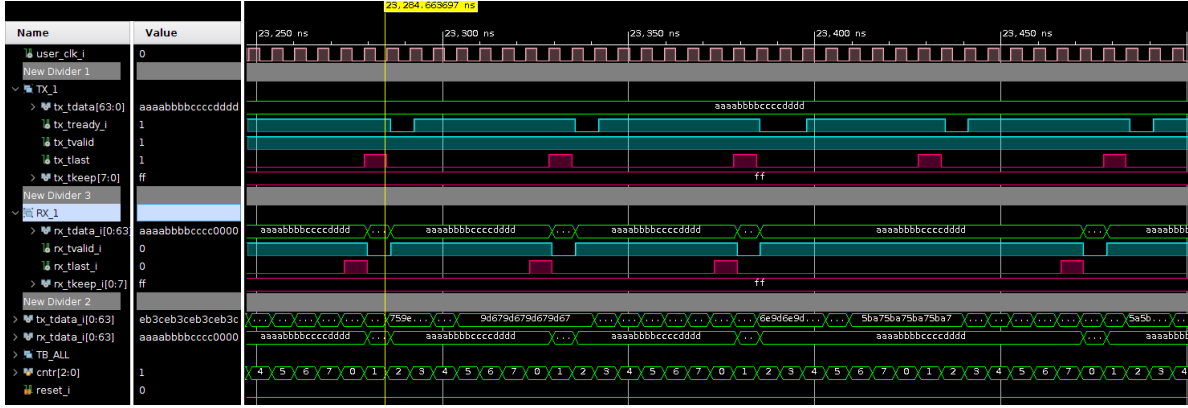
4.3.1.2 PL Developments: Gigabit Transceivers

The SoC contains several general purpose transceivers capable of transmitting data at around 16.3 Gb/s. These were intended to transfer monitoring data out of the TP. Xilinx provides an IP that implements the Aurora 64B/66B protocol [74], which was chosen for these links. Aurora transmits data in 64 bit blocks, which are encoded as 66 bits to enable clock recovery and synchronization. Blocks are grouped into arbitrary-length frames, which for example could represent the data from one detector module. Notably the ending block of a frame, called a SEP block, reserves 8 bits. This becomes noticeable if frame sizes are perfectly divisible by 64 bits. Then a completely empty SEP block must be passed to end the frame. For example, a 64 byte frame consists of 8 data blocks and 1 SEP block, leading to a 12.5% overhead. Meanwhile a 63 byte frame can fit in 7 data blocks and 1 SEP block, only a 1.6% overhead. Using longer frames would also decrease overhead. These factors were important constraints that informed the data formats explored.

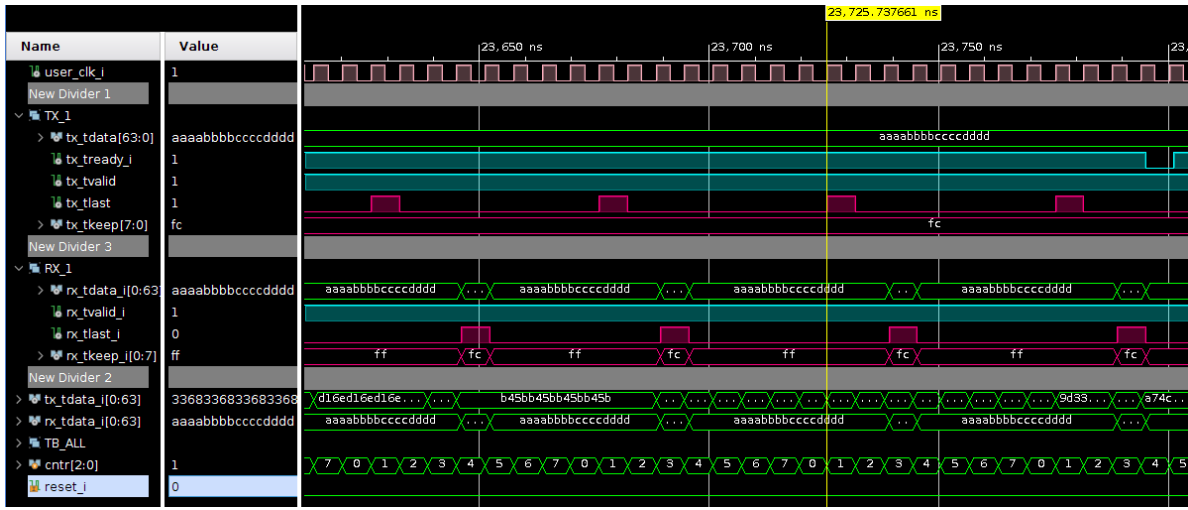
The Aurora 64B/66B IP was tested on the Zynq UltraScale via simulation. A loopback test was implemented, where one transceiver transmits a test signal to another transceiver on the same chip. Figure 4.7 shows the waveforms from these tests. The input data, seen in `tx_tdata`, was set to `0xaaaabbbbccccdddd` for every block. The line `tx_tlast` is raised high every 8 clock cycles, seen via `user_clk_i`, which marks the end of a frame. In figure 4.7a, a 64 byte frame is implemented by setting `tx_tkeep` to `0xff`, which means all 8 bits of the last block are valid data. Note that this causes `tx_tready_i`, the line used to indicate when data is ready to transmit, to go low on the subsequent clock cycle, because the IP needs to insert an empty SEP block. But in figure 4.7b, a 62 byte frame is implemented by setting `tx_tkeep` to `0xfc`. `tx_tready_i` no longer drops low every frame, reflecting the much lower overhead.

4.3.2 IPMC

ATCA uses the Intelligent Platform Management Interface (IPMI) specification for system monitoring and management [75]. It therefore mandates that each blade contain a dedicated IPMI Controller (IPMC) to handle IPMI communications, such as the message-passing protocol that happens when inserting blades into the shelf. CERN has customized their own IPMC [76], which is used on the TP boards. The card is inserted into the TP via the black connector noted in figure 4.4. Proper shelf interface behavior was validated using



(a) 64 byte frame



(b) 62 byte frame

Figure 4.7: Waveforms of a loopback test of the TP SoC's gigabit transceivers, using the Aurora 64B/66B protocol. See text for signal descriptions and interpretation.

bench-top test boards, where blade insertion and removal are mimicked by setting register flags.

In addition to handling the insertion and removal handshake, one of the major functions of an IPMC is to report board temperatures to the shelf manager. This enables the shelf manager to turn up the fans built into the shelf if a blade is getting too hot. The TP board incorporated two types of sensors, the ADT7420 temperature sensor [77] and the LTC2945 power sensor [78]. I2C drivers were written in C so that the IPMC could read and configure both sensors. The drivers for the ADT7420 in particular was tested by wiring a test board and reading the I2C responses on both a serial terminal and an oscilloscope as the temperature was varied.

4.4 Simulation and Pattern Matching Optimizations

A sophisticated simulation framework for HTT, named HTTSim, was also developed. The framework has several roles, including the generation of patterns and fit constants, measuring and optimizing performance of HTT, and of course providing the HTT functionality in the global ATLAS simulation when generating MC samples. HTTSim was implemented as a CMake package in ATLAS' offline software suite Athena [79], although due to the cancellation of HTT the code base was never merged.³ The framework was originally created by modifying the simulation code for FTK, a predecessor system to HTT that also used AM-based tracking but was also canceled. However, the code for the FTK simulation, called FTKSim,⁴ was found to be incredibly difficult to maintain so a complete rewrite was effected. Section 4.4.1 discusses the architecture implemented. However most of the HTT algorithm optimization, discussed in section 4.4.2, was done using the FTKSim-based software while the rewrite was being conducted. By the time the rewrite was ready, HTT had already been canceled. Fortunately, the simulation framework was able to be repurposed for the conceptual replacement to HTT, EF Tracking; this will be discussed further in chapter 5.

4.4.1 Framework Architecture

The framework consists of a series of sub-packages, shown in figure 4.8. The packages roughly form a linear dependency chain, reflecting the flow of data as summarized in figure 4.9. At the top level are a series of utility packages, defining common components used throughout the rest of HTTSim. These include type definitions and enumerations in HTTUtils, while HTTObjects contains classes representing the data elements that are passed within HTT, such as hits and roads. Note that HTTSim is not a bit-level simulation of the HTT firmware, but a functional one. Therefore the object classes are not constrained to the representation to be used in the firmware, allowing much more flexible class structures. A dedicated HTTPByteStream sub-package handled converting the software objects into firmware representations, as well as generating firmware test vectors. Finally HTTPConfig manages a python configuration system, used by all subsequent sub-packages, to keep track of the multitude of adjustable parameters and enable easy command-line interfacing.

4.4.1.1 Maps and Wrappers

The second layer of modules deals with the interface to the rest of ATLAS. The HTTPMaps module contains a series of mapping classes that provide the relationship between the detector geometry and the HTT hardware organization. The region map lists a set of modules that defines the $\eta \times \phi$ regions that each HTT unit oversees. Similarly the sector map defines the sectors by listing the modules in each sector. Meanwhile the plane map defines which hardware layers correspond to the logical layers used in pattern matching.

³While Athena is open source, the HTTSim fork was not made public. However the repurposed codebase for EF Tracking can be found in the **Trigger/EFTracking** package.

⁴Can be found in the **Trigger/TrigFTK** package of the 21.2 branch of Athena, but was deleted from the master branch.

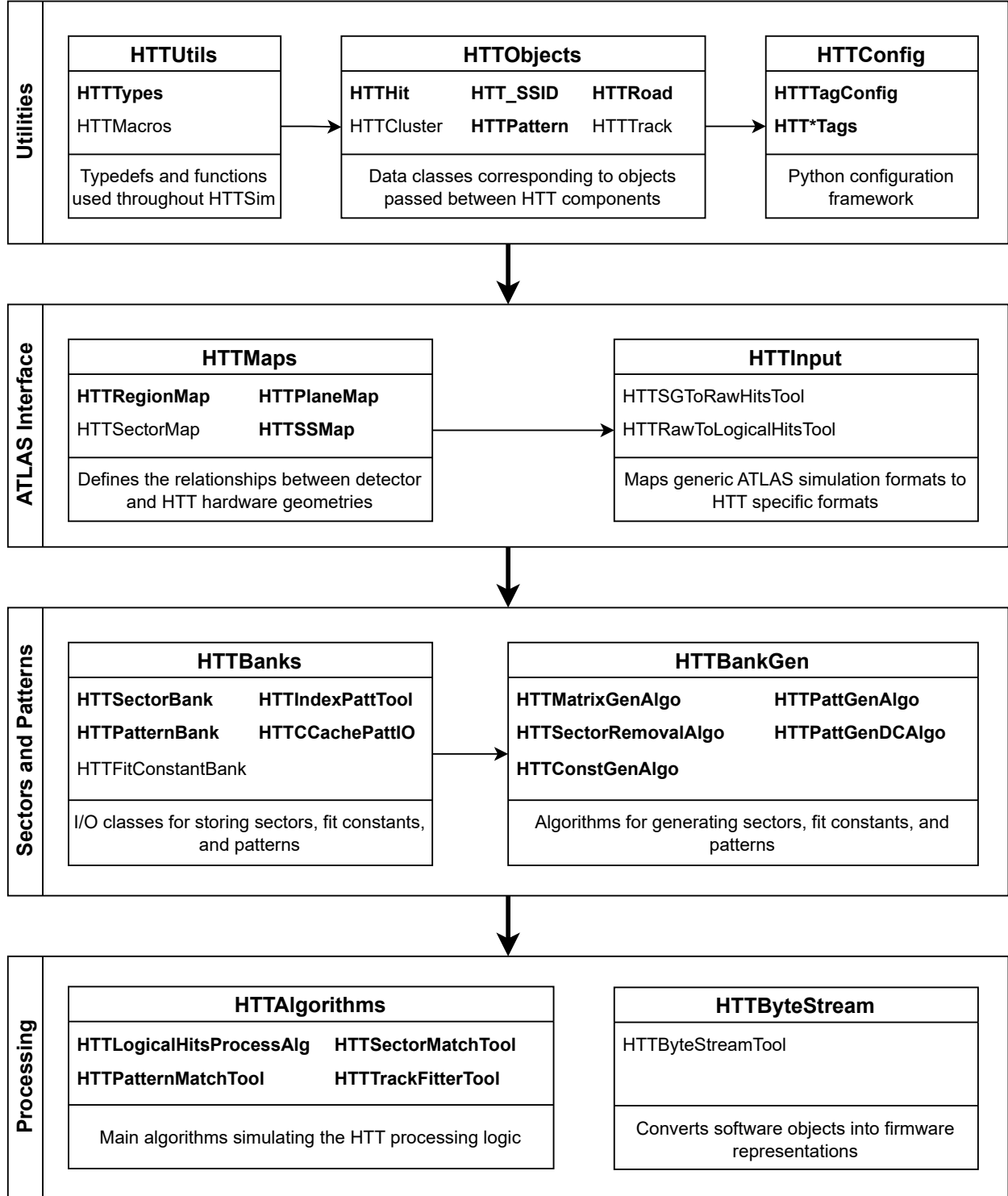


Figure 4.8: Block diagram of the HTTSim architecture. Each block corresponds to a sub-package, organized into four conceptual groups. The arrows roughly follow the dependency hierarchy of the packages. Principal classes are listed for each sub-package. Items in bold are those for which I was a major contributor.

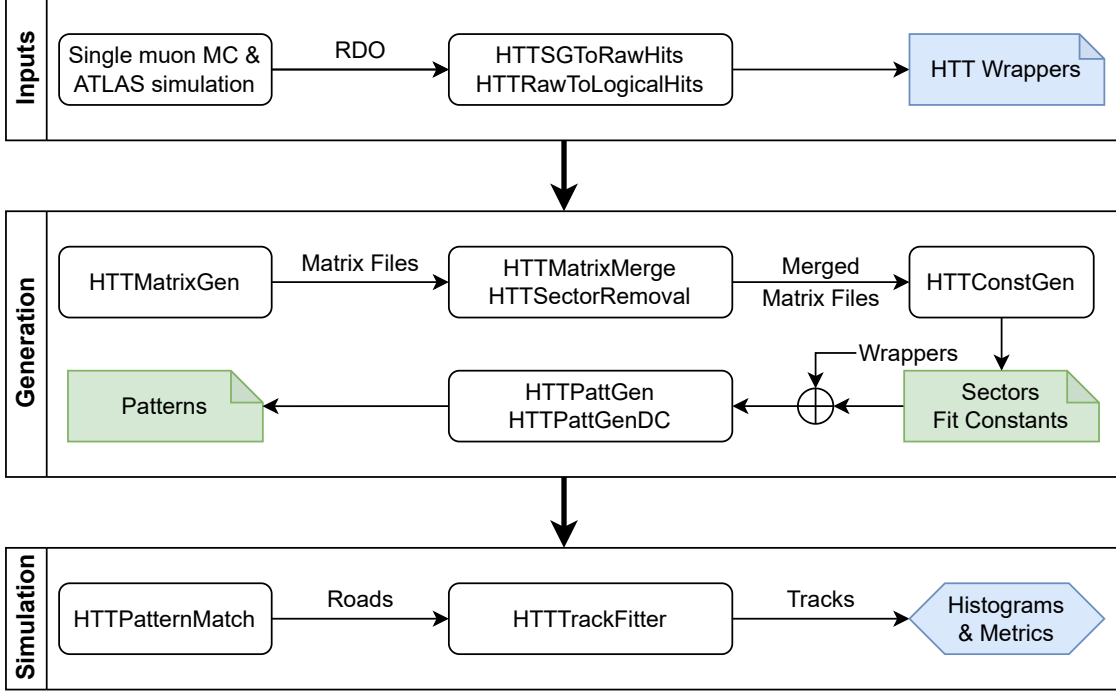


Figure 4.9: Flowchart of the HTTSim algorithms, grouped into three conceptual stages. The colored blocks correspond to the outputs of each stage, which are used in the following stage. In particular, the patterns, sectors, and fit constants, highlighted in green, are components to be used in the actual HTT firmware.

Finally a superstrip map encodes the sizing of superstrips. The mappings themselves are stored in plain text files as space-separated values. The region and sector maps are generated in the bank generation step, while the plane and superstrip maps are edited by hand. The C++ classes are used to read, write, and expose these files to the rest of the framework.

This setup keeps the hardware configuration incredibly flexible instead of being baked into the framework. For example, instead of hard coding the number of logical layers to eight everywhere in the framework, all components were forced to retrieve the value from the plane map class, which could be easily adjusted to use any number of layers. While such an indirection may have seemed unnecessary at first, as it was implausible that the AM chip would change the number of words per pattern, this modularity proved critical after the transition to EF Tracking, where there was no hardware restriction to eight layers and alternatives were able to be tested. Without a single source of truth in the plane map, changing the number of layers would have been an enormous and highly bug-prone endeavor.

Given defined maps, the HTTInput module then transforms generic ATLAS objects, Raw Data Objects (RDOs), into HTT specific ones. For example, hits are augmented with SSIDs and logical layer coordinates. These output files are termed HTT *wrappers*. Note that RDOs are generated through Athena, starting from MC simulated events and include the effects of showering, detector interactions, and readout behavior.

For HTT, wrappers containing single muon events are primarily used as the training tracks for sectors and patterns. However single electrons, pions, $t\bar{t}$, and other samples were also tested. Wrappers with the hard events overlaid on $\mu = 200$ pileup were also created to test dataflows. As the HTT units are generally independent, wrappers are created separately for each unit, and all subsequent steps are also done per unit.

4.4.1.2 Banks

After producing wrapper files, the sectors, fit constants, and patterns can be trained. These are elements that will be input to the firmware, and the collections are often referred to as *banks*. The HTTBanks package handles I/O of bank files and in-memory representations, while HTTBankGen generates the banks following the flowchart shown in figure 4.9. First, single muon wrappers are used to generate sectors and the fit constants used in equation (4.1). But because billions of training tracks are needed to create sectors and patterns, these are split into thousands of jobs whose outputs are then merged. Therefore the sectors are stored in intermediary files called *matrix files* which store accumulators that can be easily added or appended during the merging step. After merging, filtering, and duplicate removal algorithms are run, the matrix files are converted to the final sector definitions and fit constants. These are stored relatively simply; each sector contains only a list of modules and a set of matrices. In memory these are retrieved via hash maps with the sector ID as a key for $\mathcal{O}(1)$ access.

Pattern generation and storage is handled similarly, also using two formats, built off of formats used by FTKSim. The first, called the *indexed* format, was optimized for the pattern generation step. Again billions of training tracks are split into multiple jobs and merged. However, a major restriction in merging patterns is that the banks can exceed typical system memories, so the indexed format stores the patterns such that they can be read and merged sector by sector, which limits the amount of memory needed. The format also counts, for each pattern, the number of different training tracks that generated that pattern, called the *coverage*. Since bank sizes are a key limiting factor, patterns with higher coverage need to be prioritized, as they can match more cases and increase efficiency. Patterns in each sector are therefore also sorted by decreasing coverage to enable quick filtering during and after the merges. Note that HTTSim also has an alternative method of generating patterns aside from using MC wrappers as the training tracks. This method is called track inversion, where fake tracks are created by randomly sampling track parameters in each sector and inverting the linear fit to get hit coordinates. Track inversion is much faster than creating wrapper files, but potentially loses accuracy.

Patterns generated in the above step do not utilize DC bits yet, and are called thin-strip (TSP) patterns. A dedicated algorithm therefore combines TSP patterns together to create the final DC patterns. It initializes an empty bank of DC patterns then does a single loop through the TSP patterns in decreasing coverage order. For each TSP pattern, the algorithm first checks if that pattern can be merged with any existing DC

pattern in the same sector by the addition of DC bits. If not, it adds the TSP pattern unmodified to the DC bank. Various configuration parameters control when and how a TSP pattern can be merged with a DC pattern, such as limiting the total number of DC bits per pattern. This step then saves the DC bank in a different format called CCache. This format is optimized for the matching step; instead of storing a list of SSIDs for each pattern, it inverts the relationship and stores a list of pattern IDs for each SSID and sector. Thus given a hit, one can find the corresponding pattern IDs in $\mathcal{O}(1)$ time. Note that patterns with DC bits are encoded by appearing in the entries for multiples SSIDs in a layer. Note also this means that getting the actual pattern, that is the list of SSIDs, from a given pattern ID is highly non-trivial in the CCache format, requiring essentially a full scan through the pattern bank. However this is not necessary for the pattern matching step, described further in the next section.

4.4.1.3 Algorithms

The final package HTTAlgorithms contains the actual HTT simulation functionalities, primarily the pattern matching and track fitting. The pattern matching is the most time-intensive component of the simulation, as for every event, tens of thousands of hits need to be matched to around 4 million patterns, the approximate capacity of one HTT unit. As mentioned in the previous section, the patterns are already stored in an SSID-based index, such that given a hit, all corresponding pattern IDs can be found in $\mathcal{O}(1)$ time. A map-based accumulator keyed by the pattern ID stores the corresponding hits for each pattern. Note the map avoids storing an hit vector for all 4 million patterns, equating to an $\mathcal{O}(p \cdot x)$ space complexity, where p is the average number of patterns with at least one hit in an event and x is the average number of hits per pattern, typically around 10. The hits are processed in $\mathcal{O}(h \cdot y)$, where h is the average number of hits per event and y is the average number of patterns per SSID, usually less than 10. Note p and h can vary wildly depending on the event being processed. A final loop through the accumulator checks which patterns have the required number of hit layers in $\mathcal{O}(p \cdot x)$ time to create the roads. This loop references a lookup table keyed by pattern ID to check for wildcards, which decreases the threshold for the respective patterns. The roads are then passed to the track fitter, which fits each combination of hits per layer and then applies a simple χ^2 cut.

The pattern matching and track fitting are implemented as independent tools that are managed by a single main algorithm that handles the event loop and takes the wrapper files, sectors, and pattern banks as inputs. The output of the algorithm is a final list of tracks, but it also records a plethora of diagnostic histograms after each step to measure efficiency, dataflow, and other metrics. Thanks to the HTTConfig package, the configuration of all components can be easily adjusted with command line options. This includes filepaths to the input banks, value parameters such as the χ^2 cut of the track fit, as well the slotting of alternative or additional tools in the processing chain. For example, the pattern matching tool could easily

be replaced by a sector matching tool, which both implement a common RoadFinder interface. The sector matching tool effectively treats each sector as a single ultra-wide pattern, and was used frequently to validate sector generation prior to performing pattern generation. In such instances the track fit is meaningless, but it also can be omitted with a single flag.

4.4.2 Pattern Matching Optimizations

In evaluating the performance of HTT, two key metrics are considered. The first is efficiency, the fraction of truth tracks matched correctly, where a variety of matching criteria can be used. Typically efficiency is tested using single muon MC events, in which case finding any road or track is considered a success. With other inputs, especially those including pileup, a truth track is considered found if it shares more than half of its hits with a HTT output road or track. Note efficiency can be measured separately at both the road and track stages, and by these measures the road efficiency will always be higher than the track one. The second metric is the dataflow, or the average amount of roads and tracks generated, and the number of track fits needed, per event. If too many roads or tracks are being found, the hardware will not be able to finish processing within the latency requirements. This number also corresponds roughly with the fake rate, which can affect trigger rates and Event Filter performance. Dataflow is typically measured using single muon MC events overlaid on $\mu = 200$ pileup. Other metrics are also measured, such as the resolution of track parameters after the fit.

A benchmark configuration for HTT was developed in prior unpublished studies, consisting of the superstrip and layer choices detailed previously, as well as a 7/8 threshold in the pattern matching. But importantly the configuration used during the DC pattern generation step had yet to be thoroughly optimized. Therefore several setups were tested in a central barrel region of $0.1 < \eta < 0.3$, summarized in figure 4.10. The DC bits were limited to the last n bits in each pattern word, where n varies in each layer. A baseline configuration was chosen with 2 allowed bits in the pixel layer and the last two strip layers, and 1 in all other layers. This is denoted as "21 11 11 22" based on the layers in inside-out order, and shown as the blue circle in figure 4.10. Alternative bits per layer allotments were tested to see if the road rate could be reduced while maintaining efficiency, shown in the brown squares. Similarly a limit on the total number of DC bits per pattern was also tested, with the threshold scanned via the orange squares. And finally only considering TSP patterns with some minimum coverage is shown in the green triangle series. The latter two series use the baseline per-layer limit as well.

The max DC bit limit per pattern was found to be most effective, able to cut the number of roads by 25% with less than half a percentage point loss in efficiency. Meanwhile, requiring a minimum coverage performed rather poorly, which motivated moving in the opposite direction. During the DC generation step, instead of halting the production once the pattern bank was full, the algorithm was allowed to continue

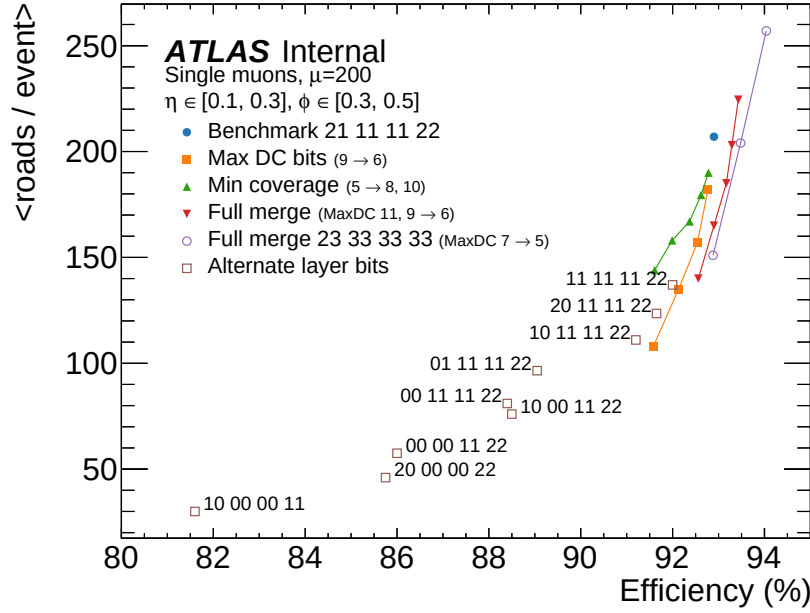


Figure 4.10: HTT pattern matching performance with various DC bit configurations. See text for details on the different series. The number ranges in the legend give the values of the parameter being scanned from highest efficiency to lowest. The numbers next to the brown squares are the respective DC bit layer configurations. No wildcards are included, and patterns are generated from track inversion.

attempting to merge in low coverage TSP patterns using DC bits, down to coverage = 2 patterns inclusive. However, the coverage = 2 patterns were uniquely only allowed to merge with patterns without DC bits yet, to prevent exploding the number of DC bits used. This new method was combined with a limit on the total number of DC bits per pattern, and is shown in the red triangle series of figure 4.10. Finally, the number of allowed DC bits per layer was increased to 3 in all strip layers, shown in the purple circle series. This final configuration was able to reduce road rates by 25% without losing any efficiency, compared to the baseline.

Quite noticeably though, the tests in figure 4.10 all have efficiency below 94%, which is far from ideal. Much effort was exerted to improve the efficiency, and two major causes were discovered. Most importantly, wildcards had yet to be implemented in the simulation, so some efficiency loss was expected. This was believed to be small, but it turned out that a sizeable portion of the training tracks do not cross all eight layers. After implementing wildcards, efficiency drastically increased, though so too did the number of roads. The second problem was that the track inversion method of pattern generation, which all the studies in figure 4.10 used, turned out to be rather inefficient at covering edge cases. While efficiency could be recovered by generating more samples, the method now loses its advantage of saving computation time compared to using MC muons directly. Figure 4.11 now summarizes updated studies using MC-generated patterns with wildcards. These use the final DC bit algorithm mentioned earlier, but still scan the max

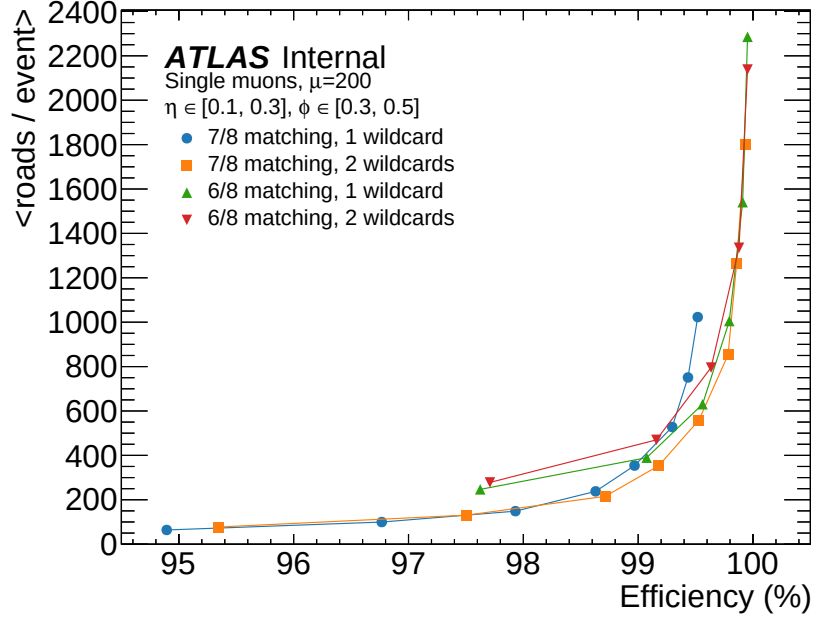


Figure 4.11: HTT pattern matching performance with various wildcards and matching criteria. Each series scans the max number of DC bits allowed per pattern. Patterns are generated directly from truth muons.

number of DC bits per pattern threshold. The number of wildcards allowed per sector is varied across the series along with the matching threshold, the number of matched words per pattern needed to create a road. Given the results in figure 4.11, a suitable working point can then be chosen depending on the dataflow restrictions from firmware.

CHAPTER 5

Hough Transforms

ATLAS held its 95th triannual Collaboration Board meeting on February 28th, 2020, with the threat of the COVID-19 pandemic looming large. A key topic in discussion was the Phase-II upgrade, as several components were already behind schedule or at severe risk of falling behind. In particular, the ITk pixel upgrade was determined to be unable to support the 4 MHz readout rate required by the evolved TDAQ scenario outlined in section 4.1.1.1. A series of remedies were considered, such as downsizing the L1 trigger rate or adding a regional readout functionality to the pixel front-end chips. But by the 97th Collaboration Board meeting on October 16th, ATLAS had decided to cancel the evolution scenario entirely, weighing the physics benefits versus the technical risk. However, without the strict latency requirement of the L1 trigger, hardware-based tracking is no longer strictly necessary. Therefore a new baseline for tracking in the Event Filter needed to be developed, and three task forces were created to investigate different approaches:

1. Custom: Use custom hardware similar to HTT but re-optimized for a L0-only scenario.
2. Heterogeneous Commodity: Use commercially available FPGAs on PCIe cards as accelerators.
3. Software: Use CPU clusters only.

This chapter discusses studies conducted on behalf of the Heterogeneous Commodity task force. These focus on the use of a Hough transform as the main tracking algorithm, which is a highly efficient method that can be processed rapidly on FPGAs. All task forces submitted their reports in June 2021, and were published together in as an amendment to the 2018 TDAQ Technical Design Report (TDR) [80] .

5.1 Hough Transform Algorithm

The Hough transform is a technique originally developed for imaging and computer vision [81, 82], but has been previously applied to tracking in high energy physics [63, 83, 84]. It can be derived as follows. The ATLAS solenoid provides an approximately constant 2T field in the z direction throughout the inner detector. Charged particles will therefore follow approximately circular trajectories in the transverse plane. By setting the Lorentz force F_L equal to the centripetal force F_C , we find a relationship between the radius

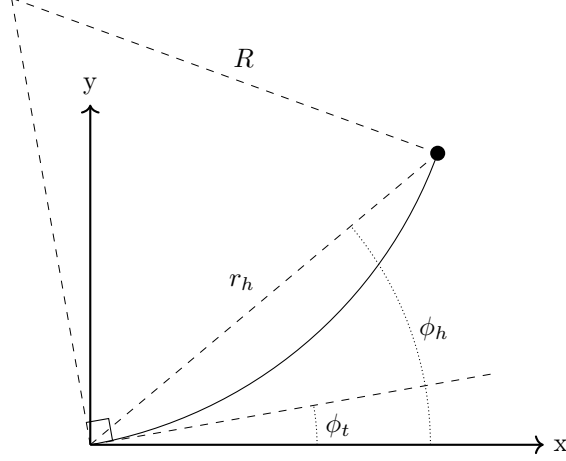


Figure 5.1: Diagram of an idealized track in the Hough transform. The track has a radius of curvature of R and is shown as the solid curve, while the point represents a single hit at (r_h, ϕ_h) .

of the circle R and the particle's transverse momentum,

$$F_L \equiv |q|v_TB = p_T v_T / R \equiv F_C \quad (5.1)$$

$$p_T = |q|BR. \quad (5.2)$$

We also assume that the track originates from the origin with a tangential azimuthal angle of ϕ_t , as seen in figure 5.1. Then note that the center of the circular trajectory is at an azimuthal angle of $\phi_t \pm \pi/2$ depending on which direction the track bends. By the definition of a circle, we find that any point on the trajectory, given in polar coordinates as (r_h, ϕ_h) , satisfies

$$R^2 = (r_h \cos \phi_h - R \cos(\phi_t \pm \pi/2))^2 + (r_h \sin \phi_h - R \sin(\phi_t \pm \pi/2))^2 \quad (5.3)$$

$$R^2 = (r_h \cos \phi_h \pm R \sin \phi_t)^2 + (r_h \sin \phi_h \mp R \cos \phi_t)^2 \quad (5.4)$$

$$R^2 = r_h^2 + R^2 \pm 2r_h R \cos \phi_h \sin \phi_t \mp 2r_h R \sin \phi_h \cos \phi_t \quad (5.5)$$

$$0 = r_h \pm 2R(\cos \phi_h \sin \phi_t - \sin \phi_h \cos \phi_t) \quad (5.6)$$

$$0 = r_h \pm 2R \sin(\phi_t - \phi_h). \quad (5.7)$$

Note negative charges curl counterclockwise, so have the positive sign in equation (5.7). Then substituting equation (5.2) for R gives

$$\frac{\sin(\phi_t - \phi_h)}{r_h} = \frac{qB}{2p_T}. \quad (5.8)$$

In practice, we know the hit coordinates r_h and ϕ_h , and want to find the track parameters ϕ_t and q/p_T . Equation (5.8) then gives a series of all possible track parameters consistent with each hit. Note that $\phi_t \approx \phi_h$, so equation (5.8) is approximately a line in $q/p_T - \phi_t$ space. For hits from the same track, their set of consistent track parameters will intersect precisely at the track's true parameters. This is shown

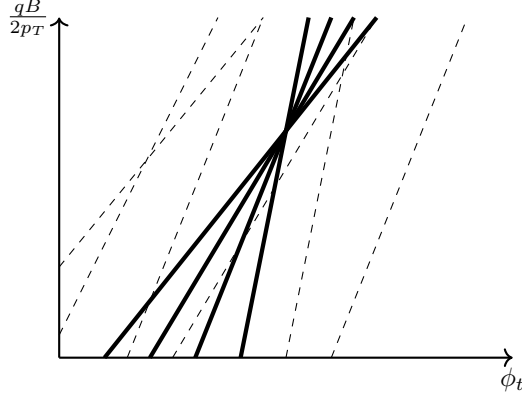


Figure 5.2: Schematic example of using a Hough transform to find tracks. Each line corresponds to a single hit, given by equation (5.8). Hits from the same track, shown in bold, will intersect at a common point. The track parameters ϕ_t and q/p_T can then be read off the intersection coordinate. Background hits are shown as the dashed lines, which can produce spurious intersections.

schematically in figure 5.2. Therefore the tracking algorithm can proceed as follows: create a binned $q/p_T - \phi_t$ space that stores a list of hits. Then for each hit in a region of interest, find the corresponding bins using equation (5.8), and append the hit to the bins' accumulators. After every hit is processed, simply iterate through the bins and select those where multiple hits are present. The corresponding hits can be taken as track candidates, and a preliminary q/p_T and ϕ_t can be read off the bin values. Note that background hits can also accidentally intersect, so a threshold criteria as well as a subsequent fake rejection method are needed to filter these out. Detector resolutions as well as the reliance on approximating assumptions in the above derivation will also cause track hits to not intersect perfectly, so some smearing effect must be applied. Various procedures were tested and are described in the following sections.

5.1.1 1D Hough

In addition to the method outlined above, termed the 2D-Hough transform, a second method was invented by Elliot Lipeles that uses a simpler 1D accumulator. This saves significant space in the eventual FPGA implementation, and is called the bit-shift method or 1D-Hough transform. The method assumes that hits are organized in layers similar to HTT, and that each layer is at a constant radius r . Then for a given layer, all hits satisfy the same relationship

$$\phi_t - \phi_h = \frac{r_i B}{2} \frac{q}{p_T} \quad (5.9)$$

in the small angle approximation. That is, each value of q/p_T corresponds to a uniform offset in ϕ across the layer. Therefore a 1D binned accumulator in ϕ_h space is created for each layer, forming effectively a bitmask of hit positions. Then q/p_T is scanned, where at each step the corresponding shift in ϕ is applied to all bitmasks. The shifted bitmasks now encode values of ϕ_t consistent to the given q/p_T . If multiple hits

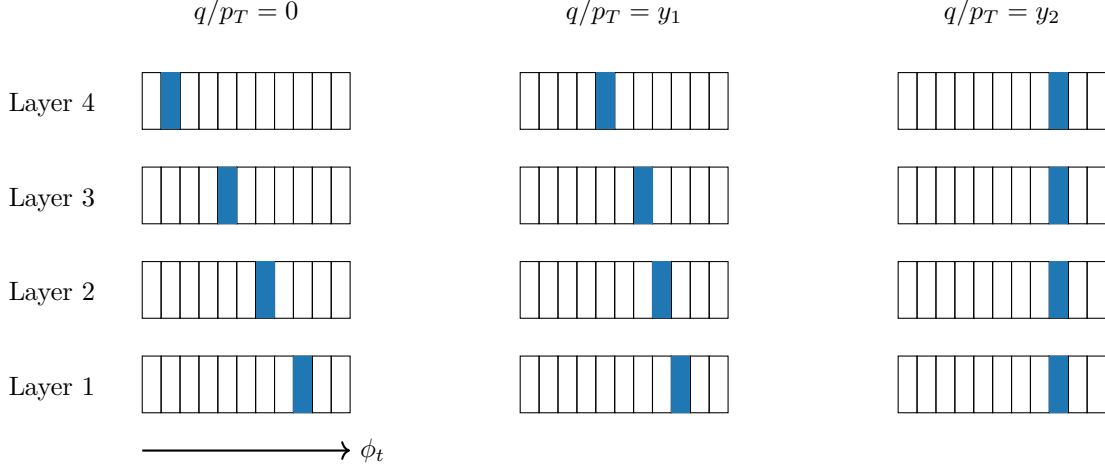


Figure 5.3: Schematic example of the 1D-Hough transform. Here four layers are split into ten ϕ bins. Hits are represented by the blue bins. The bins are filled using ϕ_h , corresponding also to infinite p_T . As q/p_T is scanned to two other values, the hit bins are shifted to the right. In the third segment, all four layers have a bit set in the eighth bin, and a track candidate is declared found.

coincide at the same bit across several layers, these are taken to be a track candidate. An example of this procedure is shown in figure 5.3. Note that since the shift amounts are monotonic in r , the minimum step size in q/p_T is given by a ϕ shift of exactly one bin in the outermost layer,

$$\Delta\left(\frac{q}{p_T}\right) = \frac{2}{\max_i(r_i)B} \Delta\phi, \quad (5.10)$$

where $\Delta\phi$ is the bin width.

5.1.2 Slicing

So far the z coordinate of the hits have been completely ignored, as well as the z_0 of the track. This would lead to an exploding data rate since hits across the detector would be effectively piled on top of each other. Therefore, like HTT, the Hough transform is split into units that target small regions of track ϕ and η . In addition, z -slicing is used to further remove redundancy. A key layer is chosen and divided along z into consecutive segments, split along module boundaries. Corresponding sub-regions are created by taking the envelope of all tracks that pass through each respective segment, within the ϕ and η region limits. An example is shown in figure 5.4. An independent Hough accumulator is used for each sub-region, only reading hits from relevant modules. By using a key layer instead of z_0 (i.e. the beamline) for the slicing, the sub-regions are smaller and intersect less in the layers of interest. Note that slicing preserves efficiency and reduces dataflows at the cost of increased FPGA resource consumption.

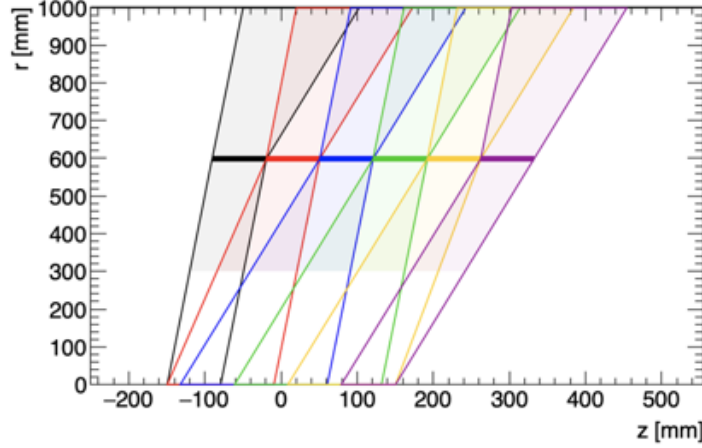


Figure 5.4: Example of key-layer z -slicing for the Hough transform, figure by Elliot Lipeles. The key layer is chosen to be the second physical strip layer, shown as the thick bars at $r \approx 600\text{mm}$. It is split into six segments denoted by the colors, and the lines show the sub-region boundaries, or the range of viable tracks that pass through each segment.

5.2 Optimization Studies

The Hough transform was optimized in the same manner as the pattern matching in HTT. Thanks to the modularity of HTTSim, switching to Hough transforms was rather simple. The bank I/O and generation steps are simply skipped, and banks are no longer input to the final processing algorithm. Meanwhile, since the pattern matching component was already compartmentalized through an agnostic RoadFinder interface, it could be trivially replaced with new Hough transform tools implemented to follow the same interface. Thus, despite a drastic shift in hardware, no major changes were required to the HTTSim architecture. The entire transition was limited just by the implementation of the Hough algorithms, with the 2D and 1D versions as separate tools, taking less than two weeks. As with HTT, the main metrics used to evaluate the Hough transform are efficiency and dataflow. Initially, the Hough transform was also tested using the same eight layers as the pattern matching, also focusing on the region $0.1 < \eta < 0.3$ and $0.3 < \phi < 0.5$.

5.2.1 Binning and Smearing

The primary parameters in both the 2D and 1D Hough algorithms are the binning and the smearing used, both of which account for the imprecision of the detector as well as the idealistic assumptions used in the Hough transform derivation. Coarser binning and larger smearing tend to increase both efficiency and dataflow. A summary of tests is shown in table 5.1. The binning was originally chosen to be quite fine, motivated by previous works [63, 83, 84]. These did not perform too well, having relatively low efficiencies, and would also drastically increase the resource consumption on FPGAs. Coarser binning values, chosen somewhat arbitrarily, were therefore favored and tested more thoroughly. Note that the

binning reported in table 5.1 for the 2D-Hough is the number of bins spanning the nominal $0.3 < \phi < 0.5$ and $-1 \text{ GeV}^{-1} < q/p_T < 1 \text{ GeV}^{-1}$ ranges. However, it was discovered that the ranges of the accumulator needed to be broadened slightly to account for resolution effects. The true ranges are *padded* with extra bins in each direction, determined empirically to fully include all truth tracks in the nominal range. The amount of padding is typically no more than 6% on each side. Meanwhile for the 1D-Hough, because the accumulators store ϕ_h and not ϕ_t , a range of $0 < \phi_h < 0.8$ is used.

The smearing was implemented by accumulating each hit to extra ϕ bins for each q/p_T bin. That is, if for a given q/p_T bin equation (5.8) equates to ϕ bins in the range $[i, j]$, then the hit would be added to the accumulator in the range $[i - x, j + x]$ instead. This is termed a *hit extension* of x bins. The size of the extension was chosen separately for each layer. Note this can be applied equivalently in the 1D version, where a hit typically is only added to a single bin per layer, call i , but can be extended to fill a range $[i - x, i + x]$. Similarly to HTT, there is also a threshold parameter, the minimum number of hit layers in the same bin required to create a road. Thresholds of 7/8 were typically used, providing already good efficiency; 6/8 was briefly tested but expectedly produced an untenable number of roads. In addition to the baseline 8 layers used in HTT, the Hough transform was also tested using 5 layers, the last pixel layer and the inner side of all strip layers, and 9 layers, the last pixel layer and both sides of all strip layers. Finally, note that all values in table 5.1 used the 6 z -slices as shown in figure 5.4; the results provided are for the entire region after aggregating the individual slices.

The 2D-Hough transform generally performs quite well, easily reaching above 99% efficiency. The 1D-Hough tended to perform more poorly, struggling to get equal efficiency without sacrificing the dataflow; however, in these studies the 1D version was much less studied and optimized, and significant improvements were later made by other analyzers. Meanwhile, the dataflow naively appears quite high, but note that no duplicate removal had been applied yet. Even on single muon wrappers with no pileup, the 2D transform would find on average around 5 roads. A simple window-based rejection was therefore implemented based on recommendations by Sasha Paramonov: before creating a road from a bin, the algorithm checks all the bin's neighbors in a $n \times n$ grid; if any neighbor has more hits, or has equal hits but is a lower p_T , the road is not created. This prioritizes the bins with the most hits, with the latter choice acting as a tie breaker. Preliminary tests demonstrated reductions in the number of roads by roughly a factor of three, with no loss in efficiency. Subsequent analyzers were then able to reduce the final number of tracks to a few dozen by adding a neural network fake rejection; see reference [80] for details.

5.2.2 Magnetic Field Correction

The above studies all focused on a central barrel region, $0.1 < \eta < 0.3$. The Hough transform was also tested in $0.7 < \eta < 0.9$, and similar performances were found. However, in a much more forward region

2D-Hough							
Bins ϕ	Bins q/p_T	Layers	Thresh.	Extension	Eff. (%)	<Roads>	<Combs.>
1200	216	8	7	-	91.3	123	640
1200	216	8	7	[2, 0*7]	92.1	228	-
1200	216	8	7	[38, 0*7]	97.1	569	-
1200	216	8	7	[12, 6, 0*6]	98.2	819	3181
216	216	8	7	[2, 1, 0*6]	98.8	361	1892
216	216	8	6	-	99.71	1091	5449
216	216	8	6	[1, 0*7]	99.81	1602	9727
216	216	8	6	[2, 0*7]	99.83	1938	14107
216	216	8	6	[1, 1, 0*6]	99.84	2258	15519
216	216	8	6	[2, 1, 0*6]	99.86	2722	22445
100	216	8	7	[2, 0*7]	98.5	751	8287
100	216	8	7	[2, 1, 0*6]	99.0	1172	11696
64	216	8	7	-	99.0	906	22103
64	216	8	7	[2, 0*7]	99.0	1406	27550
64	216	8	6	-	99.9	-	-
64	64	8	7	-	99.0	2604	-
64	64	8	7	[2, 0*7]	99.0	3358	-
64	64	8	6	-	99.9	-	-
1200	216	5	5	-	61.6	245	-
1200	216	5	4	-	96.3	5667	-
64	216	5	4	-	99.7	6799	-
64	64	5	4	-	99.9	-	-
64	216	9	8	[2, 0*8]	98.5	795	23125
216	216	9	7	-	99.12	401	2481
216	216	9	7	[1, 0*8]	99.47	583	4412
216	216	9	7	[2, 0*8]	99.55	702	6417
216	216	9	7	[1, 1, 1, 0*6]	99.65	1175	11467
216	216	9	7	[2, 1, 1, 0*6]	99.71	1410	16554

1D-Hough							
Bins ϕ	Bins q/p_T	Layers	Thresh.	Extension	Eff. (%)	<Roads>	<Combs.>
200	-	8	7	0	60.6	326	4787
200	-	8	7	[0, 0.5*7]	95.7	2031	119001
200	-	8	7	0.5	98.2	2621	211874
200	-	8	7	[0, 1*7]	98.7	3977	786823
200	-	8	7	1	99.1	5418	2M
400	-	8	7	[0, 1*7]	96.3	1487	37060
400	-	8	7	1	97.7	3187	-
400	-	8	7	2	98.8	10690	-
1200	-	8	7	2	92.3	-	-

Table 5.1: Summary of Hough transform performance with various configurations. The rightmost three columns are the output metrics, while the other columns are adjusted parameters. Note the hit extensions are lists by layer from inside-out, or a common value for all layers. The rightmost column is the number of hit combinations per event, i.e. the number of track fits needed. Dashes in the output columns represent values that were not measured.

Bins ϕ	Bins q/p_T	Extension	B -corr.	Eff. (%) per p_T [GeV]			<Roads>	<Combs.>
				[1-2]	[2-4]	[4+]		
64	128	[1, 1, 0*6]	-	87.0	98.0	99.5	286	13k
64	128	[1, 1, 0*6]	✓	97.3	98.9	99.5	280	10k

Table 5.2: Effect of the magnetic field correction on efficiency and dataflow. The first row is without the correction and the second row is with the correction. The efficiency is binned in p_T of the truth track.

of $2.0 < \eta < 2.2$, the efficiency degraded considerably for low p_T tracks. It was speculated that the loss in efficiency was due to the non-uniformity in the ATLAS solenoid's magnetic field. As seen in figure 3.15, the field achieves the nominal 2T in the z direction in the barrel region, but tapers considerably closer to the end-caps, where it also begins to develop transverse components at larger radii. However, the Hough transform assumes the nominal field throughout the detector.

To verify this hypothesis, equation (5.8) was inverted to calculate the expected ϕ_h in each layer given a known ϕ_t , q , and p_T from an MC truth particle. The difference with the ϕ position of the actual hit, $\Delta\phi$, can then be plotted in histograms. These are shown in figure 5.5 for both the central barrel region and the forward region using single muon samples with fixed p_T . In the central region, the $\Delta\phi$ is consistently small in all layers, indicating that the Hough transform is accurate. Meanwhile in the forward region a clear bimodal split occurs, with the sign of $\Delta\phi$ depending on the charge of the particle. The deviation becomes more significant at larger radii and smaller p_T , confirming the effect of the non-uniform magnetic field.

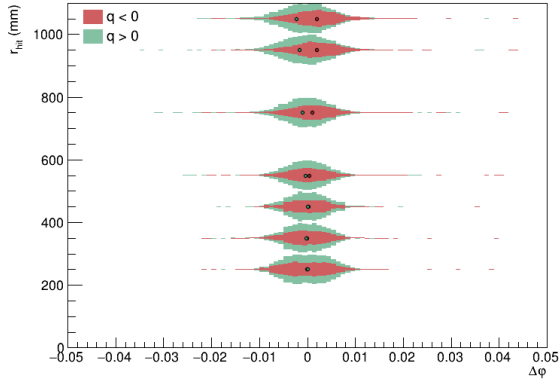
Correcting equation (5.8) for a non-uniform field would be difficult via first principles, especially considering that the magnetic field map itself is not analytically derived, and only measured at three radii. An analytic form would likely also depend on variables beyond ϕ_t and q/p_T , which would make the tracking method much more difficult. Instead, an empirical correction was attempted by reversing the observed $\Delta\phi$ obtained from the simulated tracks. The $\Delta\phi$ was assumed to be directly proportional to q/p_T and a function of r_h ,

$$\Delta\phi = f(r_h) \frac{q}{p_T}. \quad (5.11)$$

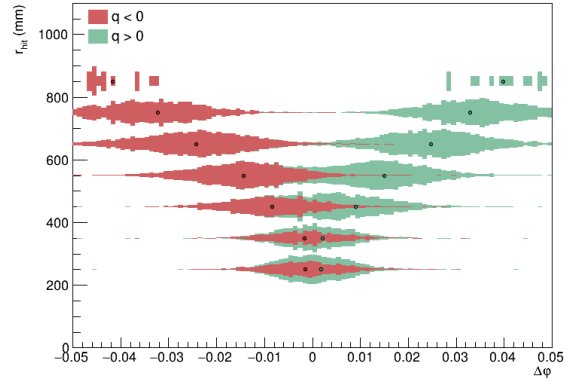
The function f was then derived by plotting the mean $\Delta\phi \cdot p_T/q$ in each bin of r_h , as shown in figure 5.6. By observation, a quadratic function was fitted to this distribution, achieving a coefficient of determination of 0.994. Equation (5.8) can then be written with the correction term added on as

$$\phi_t = \phi_h + \arcsin\left(\frac{Br_h}{2} \frac{q}{p_T}\right) + f(r_h) \frac{q}{p_T}, \quad (5.12)$$

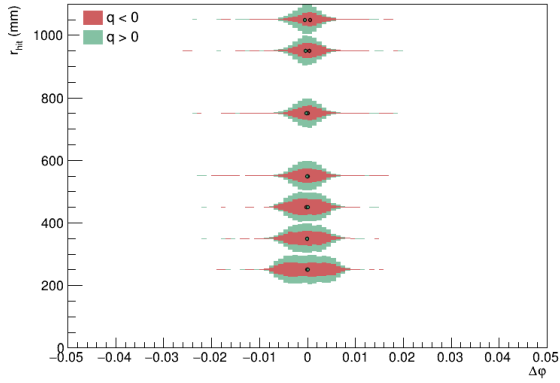
which retains ϕ_t as a simple analytic function of q/p_T and the hit coordinates. The effect of the correction can be seen in figure 5.7, where the $\Delta\phi$ deviations are now clustered around zero. This successfully restored the efficiency in the forward region, as summarized in table 5.2.



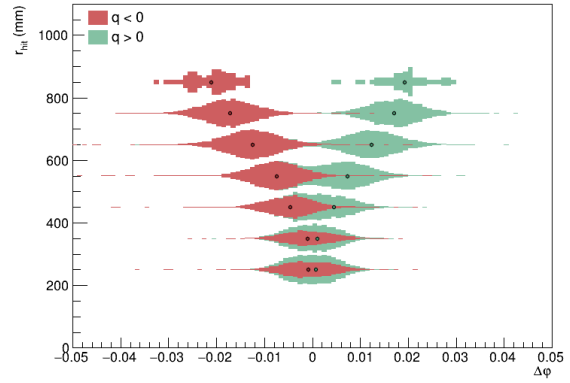
(a) $\eta \in [0.1, 0.3]$, $p_T = 1$ GeV



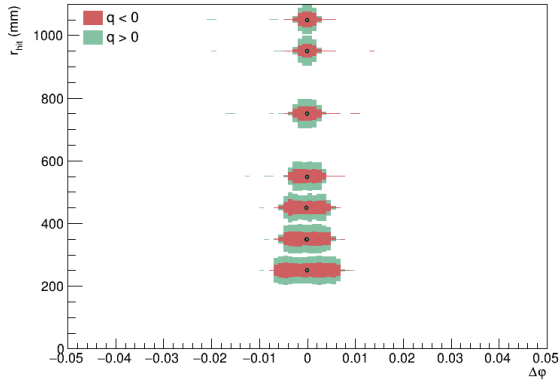
(b) $\eta \in [2.0, 2.2]$, $p_T = 1$ GeV



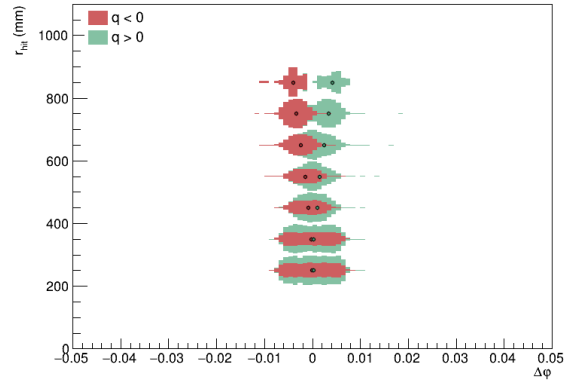
(c) $\eta \in [0.1, 0.3]$, $p_T = 2$ GeV



(d) $\eta \in [2.0, 2.2]$, $p_T = 2$ GeV



(e) $\eta \in [0.1, 0.3]$, $p_T = 10$ GeV



(f) $\eta \in [2.0, 2.2]$, $p_T = 10$ GeV

Figure 5.5: Differences between the actual and expected ϕ_h by inverting equation (5.8) using truth track parameters. These plots are violin plots, where double-sided histograms are plotted for each 100mm bin in r_h . The histograms are normalized to their maximum element. Hits from negatively charged tracks are shown in red, while those from positively charged tracks are stacked on top in green. The small circles show the means of the respective distributions.

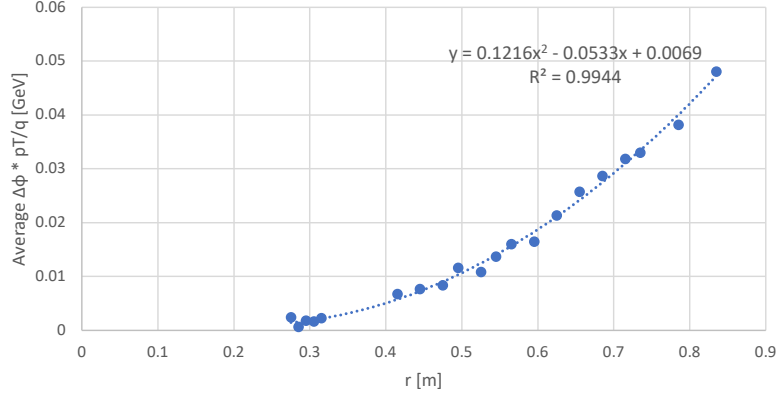


Figure 5.6: Mean $\Delta\phi \cdot p_T/q$ in bins of r_h . A quadratic function is readily fitted to the series, validating the form of equation (5.11).

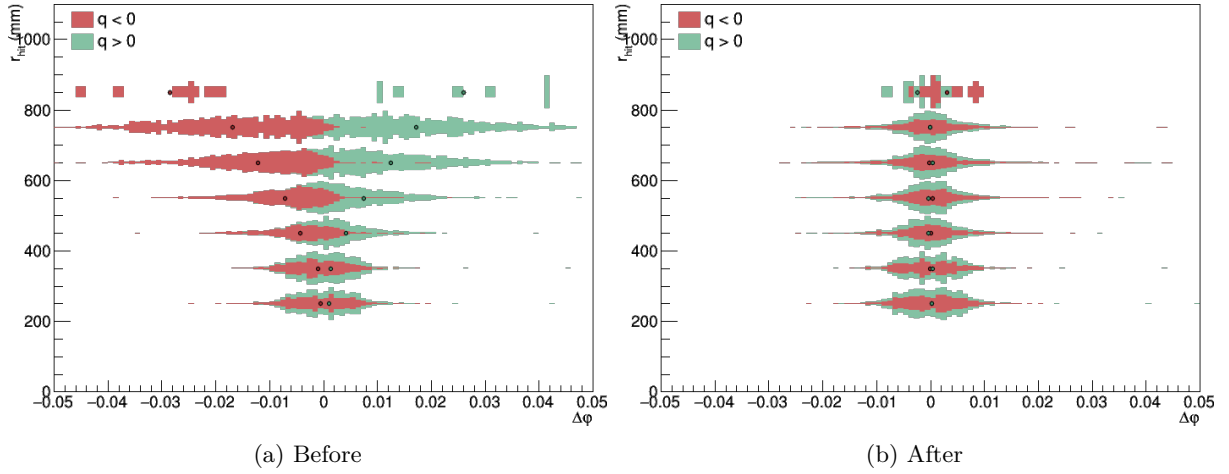


Figure 5.7: Violin plots of $\Delta\phi$ (a) before and (b) after the correction in equation (5.12) is applied in the forward region. See figure 5.5 for an explanation of the plots. Note these figures used single muons that were flat in $q/p_T \in [-1, 1]$ GeV^{-1} instead of having a fixed p_T .

CHAPTER 6

Analysis Overview

This chapter presents an overview of the $VV+VH$ Semileptonic analysis [85]. It targets the production of two bosons in the semileptonic decay channel, where a W/Z decays to leptons and a $W/Z/H$ decays to quarks, as seen in figure 6.1. This analysis is the second round and combination of separate VV semileptonic [86] and VH semileptonic [87] resonance searches. Both this analysis and the first round analyses use the full ATLAS Run-2 dataset, corresponding to an integrated luminosity of 139 fb^{-1} and a collision center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. However, this second round combination introduces many improvements, including the harmonization of the separate analyses, use of newer jet finding and tagging algorithms, and machine learning event classification. In addition, the second round includes an unfolded measurement and SMEFT interpretation of the VV process,⁵ which is the main focus of this dissertation. The resonance search and the unfolded measurement have rather distinct motivations and analysis techniques, so this part will prioritize the latter.

6.1 Motivations

As mentioned in chapter 2, SMEFT provides a systematic framework to explore BSM physics in a predominantly model-independent manner. Precisely measuring SMEFT parameters helps constrain the space of BSM theories. At the leading order of interest, dimension-6, there are 59 operator coefficients that need to be measured. Note the operators affect different physics processes, so a wide array of measurements are

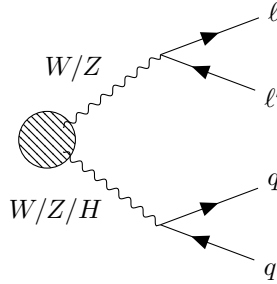


Figure 6.1: Feynman diagram of the decay signature for the $VV+VH$ Semileptonic analysis.

⁵The unfolded measurement excludes the VH process, as that has been covered by another group [88].

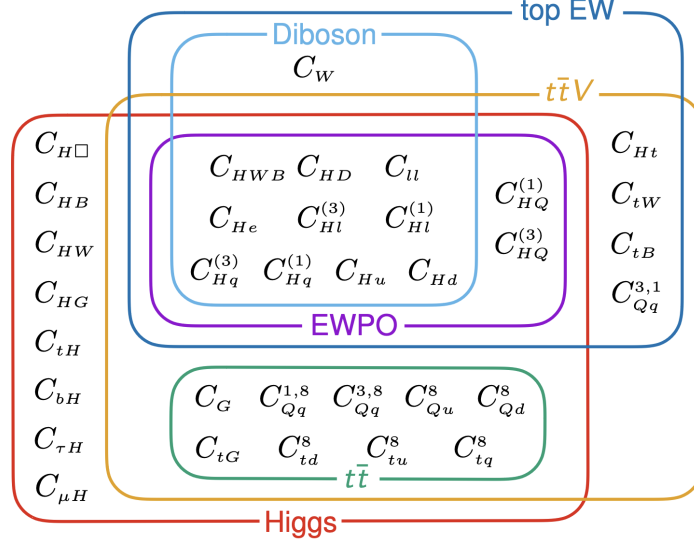


Figure 6.2: A subset of SMEFT operators organized by the processes that they primarily affect, from [89].

needed to constrain all operators. The diboson process provides sensitivity to the operators circled in the sky-blue box of figure 6.2. Notably the diboson process has relatively isolated sensitivity to the c_W operator, which affects the triple and quartic gauge couplings in figures 2.3b and 2.3c, and so this operator was heavily prioritized in studies.

Both ATLAS and CMS have conducted multiple measurements for diboson production, summarized in table 6.1. The majority of these measurements focus on fully leptonic decay channels, which are typically preferred for their cleaner event signatures and minimal backgrounds. However, the leptonic branching ratios of the vector bosons are much smaller than the hadronic ones, as seen in table 6.2. Therefore a semileptonic analysis gains access to significantly more data, which increases the energy reach of the measurement. As SMEFT operators tend to scale with E or E^2 , a semileptonic analysis can then provide a much more precise constraint of the operator coefficients [90]. The trade-off is having to deal with a much larger background, predominantly from V +jets. Modeling this background is the focus of chapter 7. Note that while a fully hadronic analysis would have the highest branching ratio, the fully jet backgrounds become incredibly difficult to deal with. In this sense the semileptonic channel strikes an ideal balance between signal and background strengths. Note also that ATLAS did perform a semileptonic measurement in 2017, but it only used data from 2012, with 20.2 fb^{-1} of collisions at 8 TeV [91]. Significant improvements are expected with the entire Run-2 dataset at 13 TeV, as well as the development of many new tagging, modeling, and reconstruction techniques.

	Process	Channel	Note	$\mathcal{L}$ [fb $^{-1}$]	Reference
ATLAS		$WW \rightarrow e\mu$		140	[92]
		$WZ \rightarrow 3\ell$		139	[93]
		$ZZ \rightarrow 4\ell$		139	[94]
		$WW + j \rightarrow e\mu$	Interference enhanced	139	[95]
		$Z + 2j \rightarrow 2\ell$	Vector boson fusion	139	[96]
		$Z + \gamma\gamma \rightarrow 2\ell$	Dim-8 aQGC limits	139	[97]
		$ZZ + 2j \rightarrow 4\ell$	Dim-8 aQGC limits, VBS	139	[98]
		$W^\pm W^\pm + 2j \rightarrow 2\ell$	Dim-8 aQGC limits, VBS	139	[99]
		$W\gamma + 2j \rightarrow \ell$	Dim-8 aQGC limits, VBS	139	-
		$VV + 2j \rightarrow \text{semilep}$	Dim-8 aQGC limits, VBS	139	-
CMS		$WW \rightarrow 2\ell$		36	[100]
		$WZ \rightarrow 3\ell$		138	[101]
		$W\gamma \rightarrow \ell$	Interference resurrection	138	[102]
		$Z\gamma + 2j \rightarrow 2\ell$	Dim-8 aQGC limits, VBS	36	[103]
		$W\gamma + 2j \rightarrow \ell$	Dim-8 aQGC limits, VBS	138	[104]
		$W^\pm W^\pm + 2j \rightarrow 2\ell$	Dim-8 aQGC limits, VBS	137	[105]
		$WZ + 2j \rightarrow 3\ell$	Dim-8 aQGC limits, VBS	137	[105]
		$ZZ + 2j \rightarrow 4\ell$	Dim-8 aQGC limits, VBS	137	[106]
		$VV + 2j \rightarrow \text{semilep}$	Dim-8 aQGC limits, VBS	36	[107]

Table 6.1: Recent ATLAS and CMS measurements in diboson channels. Not necessarily comprehensive. The process column lists the boson types in consideration along with associated productions. Notably, the vector boson scattering (VBS) measurements target anomalous quartic gauge couplings (aQGCs), which are present only in dimension-8 terms and above. The last two analyses in the ATLAS section have yet to be published.

Mode	Branching Ratio (%)	Mode	Branching Ratio (%)
$e\nu$	10.71 ± 0.16	e^+e^-	3.3632 ± 0.0042
$\mu\nu$	10.63 ± 0.15	$\mu^+\mu^-$	3.3662 ± 0.0066
$\tau\nu$	11.38 ± 0.21	$\tau^+\tau^-$	3.3696 ± 0.0083
hadrons	67.41 ± 0.27	invisible	20.000 ± 0.055
		hadrons	69.911 ± 0.056

(a) W boson
(b) Z boson

Table 6.2: W and Z branching ratios for each decay mode [10].

6.2 Object Selection

Physics objects are reconstructed using the algorithms mentioned in section 3.2.7. The analysis chooses working points and imposes additional requirements to maximize sensitivity to the target signal processes. For electrons and muons, the analysis creates two selection definitions called Loose and Tight, derived from the general ATLAS quality working points. A summary of the most important selection cuts are shown in table 6.3. For jets, the analysis uses the typical ATLAS small-R and large-R jets with additional kinematic cuts, summarized in table 6.4. These use the PFlow and UFO inputs respectively. Additionally, the analysis utilizes the DL1r b -tagger with the 77% efficiency working point to help identify jets originating from b quarks [108].

Feature		Loose	Tight
Electrons	Quality	Loose	Tight
	η	< 2.47	
	p_T	$> 7 \text{ GeV}$	$> 30 \text{ GeV}$
	$ d_0 /\sigma_{d_0}$	< 5	
	$ z_0 \sin \theta $	$< 0.5 \text{ mm}$	
Muons	Quality	$p_T < 300 \text{ GeV}$ $p_T \geq 300 \text{ GeV}$	Loose Medium High- p_T
	η	< 2.5	
	p_T	$> 7 \text{ GeV}$	$> 30 \text{ GeV}$
	$ d_0 /\sigma_{d_0}$	< 3	
	$ z_0 \sin \theta $	$< 0.5 \text{ mm}$	

Table 6.3: Primary selections for the electron and muon definitions used in the analysis. The analysis defines separate Loose and Tight selections that should not be confused with the generic ATLAS quality working points, which are precursors and listed in the first rows of each block.

Feature	Small-R	Large-R
Anti- k_T R	0.4	1.0
$ \eta $	< 4.5	< 2.0
p_T	$> 20 \text{ GeV}$ for $ \eta < 2.5$ $> 30 \text{ GeV}$ for $ \eta > 2.5$	$> 200 \text{ GeV}$
m	-	$> 50 \text{ GeV}$

Table 6.4: Primary jet selections used in the analysis.

6.3 Event Selection and Classification

Six main analysis channels are defined based on the reconstruction method for each boson leg. For the leptonically decaying boson, denoted V_ℓ , three channels are created based on the number of leptons in the event, covering all the cases $V_\ell = Z \rightarrow \nu\nu$, $V_\ell = W \rightarrow \ell\nu$, and $V_\ell = Z \rightarrow \ell\ell$. These are referred to as the 0-, 1-, and 2-lepton channels respectively. Meanwhile for the hadronically decaying boson, denoted V_h , two strategies are used: the two quarks can be reconstructed individually in separate small-R jets, or together as a single large-R jet. The latter is more effective at higher p_T because the two sub-jets become much more collimated. These are referred to as the resolved and merged channels respectively. Note however that the unfolding side of the analysis only uses the three merged channels and ignores the resolved ones.

In the resonance side of the analysis, these channels are further split to increase sensitivity to specific BSM processes. It first uses a recurrent neural network to isolate events originating from the VBS production mode, and then a deep neural network to orthogonalize the VV and VH analysis regions. The VBS classifier was developed for the first round VV analysis [86], while the orthogonality classifier was newly developed for the second round given the merging of the two signatures. The resonance side also splits out b -enhanced regions based on the number of b -tagged jets. Finally it also divides the VV merged region into high purity (HP) and low purity (LP) regions for events that respectively pass or fail a substructure tagger, discussed in

0-lepton	1-lepton	2-lepton
Lepton Selection		
0 “Loose” leptons	1 “Tight” lepton and 0 “Loose” leptons	2 opp.-sign same-flavor “Loose” leptons
$E_T^{\text{miss}} > 200 \text{ GeV}$ $E_T^{\text{miss}} \text{ sig.} > 10$	$p_T(\ell) > 30 \text{ GeV}$ $E_T^{\text{miss}} > 100 \text{ GeV}$ $p_T(W) > 200 \text{ GeV}$	Leading ℓ $p_T > 27 \text{ GeV}$ Subleading ℓ $p_T > 25 \text{ GeV}$ $83 < m_{ee} < 99 \text{ GeV}$ $m(\ell\ell)$ $m_{\mu\mu} > 85.6 - 0.0117p_T(\ell\ell) \text{ GeV}$ $m_{\mu\mu} < 94.0 + 0.0185p_T(\ell\ell) \text{ GeV}$
Jet Selection		
Leading large- R jet		
$p_T(J) > 300 \text{ GeV}$ $72 \text{ GeV} < m(J) < 102 \text{ GeV}$ Pass W/Z substructure tagger		
Topology Requirements		
$\Delta\phi(\text{jets}, E_T^{\text{miss}}) > \frac{\pi}{9}$	$E_T^{\text{miss}}/p_T(W) > 0.2$ (electron-only)	-
-	No additional b -jets	-
-	$p_T(V)/m(VV) > 0.35$	

Table 6.5: Event selection cuts in the unfolding signal region channels.

more detail in chapter 7. Meanwhile the unfolding side of the analysis does not isolate VBS or b -enhanced regions, and does not need the orthogonalization since only the VV process is considered. It also only considers HP events. The unfolding channels are therefore referred to as the “Merged HP Inclusive” regions in the context of the rest of the analysis, where the inclusive term refers to the agnosticism to b -tags and VBS.

A series of kinematic cuts are then applied to isolate the diboson signal from the primary backgrounds: V +jets, $t\bar{t}$, and single top production. A summary of the cuts in the three unfolding regions are shown in table 6.5, while tables of the resonance-side selections are shown in appendix B. Note that the unfolding selections mirror the resonance VV merged selections as closely as possible, with two major exceptions. The substructure tagger was switched to an adversarial neural network (ANN) tagger instead of the three-variable tagger used in the resonance side, which also required increasing the jet p_T cutoff to $p_T(J) > 300 \text{ GeV}$ instead of 200 GeV . Since the ANN tagger does not include a mass cut like the three-variable tagger, a cut of $m(J) \in [72, 102] \text{ GeV}$ is used instead. Details of this switch are discussed in chapter 7.

In addition to the signal regions defined in table 6.5, the analysis uses two types of control regions to constrain backgrounds remaining after the cuts. The top control region (TCR) is defined in the 1-lepton channel by mirroring the signal region selections but inverts the additional b -jet veto. It is very pure in $t\bar{t}$ and single top, which contribute 98% of the total yield. Note that a 2-lepton TCR was also defined by inverting the same-flavor lepton requirement; however it contains barely any events so it was not used. The mass-sideband control region (MCR) is similarly defined by instead inverting the $m(V_h)$ cut, now for all

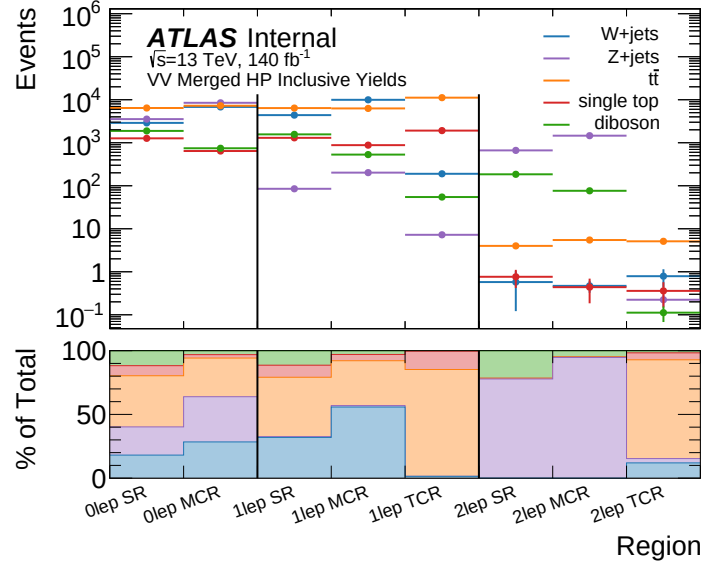


Figure 6.3: Nominal MC yields from signal and background events in the eight regions used in the unfolding side of the analysis.

lepton channels. This region is primarily V+jets, though receives significant contributions from $t\bar{t}$ in the 0- and 1-lepton channels. A summary of the yields of the different processes in each unfolding region is shown in figure 6.3.

CHAPTER 7

Fitting the V+jets Background

The main background signature in this analysis is the production of a leptonically-decaying W or Z boson in association with QCD jets. Collectively, these events are referred to as V+jets. The QCD jets can mimic a hadronically-decaying boson, creating a reducible background in the diboson signal region. The selections listed in table 6.5 help reduce this background significantly, especially the W/Z tagger, but V+jets still contributes around 40% of the total Standard Model yield in the signal region, based on MC simulations.

However, MC simulations are relatively inaccurate at modeling QCD events. Previous studies by ATLAS have shown that MC modeling of V+jets can have up to a 40% systematic error in similar phase spaces [109, 110]. These errors are generally dominated by uncertainties in the QCD scale parameters, and confirmed to be of equal order in this analysis. Relying only on MC to model this background would thus severely reduce sensitivity. This motivated the use of a data-driven approach to instead model the V+jets background.

The crux of the data-driven approach is the expectation that V+jets produces a generally smooth distribution in the jet mass $m(J)$, whereas true diboson decays will have a sharp resonance around the W and Z masses, as seen in figure 7.1. Therefore the signal region is defined by a window cut on $m(J)$, and the MCR is defined by inverting this window but keeping all other selections. This should make the MCR a region dominated by V+jets with relatively little signal contamination, but still kinematically similar to the signal region. The contribution of V+jets in the signal region can then be inferred by interpolating a smooth function between the sidebands into the signal region. Note that the MCR is also called the *sidebands*, referring to the upper and lower mass intervals surrounding the signal region window.

In this chapter, section 7.1 covers initial attempts at fitting the V+jets sideband using a functional form, and discusses the choice of W/Z taggers. Then section 7.2 introduces Gaussian process regression as an alternative fitting method. Since the mathematics can be quite confusing, section 7.3 uses toy examples to build intuition on Gaussian process. Finally section 7.4 presents studies on using the Gaussian process to fit the sidebands.

7.1 Functional Fits

Initial attempts at fitting the V+jets sideband used a functional fit, where a fixed functional form f for the $m(J)$ distribution is assumed. The parameters p of the function are optimized by minimizing the χ^2 (the

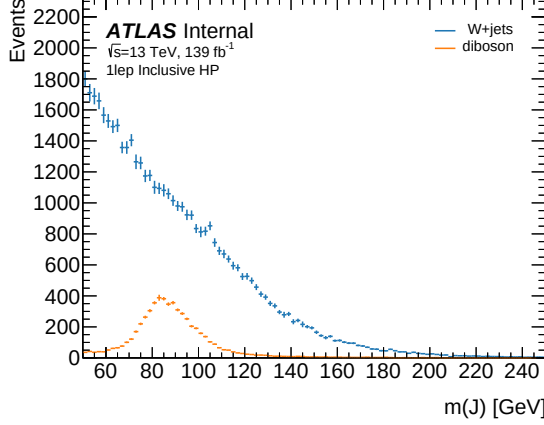


Figure 7.1: MC simulations of W+jets and diboson events in the 1-lepton channel. The W+jets background is a smooth decay, whereas the diboson signal is peaked around the W and Z masses.

least squares method):

$$\chi^2 = \sum_i \left(\frac{y_i - f(x_i|p)}{e_i} \right)^2, \quad (7.1)$$

where the sum scans over each bin in $m(J)$, the event count and error in each bin are y_i and e_i respectively, and $f(x_i|p)$ is the function evaluated at the bin center x_i for a given choice of the parameters p .

Three functional forms were chosen to be tested, and are listed below. They were motivated by similar functions used previously by ATLAS and CMS for fitting a jet mass distribution [111, 112].

- A single exponential with a bias

$$f(x) = p_1 + p_2 e^{-x/p_3} \quad (7.2)$$

- A linear term plus an exponential

$$f(x) = p_1 + p_2 x + p_3 e^{-x/p_4} \quad (7.3)$$

- The sum of two exponentials

$$f(x) = p_1 e^{-x/p_2} \left(1 + p_3 e^{-x/p_4} \right) \quad (7.4)$$

- Chebyshev polynomials of the first kind, up to order 3 or 4

$$f(x) = \sum_{n=0}^{3,4} p_n \cos(n \cdot \cos^{-1}(x)) \quad (7.5)$$

In all cases, the input variable is $x \equiv m(J)/m_{\max}$ where $m_{\max}$ is the max end of the fit range.

Before proceeding with fitting though, it must be noted that the fit procedure is intimately connected to the choice of W/Z tagger. A W/Z tagger seeks to distinguish fatjets originating from a W or Z boson from those originating from background processes, and defines the $m(J)$ cut for the signal region. Hence the

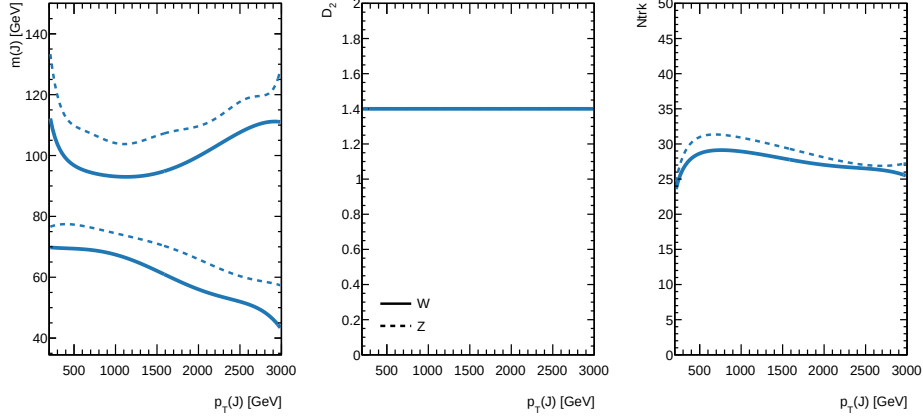


Figure 7.2: The three-variable tagger cuts. The left pad shows the $m(J)$ window; the area between the two lines is taken as the signal. The center and right pads show upper cuts on D_2 and the number of tracks in the jet, respectively. The area below these lines are taken as the signal.

MCR, which inverts the $m(J)$ cut, also depends on the W/Z tagger. Fits were first tested using ATLAS’ older three-variable tagger, which prompted the analysis to switch to ATLAS’ newer ANN tagger [113].

7.1.1 Fits using the Three-variable Tagger

The VV+VH Semileptonic analysis originally used ATLAS’ three-variable tagger. This is actually a pair of separate W and Z taggers, which utilize $p_T(J)$ -dependent cuts on three variables: m , a substructure variable D_2 , and the number of tracks in the jet (N_{trk}) [113]. These cuts are shown in figure 7.2. The D_2 and N_{trk} cuts combine to form a substructure selection; events passing both cuts form the HP region. The MCR is then defined as the HP region with the $m(J)$ cut inverted.

But this analysis does not distinguish between W and Z jets, so the $m(J)$ sidebands need to be defined using the exclusion of both taggers’ signal regions. At the largest $p_T(J)$ the W -tagger has a minimum $m(J)$ cut at around 45 GeV, and at the smallest $p_T(J)$ the Z -tagger has a maximum $m(J)$ cut at around 135 GeV. Then the sidebands could only be defined as $m(J) \notin [45, 135]$ GeV. Since the analysis uses a preselection cut of $m(J) > 50$ GeV, this definition would completely eliminate the lower sideband. Unfortunately, it is not simple to decouple the $m(J)$ cut of the three-variable tagger and use another, since the tagger is calibrated with all three cuts simultaneously. Instead, it was decided to bin the $m(J)$ fits in $p_T(J)$. For a given $p_T(J)$ bin, the $m(J)$ cuts vary less wildly, and the overall signal region window is therefore much tighter. Still, at $p_T(J) \gtrsim 1500$ GeV the signal region window becomes unmanageably large.

A second issue of the three-variable tagger is its non-uniform $m(J)$ response, as shown in figure 7.3. Since D_2 is correlated with $m(J)$, an inflection is created in the $m(J)$ distribution after applying the D_2 cut, which empirically appears to be located around $m(J) = p_T(J)/2$. This point also corresponds to the jet radius

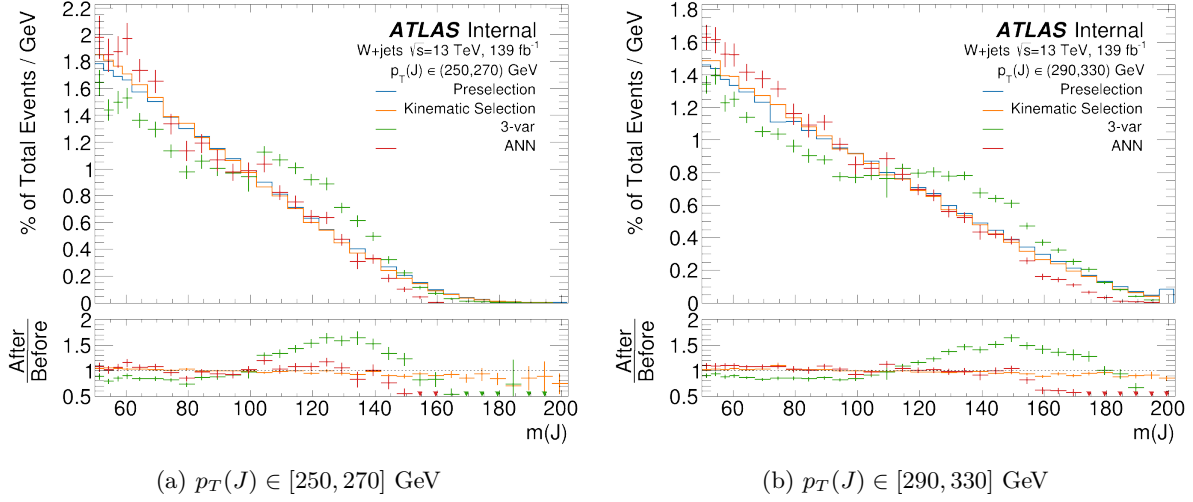


Figure 7.3: $m(J)$ response curves of the three-variable and ANN taggers in two $p_T(J)$ bins. The curve for the three-variable tagger only applies the D_2 and N_{trk} cuts.

cutoff $\Delta R \leq 1.0$ where $\Delta R \approx 2m(J)/p_T(J)$ [114]. To avoid this shoulder in the $m(J)$ distribution, which can be hard to model, a cut-off to the fit was introduced at $m(J) < \max(120 \text{ GeV}, p_T^{\min}/2)$, where $p_T^{\min}$ is the lower edge of each bin. The lower bound at 120 GeV is necessary to make sure the upper sideband always has a suitable number of data points to fit.

Fits were then tested using the single exponential function in equation (7.2) in several $p_T(J)$ bins. Figure 7.4 showcases fits from two bins, and the rest of the fits can be found in appendix C.1.1. Note that in general the sideband boundaries do not lie along bin boundaries, so only bins completely contained in the MCR were used for the fits. As seen in figure 7.4a, the inflection in the mass distribution is quite noticeable, and ruins the simple single exponential fit. However, at higher $p_T(J)$ bins the inflection point is shifted to larger $m(J)$ and can be avoided, as seen in figure 7.4b. The fit in this bin closes quite well against MC, with a reduced $\chi^2 < 1$ and ratios generally within one standard deviation of unity.

7.1.1.1 Two-dimensional Fit

Technically, one could attempt to model the inflection with a more complex functional form. But when the inflection is located inside the signal region window, the fit would become relatively unconstrained. Instead, a two-dimensional fit was proposed. Here, the regression is performed simultaneously over both $m(J)$ and $p_T(J)$. Ideally, $p_T(J)$ bins that are problematic, either due the location of the inflection or the size of the sidebands, could use information from other $p_T(J)$ bins to help constrain the fit.

As defining a fit functional form over two dimensions is somewhat difficult, a two-step empirical procedure was devised. First, the $p_T(J)$ distribution is fitted in 2 GeV bins of $m(J)$. The $p_T(J)$ distribution is much more well-behaved than the $m(J)$ distribution, but has a turn-on shape. A piecewise function was used,

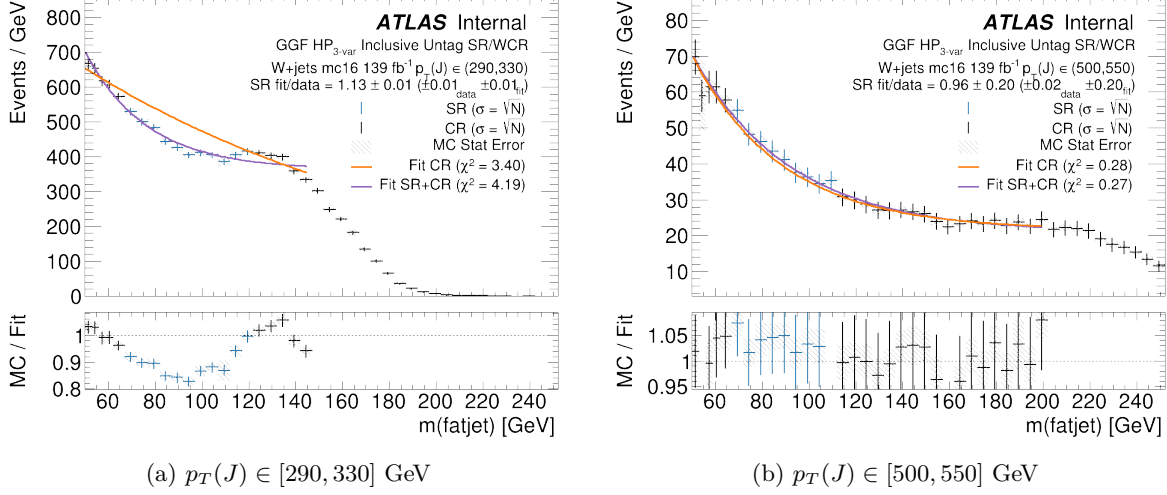


Figure 7.4: Fits to W +jets MC using the three-variable tagger in bins of $p_T(J)$, using a single exponential functional form, equation (7.2). The MC yields are shown by the points, where the error bars show $\sqrt{N}$ errors. The black points are the control region, and the blue points show the inverted control region. The orange line shows the fit to the control region, and the purple line shows a fit to all points in the fit range as a closure test.

consisting of a Gaussian on the left side of the turn-on and a decaying exponential on the right,

$$f(p_T) = \begin{cases} p_0 \exp\left[-\frac{1}{2} \left(\frac{p_T - p_1}{p_2}\right)^2\right] & p_T < p_1 \\ p_0 \exp\left[-\frac{p_T - p_1}{p_3}\right] & p_T > p_1 \end{cases} \quad (7.6)$$

Note this function is continuous but not differentiable at the peak position, defined by the parameter p_1 . The function has four parameters total, and was chosen empirically from inspecting the $p_T(J)$ shape. A sample fit is shown in figure 7.5a; the fit generally performs quite well, with reduced χ^2 averaging around 1. Then, the four parameters from each fit were plotted against the $m(J)$ bin of the fit, shown in figure 7.5b. An empirical relationship between each parameter and $m(J)$ was derived; p_1 , p_2 , and p_3 were identified as approximately linear functions of $m(J)$ and p_0 as a decaying exponential. Replacing the parameters with these functions of $m(J)$ gives a full two-dimensional functional form,

$$f(p_T, m) = \begin{cases} p_0(m) \exp\left[-\frac{1}{2} \left(\frac{p_T - p_1(m)}{p_2(m)}\right)^2\right] & p_T < p_1(m) \\ p_0(m) \exp\left[-\frac{p_T - p_1(m)}{p_3(m)}\right] & p_T > p_1(m) \end{cases} \quad (7.7)$$

$$\begin{aligned} p_0(m) &= q_0 e^{-m/q_1} \\ p_1(m) &= q_2 m + q_3 \\ p_2(m) &= q_4 m + q_5 \\ p_3(m) &= q_6 m + q_7. \end{aligned}$$

Here, the q_i are now the fit parameters that are optimized.

As before, for each $p_T(J)$ bin the inclusive signal region window across the bin is inverted to obtain the MCR. But for the two-dimensional fit, fine uniform $p_T(J)$ bins of size 20 GeV were used. A ratio plot of the

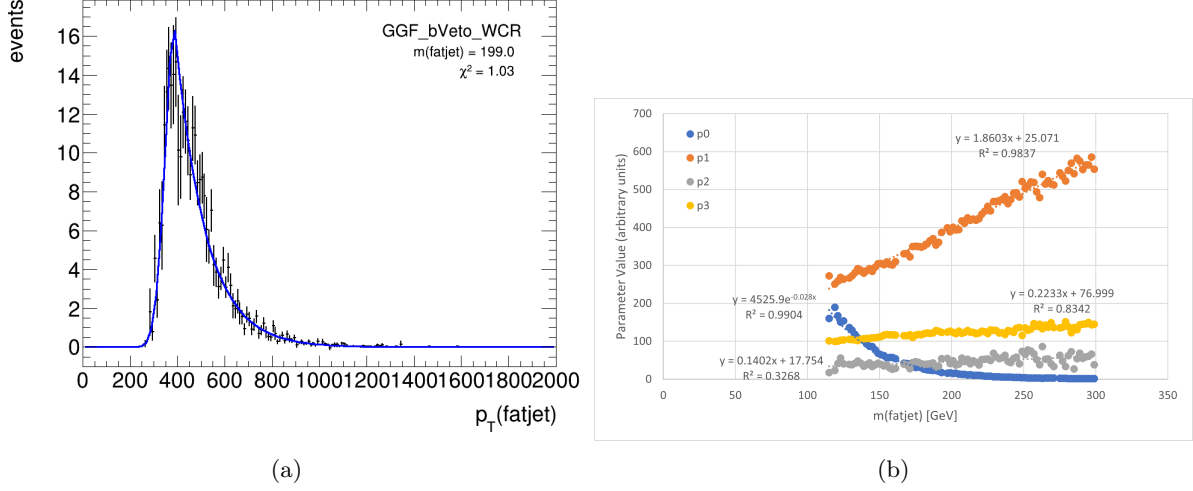


Figure 7.5: Supporting plots for the empirical derivation of the two-dimensional functional form in equation (7.7). (a) A fit to $p_T(J)$ using equation (7.6) in the $m(J) \in [198, 200]$ GeV bin. (b) The parameters from the $p_T(J)$ fits plotted as a function of the $m(J)$ bin. p_0 is seen to follow a decaying exponential trend, while the remaining parameters follow linear trends. Note the trendline functions shown in the plot are not used in the final fit.

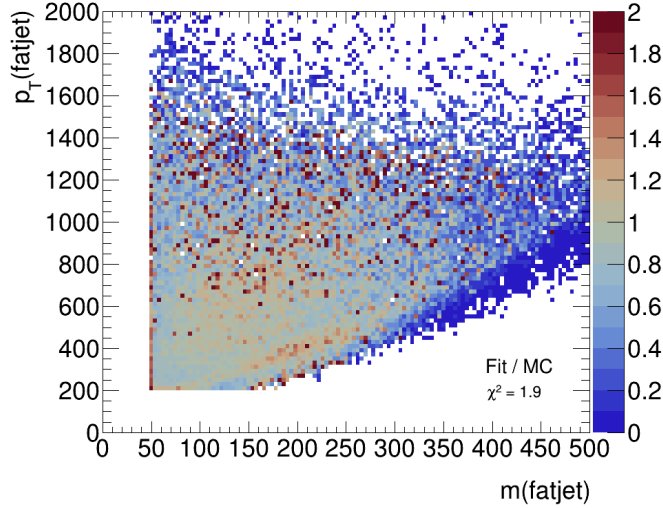


Figure 7.6: A ratio plot of the two-dimensional fit to W +jets MC divided by the MC yield in each bin. While the ratio generally hovers around unity, the deviation can large in certain areas, especially at the tails of the $p_T(J)$ and $m(J)$ spectrums. The closure in the signal region however is generally quite good.

fit result divided by the MC yields is shown in figure 7.6. The ratios generally are close to unity, including inside the signal region, but the function drastically underestimates events in the high $m(J)$ tail. This is not a major problem though since the fit result is only used to interpolate the signal region; however, the uncertainty on the interpolation was generally quite large. The reduced χ^2 was reasonable at $\chi^2_\nu = 1.9$.

Compared to the 1D fits, overall the 2D version tended to perform better at the low and high ends of the

$p_T(J)$ range, where the signal region window is the widest, but at central $p_T(J)$ values its closure against MC and uncertainties were generally worse. Fortunately, at this time ATLAS had begun developing a new W/Z tagger that was decorrelated with mass. This would eliminate the two major problems of the three-variable tagger: the wide $p_T(J)$ -dependent mass cuts and the non-linearity in its $m(J)$ response. Given the complexities and mediocre performance of the current fit tests, the analysis team decided to switch to this new tagger.

7.1.2 Fits using the ANN Tagger

ATLAS' newest W/Z -tagger was the ANN tagger, which combines a deep neural network tagger, trained on jet kinematic and substructure variables, with an adversarial component that decorrelates it from $m(J)$ [113]. Figure 7.3 shows the improvement in the $m(J)$ response compared to the three-variable tagger. The decorrelation avoids the inflection problem from the three-variable tagger, and also allows the analysis to define its own mass cuts; the tagger itself is calibrated as only a substructure cut. A signal region window agnostic to $p_T(J)$ and to W versus Z bosons can then be defined, removing many of the hassles of the three-variable tagger.

The window definition was optimized in the 1-lepton channel with MC W +jets and diboson samples, based on the significance

$$Z = \sqrt{2 \left(n \cdot \log \frac{n}{b} + b - n \right)} \quad (7.8)$$

where n and b are the number of diboson and W +jets events, respectively, that pass both the ANN and mass cuts. Also considered are the signal efficiency and background rejection, which are respectively

$$\epsilon_{\text{sig}} = \frac{n}{N} \quad (7.9)$$

$$1/\epsilon_{\text{bkg}} = \frac{B}{b} \quad (7.10)$$

where N and B are the total number of diboson and W +jets events, respectively, before applying either cut. The lower and upper mass window cut were varied in a two-dimensional scan and the above metrics measured, shown in figure 7.7. The significance peaks at a signal region definition of $m(J) \in [72, 102]$ GeV, which is used in all subsequent studies. This procedure was repeated in bins of $p_T(J)$ to validate that the optimal window is not too correlated with $p_T(J)$; these plots are listed in appendix C.1.1.2.1.

The MCR is now simply defined as the inverted mass cut, $m(J) \notin [72, 102]$ GeV, independent of $p_T(J)$. Fits against the ANN tagger no longer need to be binned in $p_T(J)$, but most of the subsequent tests were still $p_T(J)$ binned for ease of comparison. The cut-off at $m(J) < \max(120 \text{ GeV}, p_T^{\text{min}}/2)$ from before is still used, where p_T^{min} is the lower edge of each bin. This is still due to the sharply falling tagger response attributed to the jet radius cutoff, as mentioned in section 7.1.1. Fits not binned in $p_T(J)$ use a fit range up to $m(J) < 250$ GeV.

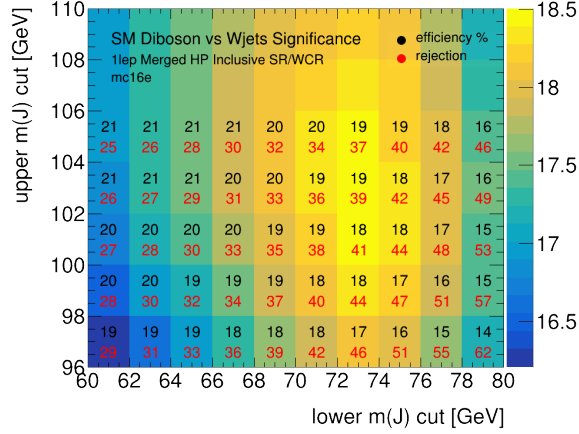


Figure 7.7: A scan of the lower and upper $m(J)$ cuts to be used as the signal region definition with the ANN tagger. Each bin corresponds to cuts at the bottom-left corner of the bin, so for example the bottom-left bin corresponds to a signal region $m(J) \in [60, 96]$ GeV. The z axis shows the significance, and the efficiency (as a percentage) and background rejection are overlaid as text.

Fits were tested using all functions listed in equations (7.2)–(7.5). Figure 7.8 displays fits in two $p_T(J)$ bins using the third-order Chebyshev polynomials; other fit plots can be found in appendix C.1.1.2. One immediately notices that at $p_T(J) < 300$ GeV, the mass distribution is not smooth and the fits perform poorly. The bumpy distribution is caused by poor performance of the adversarial component of the ANN at lower $p_T(J)$. This feature was relayed to the team developing the ANN tagger, but they were unable to identify the cause or fix the issue. However, the sensitivity of the analysis is not significant in the low $p_T(J)$ region to begin with, so a cut at $p_T(J) > 300$ GeV was introduced in all analysis signal regions. Fits at higher $p_T(J)$ bins performed much better; the ratio against the MC yield is uniform around unity and the interpolated signal region yield agrees well with the MC. Errors on the fitted yield are also generally small. The Chebyshev polynomials tended to produce a marginally better fit than the other two functional forms.

Despite the generally good performance of the functional fits, a key issue remained: since the functional forms were chosen completely arbitrarily, such choice could introduce a bias to the fitted yields. Figure 7.9b exemplifies the importance of addressing this systematic error. The fit using a 3rd order polynomial predicts a signal region yield 8% higher than that of the 4th order polynomial, despite both individually seeming to fit well. A common practice is to likewise compare fits of different functional forms, and take the largest difference in their yields as the uncertainty (see for example references [112, 115]). However, this is just as arbitrary as the choice of the original function. If insufficient alternatives had been considered, the error may be underestimated. And if a poorly motivated alternative was included, the error may be overestimated.

In addition, functional forms that work well in some bins may fit poorly to others, as seen in figure 7.9. Naively the function with the best closure in figure 7.9b is the linear+exponential of equation (7.3), but this performs very bad in the lower $m(VV)$ bin of figure 7.9a with a reduced χ^2 of 3.7. Meanwhile the function

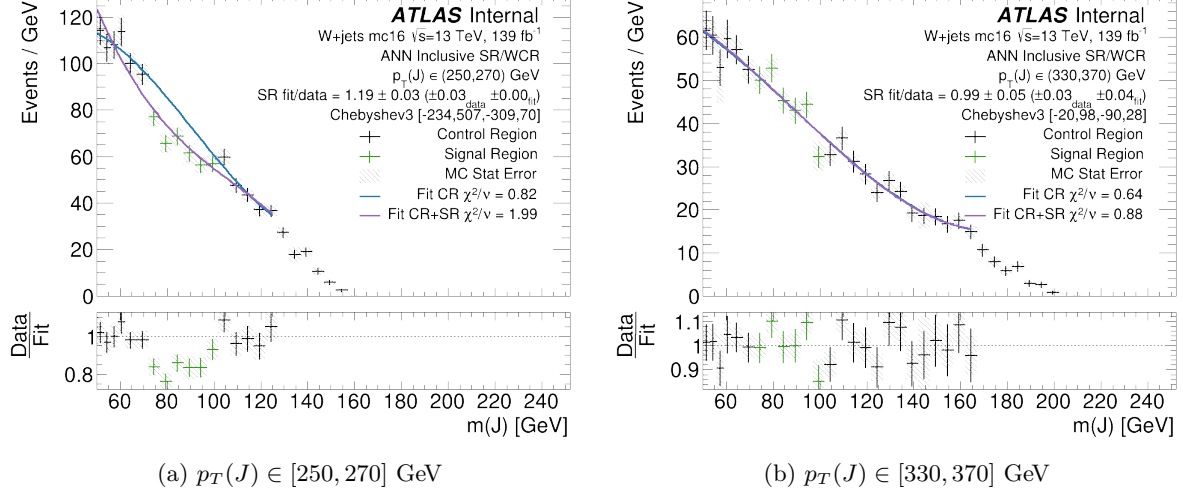


Figure 7.8: Fits to the W+jets MC in the ANN HP region in two bins of $p_T(J)$. These fits use equation (7.5) as the functional form. The MC yields are shown in the black points for the control region and green points for the signal region, where the error bars show $\sqrt{N}$ errors. The blue line shows the fit to the control region, and the purple line shows a fit to both the control and signal regions as a closure test.

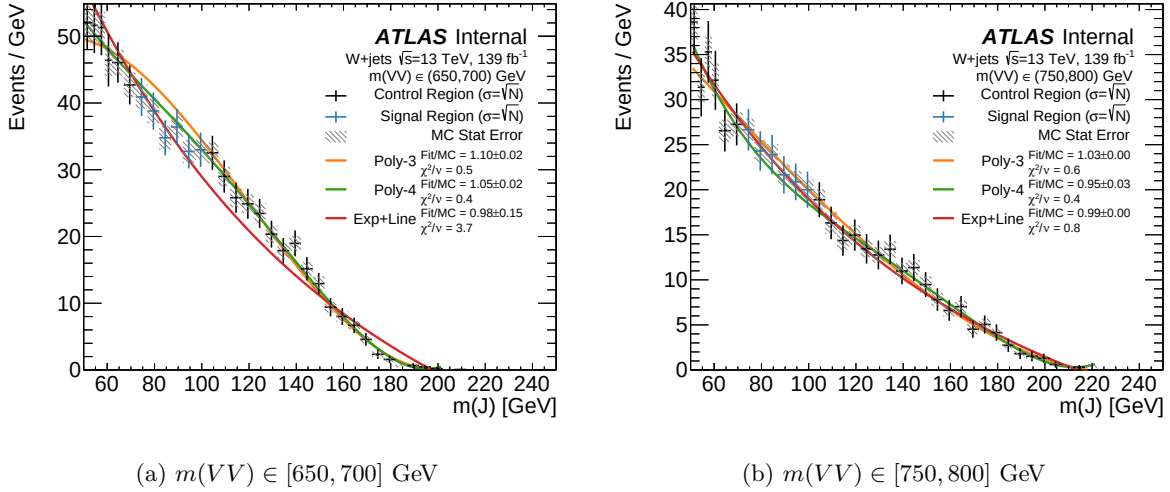


Figure 7.9: Fits to the W+jets MC in the ANN HP region in two bins of $m(VV)$. Three functional forms are used: Chebyshev polynomials of order 3 and 4, and the linear+exponential function from equation (7.3). Error bars on the points represent $\sqrt{N}$ errors as input to the fits, while the shaded box shows the actual MC statistical error. The fit/MC ratio errors in (b) are not accurate due to a non-positive definite covariance matrix.

with the worst closure in figure 7.9b, the 4th order polynomial, is the one with the best fit in figure 7.9a. One could optimize the functional form bin-by-bin, but such a procedure would be tedious not to mention inadvisable, as the variations between the $m(VV)$ bins is likely more statistical than physical. Therefore a procedure optimized for closure on the MC is unlikely to be applicable to the same bin in real data. These problems therefore motivated the transition from using functional fits to the Gaussian process.

7.2 Gaussian Process Regression

A key problem with a functional fit is that it introduces significant bias due to the arbitrary choice of a functional form. Estimating this bias (i.e. the systematic uncertainty) is difficult and imprecise. Gaussian process is a Bayesian regression method that addresses this shortcoming. Instead of choosing a specific class of functions, one assigns a Bayesian prior to the entire space of functions. Applying Bayes' Rule, given some training data, results in a posterior over the space of functions. Hence, unlike a functional fit which outputs a single value, the Bayesian posterior returns a distribution over output values. This distribution intrinsically incorporates both statistical and systematic errors. In other words the systematic error comes “for free” in a mathematically consistent framework.

The majority of this section will introduce the formalism of the Gaussian process. Much of this references the textbook *Gaussian Processes for Machine Learning* by Rasmussen and Williams [116]. Section 7.3 rehashes the mathematics with toy examples, which may be more accessible. Studies on using the Gaussian process are presented in section 7.4.

7.2.1 Bayesian Regression

In a typical regression analysis, we have experimental observations at inputs $\mathbf{x}_i$ in some multidimensional space $\mathcal{D}$ that correspond to target values $y_i \in \mathbb{R}$, where i indexes n observations. The target values may also have some experimental noise σ_i . We assume that the target values are produced via some stochastic process $\{f_* : \mathcal{D} \rightarrow \mathbb{R}\}$, and are interested in finding a function f that best approximates this underlying process. For notational simplicity, we collate the inputs $\mathbf{x}_i$ into a matrix X and similarly the measured targets y_i and errors σ_i into vectors $\mathbf{y}$ and $\boldsymbol{\sigma}$ respectively.

As an example, for fitting V+jets the input is a jet mass $m(J)$; note in this case the inputs are one dimensional so we write them as x_i . The target is the differential event count (i.e. events/GeV) in a bin $m(J) \in [x_i, x_i + \epsilon]$. The true process is a Poisson process with a mean determined by the underlying physics, convoluted by detector and analysis efficiencies, resolutions, and biases.

In a frequentist regression framework, a single estimated function f is obtained by some minimization procedure, such as χ^2 as in section 7.1. In the Bayesian framework, we instead would like a probability distribution over all functions, $p(f|X, \mathbf{y}, \mathcal{H})$. Here, the probability is conditioned on the observed data, sometimes also called the *training data*, and the regression model $\mathcal{H}$.⁶ It can be evaluated via Bayes' Rule

⁶Of course in this section $\mathcal{H}$ will be a Gaussian process model, but this is only an umbrella term. We will see momentarily that a Gaussian process is only fully defined after selecting a specific *kernel*.

as

$$\underbrace{p(f|X, \mathbf{y}, \mathcal{H})}_{\text{posterior}} = \frac{\overbrace{p(\mathbf{y}|f, X, \mathcal{H})}^{\text{likelihood}} \overbrace{p(f|X, \mathcal{H})}^{\text{prior}}}{\underbrace{p(\mathbf{y}|X, \mathcal{H})}_{\text{marginal likelihood}}}. \quad (7.11)$$

As with all Bayesian analyses, a key ingredient is choosing the prior $p(f|X, \mathcal{H})$.⁷ This represents our prior belief of what functions f are valid, without having observed the targets $\mathbf{y}$ yet. If the prior is too strict, we may be unable to model the true process $\{f_*\}$ correctly. If the prior is too broad, the posterior will be unnecessarily imprecise. For example, the prior is where we enforce assumptions about the smoothness of the V+jets process, as mentioned at the start of this chapter.

Defining such a prior may seem daunting at first. We thus introduce the Gaussian process as a way to define the prior, and more generally to define probability density functions (PDFs) over functions. A key feature that we will discover is that, if the prior is assigned to be a Gaussian process, equation 7.11 is analytically solvable. In fact, the posterior is a Gaussian process too.

7.2.2 Functions as an Infinite Vector Space

Before we can introduce the Gaussian process, we need a better handle on describing the space of functions. Note that a function $f : \mathcal{D} \rightarrow \mathbb{R}$ is defined by its output $f(\mathbf{x})$ on every input $\mathbf{x} \in \mathcal{D}$. For example, when $\mathcal{D} = \mathbb{R}$, we define functions by writing $f(x) = x^2$ or $f(x) = \sin(x)$, as two instances. Now suppose we could enumerate $\mathcal{D}$, so we could write out the inputs as $\mathbf{x}_1, \mathbf{x}_2, \dots$. Then f could be equally defined by specifying a series of values $f_1, f_2, \dots$, which correspond to the respective function output by $f_i \equiv f(\mathbf{x}_i)$. In other words, selecting an infinite vector $(f_1, f_2, \dots)$ uniquely identifies a function f . In this sense, a function f is just a point in an infinite-dimensional space, where each dimension corresponds to an input $\mathbf{x}_i$.

We can then define probabilities over functions the same way we define probabilities over vector spaces. Functions are interpreted as a stochastic point in an infinite vector space, or equivalently an infinite set of stochastic variables $f_i \in \mathbb{R}$. Properties of functions, such as smoothness, correspond to correlations on the f_i . For example the condition $f(\mathbf{x}_1) = f(\mathbf{x}_2)$ can be imposed by making $f_1 \equiv f(\mathbf{x}_1)$ and $f_2 \equiv f(\mathbf{x}_2)$ perfectly correlated.

Note that usually $\mathcal{D}$ is some region in $\mathbb{R}^d$, and is not enumerable. However, in most practical use cases, one only cares about the value of f at a finite number of test points, which we denote $\mathbf{f}'$ and X' respectively. And what we technically calculate is $p(\mathbf{f}'|X, \mathbf{y}, X', \mathcal{H})$, as will be seen in equation 7.26. Though when the functions of interest are smooth and slowly varying, and the test points are numerous and finely separated, this distinction becomes negligible. So for simplicity we write $p(f|X, \mathbf{y}, \mathcal{H})$, which should be taken more rigorously to mean $p(\mathbf{f}'|X, \mathbf{y}, X', \mathcal{H})$, and similarly for other probabilities over functions.

⁷Technically, we are actually choosing the model $\mathcal{H}$, and that defines the prior.

7.2.3 Definition of a Gaussian Process

We now define a Gaussian process as the generalization of a multivariate Gaussian to infinite dimensions.

Definition 7.1. A Gaussian process g is an infinite set of stochastic variables in $\mathbb{R}$ with an index set $\mathcal{D}$, where any finite subset of them are jointly Gaussian.

Just like multivariate Gaussians, a Gaussian process is fully defined by its mean and covariance, which now take the form of functions $\mu : \mathcal{D} \rightarrow \mathbb{R}$ and $k : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$, respectively. Denote an arbitrary finite subset of the Gaussian process as $\{g_i\}$, and let $\mathbf{x}_i \in \mathcal{D}$ be the index of g_i , which is written $g_i = g(\mathbf{x}_i)$. The $\{g_i\}$ are jointly Gaussian by definition, and the mean and covariance of this Gaussian is given by

$$\mathbb{E}[g_i] = \mu(\mathbf{x}_i) \quad (7.12)$$

$$\text{Cov}[g_i, g_j] = k(\mathbf{x}_i, \mathbf{x}_j). \quad (7.13)$$

Note that k must be defined such that the covariance matrix is always positive definite for any selection of $\{g_i\}$. It is also referred to as a kernel function. Somewhat confusingly, k returns the covariance between the output values g_i and not the inputs $\mathbf{x}_i$, even though the it is written in terms of the latter. We denote the Gaussian process itself as

$$g \sim \mathcal{N}(\mu, k). \quad (7.14)$$

A Gaussian process defines a probability distribution over an infinite-dimensional vector space, which as noted in the previous section, corresponds to a probability distribution over functions. To sample a Gaussian process $g = \mathcal{N}(\mu, k)$, one must first pick a finite number of test points $\mathbf{x}'_j \in \mathcal{D}$. The set of corresponding outputs $g'_j = g(\mathbf{x}'_j)$ is by definition normally distributed by

$$\mathbf{g}' \sim \mathcal{N}(\boldsymbol{\mu}', K_{X'X'}) \quad (7.15)$$

where as usual we have collated the test points into vectors and matrices given by

$$\mathbf{g}'_j \equiv g(\mathbf{x}'_j) \quad \boldsymbol{\mu}'_j \equiv \mu(\mathbf{x}'_j) \quad (K_{X'X'})_{jk} \equiv k(\mathbf{x}'_j, \mathbf{x}'_k). \quad (7.16)$$

A specific $\mathbf{g}'$ is thus obtained just by sampling from a normal multivariate Gaussian. When the number of test points goes to infinity, as described in the previous section, the vector $\mathbf{g}'$ identifies a function $f : \mathcal{D} \rightarrow \mathbb{R}$.

7.2.4 Gaussian Process Regression

We are now ready to use the Gaussian process in Bayesian regression. We assign the prior to be a Gaussian process g with arbitrary mean μ and kernel k ,

$$f|X, \mathcal{H} \sim \mathcal{N}(\mu, k). \quad (7.17)$$

Generally the mean is set to a constant zero, since the training data should be normalized beforehand. The kernel is often set to the radial basis function (RBF), also called the squared exponential or Gaussian kernel:

$$k(\mathbf{x}_1, \mathbf{x}_2) = \gamma^2 \cdot \exp \left[-\frac{|\mathbf{x}_1 - \mathbf{x}_2|^2}{2\ell^2} \right]. \quad (7.18)$$

Here, ℓ is an arbitrary length scale and γ^2 is an overall variance.

We further assume the measured outputs $\mathbf{y}$ have an additive, independent Gaussian error with variance σ^2 . The likelihood is then

$$\mathbf{y}|f, X, \mathcal{H} \sim \mathcal{N}(\mathbf{f}, \text{diag}(\sigma^2)) \quad (7.19)$$

where $\mathbf{f}$ is the function f evaluated at the training inputs X . Now we can technically solve for the posterior using equation 7.11. However, this involves solving the marginal likelihood as a normalizing integral

$$p(\mathbf{y}|X, \mathcal{H}) = \int d\mathbf{f} p(\mathbf{y}|\mathbf{f}, X, \mathcal{H}) p(\mathbf{f}|X, \mathcal{H}). \quad (7.20)$$

This is not terribly difficult since both terms are Gaussians, but an even simpler method is to use a trick of conditioning the prior.

Let $X' \subset \mathcal{D}$ be a set of test points as in section 7.2.2. Consider the joint distribution between the Gaussian process outputs at the training and test points, $\mathbf{g} = g(X)$ and $\mathbf{g}' = g(X')$ respectively. By definition, this is again a multivariate Gaussian

$$\begin{bmatrix} \mathbf{g} \\ \mathbf{g}' \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}(X) \\ \boldsymbol{\mu}(X') \end{bmatrix}, \begin{bmatrix} K_{XX} & K_{XX'} \\ K_{X'X} & K_{X'X'} \end{bmatrix} \right). \quad (7.21)$$

Here $\boldsymbol{\mu}(A)$ is the vector consisting of the prior's mean function evaluated at each point in A , and $(K_{AB})_{ij} \equiv k(A_i, B_j)$ are matrices from evaluating the kernel function between the training and test points. For the time being, suppose the $\mathbf{y}$ are noiseless. Then training on $\mathbf{y}$ is the same thing as fixing $\mathbf{g} = \mathbf{y}$. That is, we simply condition equation 7.21 on $\mathbf{g} = \mathbf{y}$ to arrive at the posterior. The details are worked out in reference [116]. As it turns out, the posterior is also a Gaussian process given by

$$\mathbf{g}'|X, \mathbf{y}, X', \mathcal{H} \sim \mathcal{N}(\boldsymbol{\mu}', \Sigma') \quad (7.22)$$

$$\boldsymbol{\mu}' = \boldsymbol{\mu}(X') + K_{X'X}[K_{XX}]^{-1}(\mathbf{y} - \boldsymbol{\mu}(X)), \quad (7.23)$$

$$\Sigma' = K_{X'X'} - K_{X'X}[K_{XX}]^{-1}K_{XX'}. \quad (7.24)$$

As mentioned in section 7.2.2, in the limit of infinite test points, $\mathbf{g}'$ identifies a function f and we arrive at a posterior over functions as written in equation 7.11.

If instead the training points are noisy, we cannot fix $\mathbf{g}$ to the single value $\mathbf{y}$. Instead, we have to invert equation 7.19 and sample $\mathbf{g}$ from $\mathcal{N}(\mathbf{y}, \text{diag}(\sigma^2))$. But we assume that the σ noise is independent, so we can actually add it to the covariance in the joint distribution,

$$\begin{bmatrix} \mathbf{g}^* \\ \mathbf{g}' \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}(X) \\ \boldsymbol{\mu}(X') \end{bmatrix}, \begin{bmatrix} K_{XX} + \text{diag}(\sigma^2) & K_{XX'} \\ K_{X'X} & K_{X'X'} \end{bmatrix} \right). \quad (7.25)$$

Equation 7.25 effectively reassigns the noise on the training points to a noise on the Gaussian process outputs. Now the $\mathbf{y}$ are de facto noiseless, and the star on $\mathbf{g}^*$ indicates the modified process. We then follow the noiseless case and condition on $\mathbf{g}^* = \mathbf{y}$, which subtly modifies the posterior to

$$\mathbf{g}'|X, \mathbf{y}, X', \mathcal{H} \sim \mathcal{N}(\boldsymbol{\mu}', \Sigma') \quad (7.26)$$

$$\boldsymbol{\mu}' = \boldsymbol{\mu}(X') + K_{X'X}[K_{XX} + \text{diag}(\boldsymbol{\sigma}^2)]^{-1}(\mathbf{y} - \boldsymbol{\mu}(X)), \quad (7.27)$$

$$\Sigma' = K_{X'X'} - K_{X'X}[K_{XX} + \text{diag}(\boldsymbol{\sigma}^2)]^{-1}K_{XX'}. \quad (7.28)$$

Equations 7.26–7.28 form the main result of this section.

At this point it is useful to write out the marginal likelihood explicitly for future use. Note that from above

$$\mathbf{g}^*|X, \mathcal{H} \sim \mathcal{N}(\boldsymbol{\mu}(X), K_{XX} + \text{diag}(\boldsymbol{\sigma}^2)), \quad (7.29)$$

so the marginal likelihood can be obtained by just evaluating this multivariate normal at $\mathbf{g}^* = \mathbf{y}$:

$$p(\mathbf{y}|X, \mathcal{H}) = \frac{1}{\sqrt{(2\pi)^n |K|}} \exp \left[-\frac{1}{2} \mathbf{v}^T K^{-1} \mathbf{v} \right] \quad (7.30)$$

$$\mathbf{v} \equiv \mathbf{y} - \boldsymbol{\mu}(X) \quad (7.31)$$

$$K \equiv K_{XX} + \text{diag}(\boldsymbol{\sigma}^2), \quad (7.32)$$

where n is the number of training points. We have completely avoided the integral in equation 7.20.

7.2.5 Selecting a Kernel

The choice of the kernel and especially the choice of the length scale in the RBF kernel can dramatically affect the posterior distribution. One may rightly worry that this arbitrary choice introduces bias just like choosing a functional form in a functional fit. Do note that in the Bayesian framework, the kernel is exactly the prior, which in some sense always depends on our prior biases. Still, our prior beliefs are usually not so strict where we should fix an exact kernel. For example, we assume the V+jets background is smooth at scales larger than the W/Z mass separation (~ 10 GeV), but we are unable to state an explicit length scale. So we should scan length scales within some reasonable range. We call all auxiliary parameters in a kernel the hyperparameters, and denote them collectively as $\boldsymbol{\theta}$. The RBF kernel from equation 7.18 has two hyperparameters, $\boldsymbol{\theta} = (\ell, \gamma)$.

The choice of the RBF kernel versus another form is somewhat negligible in fitting V+jets. We only want kernels that are isotropic, that is they can be written in terms of only $|\mathbf{x}_1 - \mathbf{x}_2|/\ell$, and these tend to behave similarly. Studies in section 7.4 will quantify this more precisely, but we will ignore the error from the discrete choice of kernel form moving forward. Thus in the following equations we replace the condition

on the model $\mathcal{H}$ with just a condition on the hyperparameters $\boldsymbol{\theta}$, which are all kernel parameters that are not the inputs $\mathbf{x}_i$, leaving the model selection implicit.

Choosing a specific set of hyperparameters can be done using the marginal likelihood $p(\mathbf{y}|X, \boldsymbol{\theta})$ derived in equation 7.30. In the simplest case, the model with the highest marginal likelihood can be chosen. This approximation is known as type II maximum likelihood [116], and the selected hyperparameters we refer to as the maximum marginal likelihood estimator (MMLE),

$$\boldsymbol{\theta}_{\text{MMLE}} \equiv \underset{\boldsymbol{\theta}}{\operatorname{argmax}} p(\mathbf{y}|X, \boldsymbol{\theta}). \quad (7.33)$$

The MMLE can be thought of as choosing the hyperparameters that match the training data $\mathbf{y}$ the best. Importantly, the marginal likelihood automatically includes a trade-off between model complexity and accuracy, so the maximum will generally be well-peaked.

A more complete approach is to marginalize out the hyperparameters. Recall in equation 7.26 we found the posterior conditioned on $\boldsymbol{\theta}$, $p(f|X, \mathbf{y}, \boldsymbol{\theta})$. We now wish to find the posterior independent of our choice of $\boldsymbol{\theta}$, the marginal posterior,

$$p(f|X, \mathbf{y}) = \int d\boldsymbol{\theta} p(f|X, \mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|X, \mathbf{y}). \quad (7.34)$$

The second term in the integrand is called the “hyperposterior” (posterior over hyperparameters) and can be computed from Bayes’ Rule as

$$\underbrace{p(\boldsymbol{\theta}|X, \mathbf{y})}_{\text{hyperposterior}} = \frac{\overbrace{p(\mathbf{y}|X, \boldsymbol{\theta})}^{\text{marginal likelihood}} \times \overbrace{p(\boldsymbol{\theta}|X)}^{\text{hyperprior}}}{\underbrace{p(\mathbf{y}|X)}_{\text{normalizing constant}}}. \quad (7.35)$$

We note the reappearance of the marginal likelihood in the numerator. The normalizing constant is given by

$$p(\mathbf{y}|X) = \int d\boldsymbol{\theta} p(\mathbf{y}|X, \boldsymbol{\theta}) p(\boldsymbol{\theta}|X). \quad (7.36)$$

7.2.6 Numerical Approximation of the Marginal Posterior

The integrals in equations 7.34 and 7.36 are generally not analytic because of the non-Gaussian marginal likelihood. They can also be quite difficult to estimate as the $\boldsymbol{\theta}$ integrals scan over an infinite, multidimensional hyperparameter space, and f needs to be scanned over the space of functions to obtain the full distribution $p(f|X, \mathbf{y})$.

However a numerical estimation can be made tractable given some conditions. If we assume a hyperprior that is uniform in a rectangular region and zero elsewhere, a grid can be defined over the $\boldsymbol{\theta}$ -space. Then the $\boldsymbol{\theta}$ integrals can be approximated as Riemann sums [117] over the grid. The hyperposterior becomes merely a normalized version of the marginal likelihood,

$$p(\boldsymbol{\theta}_i|X, \mathbf{y}) \approx \frac{p(\mathbf{y}|X, \boldsymbol{\theta}_i)}{\sum_{j \in \text{grid}} p(\mathbf{y}|X, \boldsymbol{\theta}_j)}, \quad (7.37)$$

where θ_i refers to a specific point on the hyperparameter grid. We find that the marginal posterior is then obtained by weighted sum of $p(f|X, \mathbf{y}, \theta)$ over the grid, where the weights are the marginal likelihoods:

$$p(f|X, \mathbf{y}) \approx \sum_{i \in \text{grid}} p(f|X, \mathbf{y}, \theta_i) p(\mathbf{y}|X, \theta_i) \bigg/ \sum_{j \in \text{grid}} p(\mathbf{y}|X, \theta_j). \quad (7.38)$$

This can still be quite taxing, since to get the full $p(f|X, \mathbf{y})$ distribution, the numerator in equation 7.38 has to be recalculated at each value of f . Fortunately, for fitting V+jets we only care about the integral of f over the signal region, $\mathcal{I}$, and not its value at each point. The integral can be approximated by another Riemann sum over some arbitrary number of test points in the signal region, $\mathbf{x}'_i$. So if f'_i are the corresponding outputs, we want

$$\mathcal{I} \propto \sum_i f'_i = \mathbf{1}^T \mathbf{f}', \quad (7.39)$$

where $\mathbf{1} = (1, \dots, 1)^T$ and $\mathbf{f}'$ collates the f'_i into a vector. Recall the posterior is a Gaussian process too, so the output values are jointly Gaussian: $\mathbf{f}' \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu}$ and Σ are given by equation 7.26. But any linear combination of jointly Gaussian variables is also Gaussian [118], and we find

$$\mathcal{I}|X, \mathbf{y}, \theta_i \sim \mathcal{N}(\mathbf{1}^T \boldsymbol{\mu}, \mathbf{1}^T \Sigma \mathbf{1}). \quad (7.40)$$

Equation 7.38 can then be replaced with

$$p(\mathcal{I}|X, \mathbf{y}) \approx \sum_{i \in \text{grid}} p(\mathcal{I}|X, \mathbf{y}, \theta_i) p(\mathbf{y}|X, \theta_i) \bigg/ \sum_{j \in \text{grid}} p(\mathbf{y}|X, \theta_j). \quad (7.41)$$

The numerator now only needs to be scanned over values of $\mathcal{I}$, a one-dimensional scan.

An even further simplification is to restrict the sums in equation 7.41 to only those points where the marginal likelihood makes a noticeable contribution. For example, we could choose the hyperposterior 1σ or 2σ credibility region. Since the marginal likelihood is usually well-peaked, we define the credibility region as those points on the grid with the highest hyperposterior density that sum to 68% or 95% of the total grid, respectively.

7.3 Toy Examples for Gaussian Process

The math in the previous section can be quite confusing. This section will try to break down the math using some toy examples, and hopefully help develop an intuitive understanding behind Gaussian process regression. All examples in this section use a one-dimensional input space $\mathcal{D} = [-5, 5]$.⁸

⁸These examples were created in a Jupyter notebook [119], available at <https://github.com/riley-x/GaussianProcessExample>. The reader is encouraged to try using the notebook as a playground to further examine the effect of different kernels, training points, and so on. This notebook directly implements the equations in section 7.2 using only NumPy [120] matrix algebra.

7.3.1 The Probability Density Function

Since specifying the prior is the main task in Gaussian process regression, a key point is to understand what the PDF of a Gaussian process looks like.

7.3.1.1 Sampling a Gaussian Process

The easiest way to visualize a Gaussian process is by sampling it. Since a Gaussian process is a probability distribution over functions, each sample is a function, which can be readily plotted. Samples of course only offer a superficial understanding, but they enable simple comparisons between different Gaussian processes.

To sample a Gaussian process, we must first discretize the input space into a series of test points. Let there be n input points $\mathbf{x} = [x_1, \dots, x_n]^T$, evenly distributed in the input space. We then evaluate the mean and kernel functions of our Gaussian process at each input point, obtaining an n -vector of means $\boldsymbol{\mu}$ and an $(n \times n)$ covariance matrix K ,

$$\boldsymbol{\mu}_i = \mu(x_i) \qquad K_{ij} = k(x_i, x_j). \qquad (7.42)$$

These define an n -dimensional Gaussian distribution. This can be sampled using the method described in reference [121]:

1. Find the Cholesky decomposition of K , a matrix A such that $AA^T = K$.
2. Sample a standard normal (zero mean and unit variance) n times, forming a vector $\mathbf{z}$.
3. Calculate $\mathbf{f} = \boldsymbol{\mu} + A\mathbf{z}$, which has the desired distribution.

Each sampling of $\mathbf{z}$ produces a vector of output values $\mathbf{f}$, which loosely can be drawn as a smooth curve to represent a function f . By using the same $\mathbf{z}$ vectors, different Gaussian processes can be heuristically compared.

An example is shown in figure 7.10 using the RBF kernel from equation (7.18). Four samples are plotted using three different hyperparameter settings. For each of the hyperparameter settings, the same $\mathbf{z}$ vectors are used. The dashed line is the mean $\boldsymbol{\mu}$, chosen in this case to be zero. Overlaid is a shaded band representing the standard deviation at each point, $\sqrt{K_{ii}}$. In figure 7.10a, a benchmark configuration of $\ell = \gamma = 1$ is chosen. First, notice that the output values are generally constrained around the mean. In the limit of infinite samples, 68% of the output values should fall in the shaded region. Also notice that the functions are very smooth, as the RBF kernel forces nearby values to be highly correlated. A characteristic length scale can be inferred: the distance between local extrema is generally around $\Delta x \approx 2$, corresponding to twice the value of ℓ .

In figure 7.10b, ℓ is increased to 5. Since γ did not change, the variance of the output values around the mean has not changed drastically. However, the sampled functions are evidently much flatter with far fewer

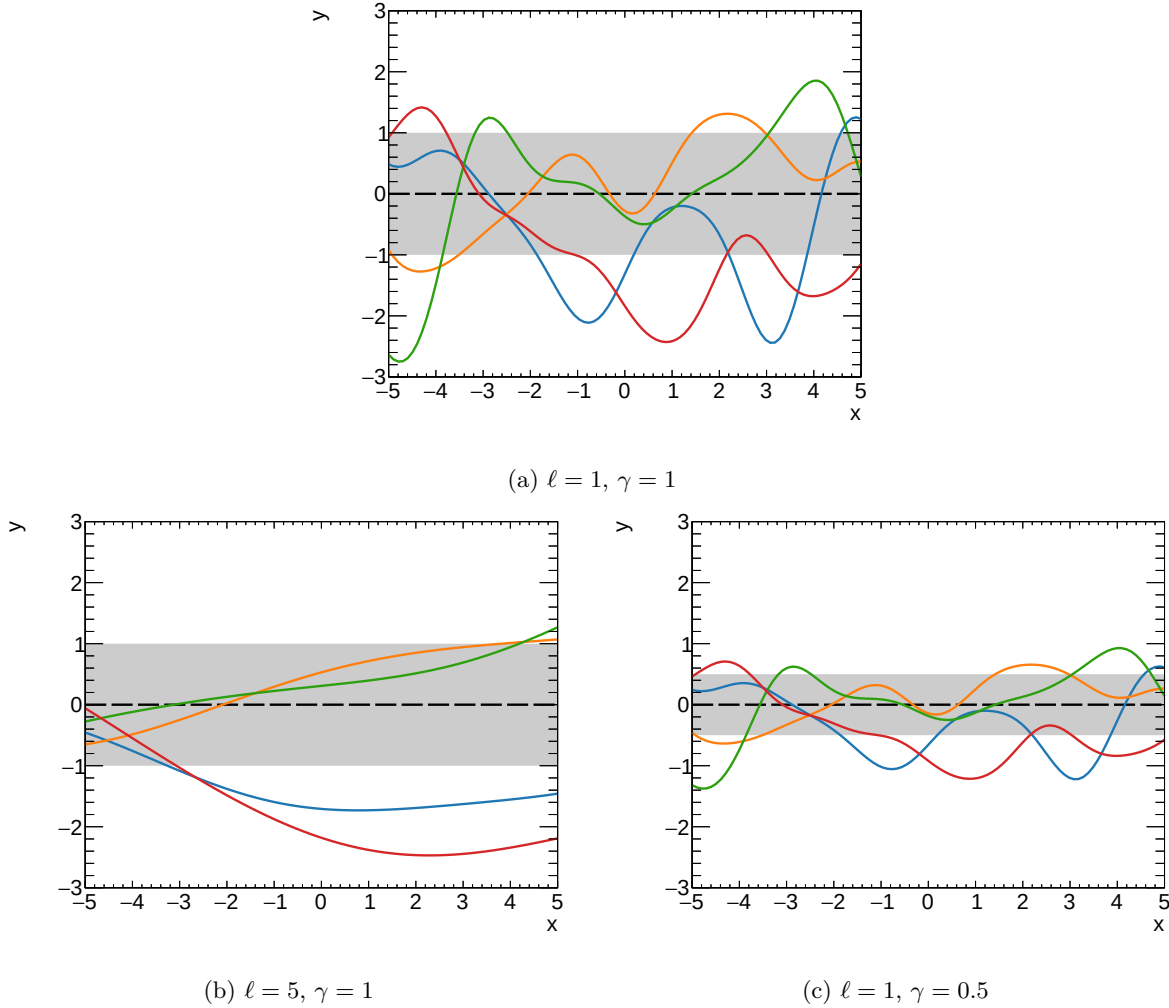


Figure 7.10: Samples drawn from a Gaussian process $\mathcal{N}(0, \text{RBF}_{\ell, \gamma})$ with three different settings of ℓ and γ . The colored lines are the outputs of each sample drawn as a smooth curve, the dashed black line is the mean, and the shaded band is the standard deviation at each x value.

local extrema, thanks to the increased length scale. Meanwhile in figure 7.10c, ℓ is restored to 1 but γ is decreased to 0.5. The sampled functions are exactly the same as in figure 7.10a, except the y values are all a factor of 2 smaller. So in summary, ℓ controls how “wiggly” the sampled functions are, while γ controls how closely they stick to the mean.

Be warned that the observations above are heuristic. Gaussian processes represent probability distributions with infinite tails. So for example, while the samples in figure 7.10b appear generally flat, rapidly varying functions are not prohibited; they are merely much less probable. Similarly, slowly varying functions are not prohibited from the Gaussian process in figure 7.10a either. In fact, they are equally probable as the rapidly varying ones shown. However, the phase space for rapidly varying functions is much larger, so the slowly varying ones are “drowned out,” so to speak. The next sections will investigate this in more detail.

7.3.1.2 PDF of a Simple Gaussian Process

In the previous section, we visualized Gaussian processes by drawing random samples from their PDFs. Inversely, we can also calculate the probability density at specific test functions. To start, consider a simple Gaussian process g_1 defined by

$$g_1 = \mathcal{N}(\mu, k) \quad (7.43)$$

$$\mu(x) = 0 \quad (7.44)$$

$$k(x, x') = \delta(x - x'). \quad (7.45)$$

We want to find the probability density $p(f)$ at an arbitrary function f .

As before, we need to first discretize the input space into n input points $\mathbf{x} = [x_1, \dots, x_n]^T$. Evaluating f at each point gives us a vector $\mathbf{f} = f(\mathbf{x}) = [f(x_1), \dots, f(x_n)]^T$. We again create an n -dimensional Gaussian from the mean and kernel functions, which we can write explicitly in the case of g_1 :

$$\boldsymbol{\mu}_i = \mu(x_i) = \mathbf{0} \quad K_{ij} = k(x_i, x_j) = \delta_{ij}. \quad (7.46)$$

The probability density $p(f)$ is then found by simply evaluating the PDF of a multivariate normal,

$$p(\mathbf{f}) = \frac{1}{\sqrt{(2\pi)^n |K|}} \exp \left[-\frac{1}{2} (\mathbf{f} - \boldsymbol{\mu})^T K^{-1} (\mathbf{f} - \boldsymbol{\mu}) \right] \quad (7.47)$$

$$= (2\pi)^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \mathbf{f}^T \mathbf{f} \right]. \quad (7.48)$$

As with all Gaussians, this probability has a maximum at the mean, that is when $f(x) = \mu(x)$. It is convenient to rewrite the probabilities as

$$\eta(\mathbf{f}) \equiv \frac{\ln p(\boldsymbol{\mu}) - \ln p(\mathbf{f})}{n} \quad (7.49)$$

$$= \frac{1}{2n} |\mathbf{f}|^2. \quad (7.50)$$

Note that η has a minimum at $\eta = 0$, and larger η corresponds to less probable $\mathbf{f}$. The normalization by n ensures η is comparable for different choices of n .

This simple analytic form shows us that as the magnitude of $\mathbf{f}$ increases, the probability of the function decreases. In other words, our Gaussian process favors functions that are close to the mean $\mu(x) = 0$. We verify this with a quick test using $n = 101$ inputs points, such that

$$\mathbf{x} = [-5, -4.9, \dots, 4.9, 5]^T. \quad (7.51)$$

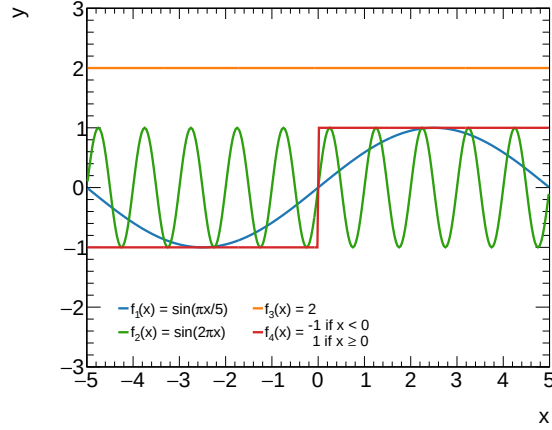


Figure 7.11: Test functions for evaluating a Gaussian process' PDF.

We then pick four sample functions to test:

$$f_1(x) = \sin(\pi x/5) \quad \eta(f_1) = 0.25 \quad (7.52)$$

$$f_2(x) = \sin(2\pi x) \quad \eta(f_2) = 0.25 \quad (7.53)$$

$$f_3(x) = 2 \quad \eta(f_3) = 2 \quad (7.54)$$

$$f_4(x) = -1 \text{ if } x < 0, 1 \text{ if } x \geq 0 \quad \eta(f_4) = 0.5. \quad (7.55)$$

These functions are shown in figure 7.11. Notice that $f_3(x) = 2$ has the worst η score, since it deviates the most from 0. Meanwhile the two sine functions have the best η scores because they hover around 0. Although f_2 is more rapidly varying, they have effectively identical η scores since the magnitude of the oscillations are the same. Finally the step function f_4 has a mediocre η score because it deviates more from 0 than the sine functions, but not as much as f_3 . It is also easy to verify that the η values do not change much when n is altered, as long as it is sufficiently large.

The importance of 0 is because we chose that to be the mean in equation (7.44). If we have an arbitrary mean function $\mu(x)$, the η score is simply modified to

$$\eta(\mathbf{f}) = \frac{1}{2n} |\mathbf{f} - \boldsymbol{\mu}|^2. \quad (7.56)$$

Thus, the choice of mean may seem critical at first, as it can drastically affect the Gaussian process' PDF. However, we can compensate for this by adding an overall scale γ^2 to the kernel, $k(x, x') = \gamma^2 \delta(x - x')$. The η score becomes

$$\eta(\mathbf{f}) = \frac{1}{2n} \frac{|\mathbf{f} - \boldsymbol{\mu}|^2}{\gamma^2}. \quad (7.57)$$

When γ is chosen to be large, relatively small changes to $\boldsymbol{\mu}$ do not affect η much. Note also that in regression, the choice of the mean becomes much less important as the data constrains the posterior.

ℓ	γ	$f_1 = \sin(\pi x/5)$	$f_2 = \sin(2\pi x)$	$f_3 = 2$	$f_4 = \text{step func.}$
1	1	0.01	19,627	0.10	1,137
0.1	1	0.10	0.12	0.81	0.23
10	1	153	24,324	0.04	6,612
1	0.1	1.36	23,379	10.22	1,398
1	10	0.0001	7,739	0.001	980

Table 7.1: η scores as defined in equation (7.49) of the four test functions from equations (7.52)–(7.55), using a Gaussian process $g_2 = \mathcal{N}(0, \text{RBF}_{\ell,\gamma})$ where ℓ and γ are varied. Larger η indicates a less probable function.

The δ -function in the kernel of this example makes the PDF easily analytic. However this ignores correlations between nearby points, which is why the rapidly varying sine function f_2 has the same η score as the slower varying f_1 . But in practice we do want correlations, so we need a more complicated kernel.

7.3.1.3 Radial Basis Function

We now repeat the analysis above with the RBF kernel,

$$\text{RBF}_{\ell,\gamma}(x_1, x_2) = \gamma^2 \cdot \exp \left[-\frac{|x_1 - x_2|^2}{2\ell^2} \right]. \quad (7.58)$$

Consider a Gaussian process g_2 with a zero mean function and the RBF kernel as the covariance function,

$$g_2 = \mathcal{N}(0, \text{RBF}_{\ell,\gamma}). \quad (7.59)$$

Although η no longer has a concise form, we can still examine the η scores of the test functions in equations (7.52)–(7.55). We try a variety of choices of ℓ and γ , which are summarized in table 7.1.

The first row uses $\ell = \gamma = 1$ as a nominal starting point. Immediately we see the rapidly varying sine function f_2 is given an extremely poor η score, since the function oscillates on scales smaller than the length scale $\ell = 1$. The slowly varying sine function f_1 oscillates on scales larger than the length scale, and so has a very good η score. It even beats the constant function $f_3(x) = 2$, because it sticks more closely to the mean $\mu(x) = 0$. Finally the step function f_4 has a bad η score due to the discontinuity, though not as bad as f_2 .

Next we try setting $\ell = 0.1$ but keeping $\gamma = 1$, shown in the second row of table 7.1. Here all the η scores become relatively good, since such a short length scale is not restrictive. Indeed, in the limit of $\ell \rightarrow 0$, the RBF kernel becomes the δ -function kernel in equation (7.45), and the η scores converge to those in equations (7.52)–(7.55). Inversely, if we increase ℓ to 10, the η score explodes for all functions except the constant function f_3 . In the limit of $\ell \rightarrow \infty$, only constant functions would be allowed.

Finally we revert to $\ell = 1$ but vary γ to 0.1 and 10. The trend of the η values follows the $\gamma = 1$ case, but the magnitude of all the scores increase and decrease, respectively. We see γ is not too important in prioritizing specific functions. But as alluded to above, it is important in the Bayesian context as it reflects the confidence in the prior’s mean. A poor choice of γ will still lead to bad regression performance.

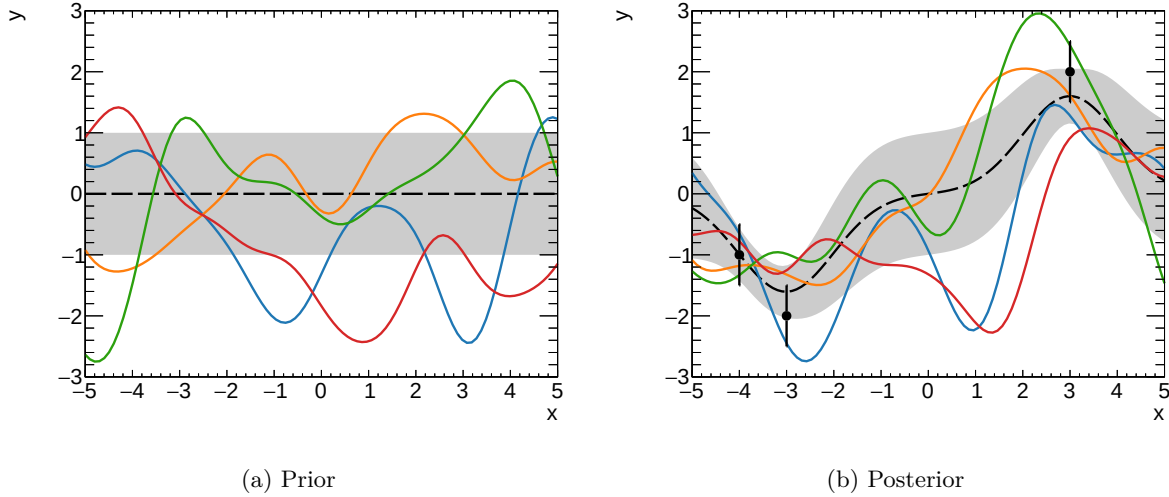


Figure 7.12: Prior and posterior of a Gaussian process $\mathcal{N}(0, \text{RBF}_{\ell, \gamma})$ with $\ell = \gamma = 1$. The dashed black line is the mean and the shaded band is the standard deviation at each x value. The colored lines are four samples from the Gaussian process plotted as smooth curves. Samples of the same color use the same $\mathbf{z}$ vector in the procedure described in section 7.3.1.1.

7.3.2 The Posterior

After specifying the prior by choosing a Gaussian process, the posterior is obtained by simply plugging everything into equations (7.26)–(7.28). We illustrate this process by using the RBF kernel with $\ell = \gamma = 1$, and three ad hoc data points. The mean is fixed to zero still. Since the posterior is also a Gaussian process, we can plot it the same way we did the prior in section 7.3.1.1. Both the prior and posterior, along with the chosen data points, are shown in figure 7.12. As before, the dashed black line represents the mean of the posterior. The mean is usually interpreted as the final predicted function f of the regression. The shaded gray band shows the standard deviation at each x value, and represents the error of the regression. Notice that the error is smaller near the data points and grows as one moves away from the data. Thus it represents both the statistical error from noisy data, and the systematic error from interpolation.

7.3.2.1 Effect of Hyperparameters

It is instructive to again examine the effect of altering the hyperparameters, but now on the posterior. We still use the RBF kernel and start with $\ell = \gamma = 1$ as before, shown in figure 7.13a. First, note that the mean of the posterior reverts to zero, the prior, around $x \in [-1, 1]$. This reflects the length scale $\ell = 1$; the data point at $x = 3$ stops providing any constraint after $\Delta x \approx 2\ell$. In figure 7.13b, ℓ is adjusted to 0.5, and the zero-region expands to $x \in [-2, 2]$. Meanwhile in figure 7.13c, ℓ is adjusted to 5, which correlates the data points enough that the mean no longer deflects to a local minimum at $x = -3$.

Next, notice in figure 7.13a the posterior mean does not go through the data points exactly, but biases

towards zero, the prior. The amount of bias is controlled by γ . When γ is adjusted to 0.5 (the same as the error on the data points) in figure 7.13d, the posterior mean only reaches about halfway to each point.⁹ Hence γ should be thought of as the inverse confidence on the prior; a small γ represents a strong prior, and the posterior will be unwilling to accommodate data points with larger error. However, one should avoid making γ too large, or else the errors of the posterior will be unnecessarily broad, as seen in figure 7.13e.

7.3.2.2 Other Kernels

So far we have focused on the RBF kernel. Some other popular kernels include the Matern 5/2 kernel,

$$k(r) = \gamma^2 \left(1 + \frac{\sqrt{5}r}{\ell} + \frac{5r^2}{3\ell^2} \right) \exp \left(-\frac{\sqrt{5}r}{\ell} \right), \quad (7.60)$$

and the rational quadratic kernel,

$$k(r) = \gamma^2 \left(1 + \frac{r^2}{2\alpha\ell^2} \right)^{-\alpha}. \quad (7.61)$$

Here these are written in terms of $r \equiv |\mathbf{x}_1 - \mathbf{x}_2|$, which emphasizes that these kernels (and the RBF kernel) are isotropic. Both of these depend on a length scale ℓ and the variance scale γ , and the rational quadratic depends additionally on the exponent α .

The posterior using these kernels is shown in figure 7.14. One immediately notices that the sampled functions (the colored lines) are dramatically different. However, the mean and standard deviations are almost imperceptibly changed. This negligible difference can be attributed to the isotropic form of all these kernels. The isotropic assumption is usually well-motivated when lacking a compelling prior, as no direction or location should be favored. Since the difference between kernel forms is marginal, the error on using only the RBF kernel will be ignored, as mentioned in section 7.2.5.

7.3.3 Marginal Likelihood

The previous section has shown that while the kernel form has a minute effect, the hyperparameters are quite important, and can drastically affect the posterior. Section 7.2.5 laid out two methods for dealing with this problem. The first and more common method is to simply use the MMLE, equation (7.33). The second and more complete method is to find the marginal posterior, equation (7.34). In both cases, the marginal likelihood takes a key role, as it is either maximized or integrated over, respectively.

Intuitively, the marginal likelihood measures how well the training data matches the prior. An important quality of the marginal likelihood is that it naturally includes a trade-off between model complexity and accuracy. This means that in hyperparameter-space, the marginal likelihood generally has a well-defined maximum, making the previously mentioned methods feasible. A schematic explanation of this trade-off is

⁹It would be exactly halfway if there was only one data point, which can be seen easily from equation (7.26).

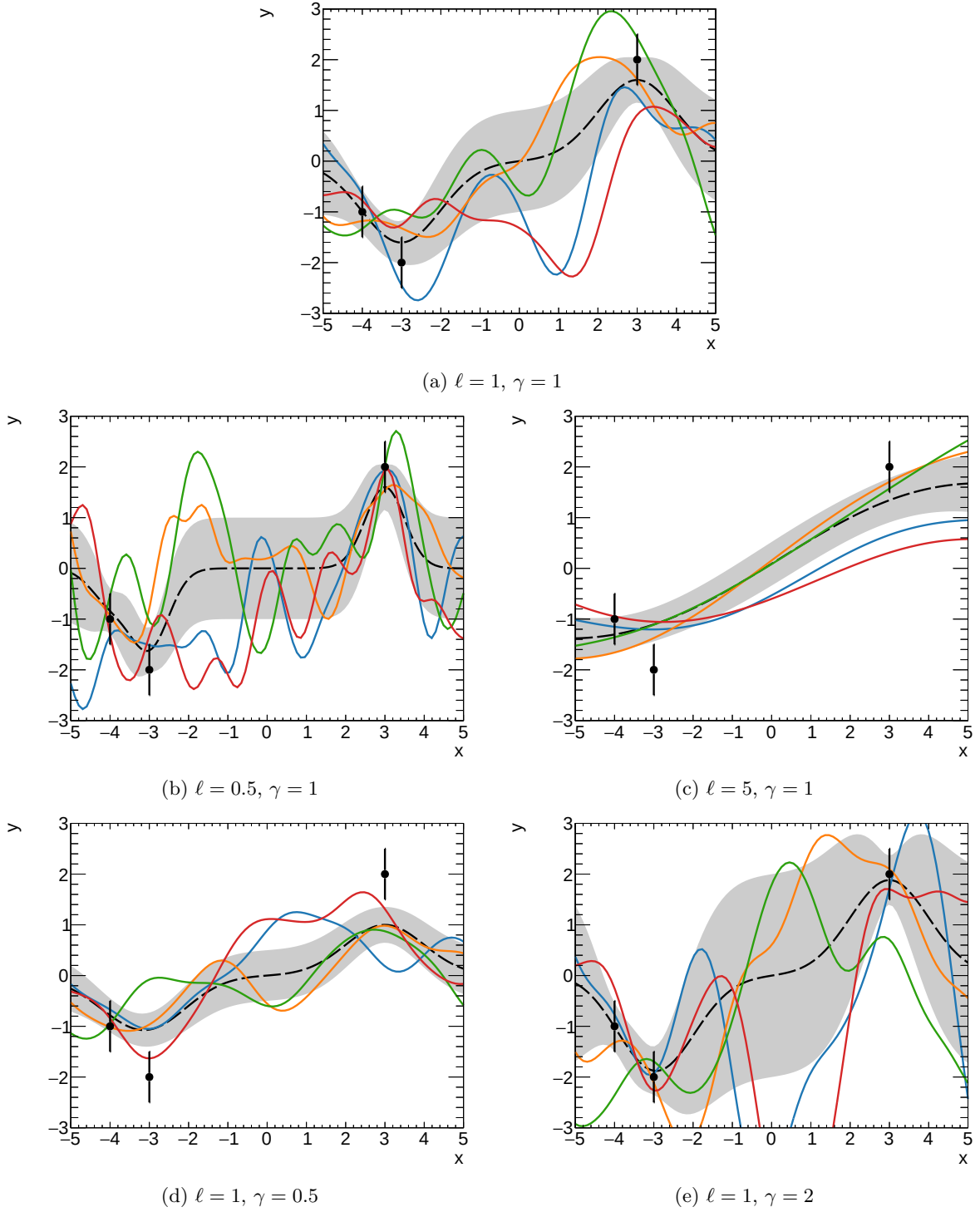


Figure 7.13: Posterior of a Gaussian process $\mathcal{N}(0, \text{RBF}_{\ell, \gamma})$, for various settings of ℓ and γ . The dashed black line is the mean and the shaded band is the standard deviation at each x value. The colored lines are four samples from the Gaussian process plotted as smooth curves. Samples of the same color use the same $\mathbf{z}$ vector in the procedure described in section 7.3.1.1.

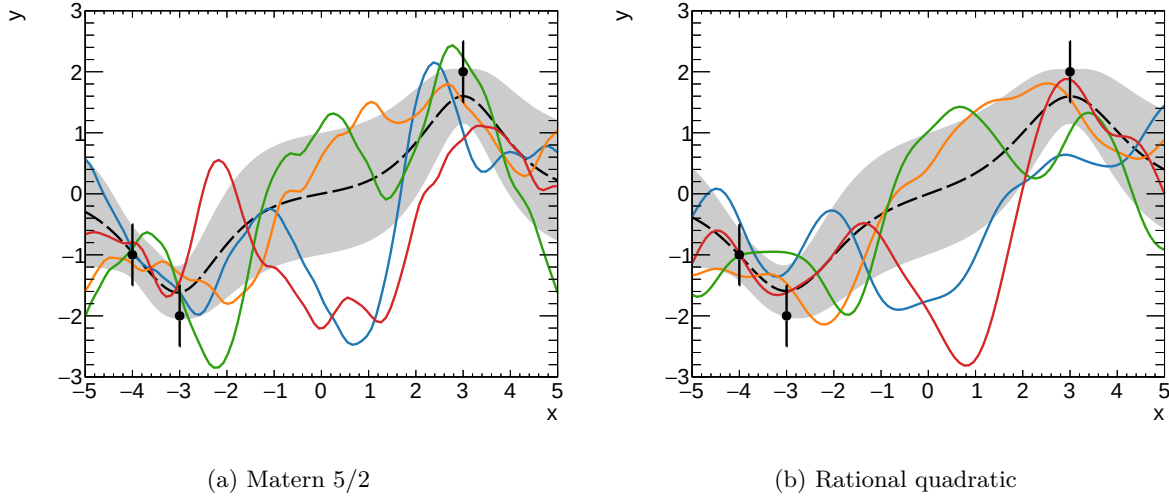


Figure 7.14: Posteriors of Gaussian processes with a (a) Matern 5/2 and (b) rational quadratic kernel. Both have $\ell = \gamma = 1$, and the rational quadratic has $\alpha = 2$. The dashed black line is the mean and the shaded band is the standard deviation at each x value. The colored lines are four samples from the Gaussian process plotted as smooth curves. Samples of the same color use the same $\mathbf{z}$ vector in the procedure described in section 7.3.1.1.

shown in figure 7.15a. Here, we plot the marginal likelihood against the space of data values $\mathbf{y}$, represented in one dimension on the x-axis. The measured data $\mathbf{y}_{\text{observed}}$ is a single instance in the $\mathbf{y}$ -space, shown as the dashed line. We then consider three illustrative models:

1. A simple model, shown in blue, is unable to accommodate a wide variety of data. Hence the marginal likelihood is very narrow, and evaluated at $\mathbf{y}_{\text{observed}}$ returns a very low marginal likelihood.
2. A complex model, shown in orange, can accommodate many types of data. So the marginal likelihood has a very wide distribution. However, because all probabilities must integrate to one, the overall height of its distribution is correspondingly much lower.
3. An intermediate model, shown in green, balances accuracy and complexity to achieve the highest marginal likelihood at $\mathbf{y}_{\text{observed}}$.

In the case of the isotropic kernels from section 7.3.2, the model complexity is controlled by the hyperparameters: small ℓ and large γ increase complexity. Using the RBF kernel and the same data points from section 7.3.2, we graph in figure 7.15b the marginal likelihood as we vary the two hyperparameters. A clear maximum can be seen around $\ell \approx \gamma \approx 2$. Note the sharp fall-off at large ℓ and small γ , caused because the models are too simple to accurately describe the training data. Meanwhile the fall-off at small ℓ and large γ , caused by the complexity penalty, is generally less steep.

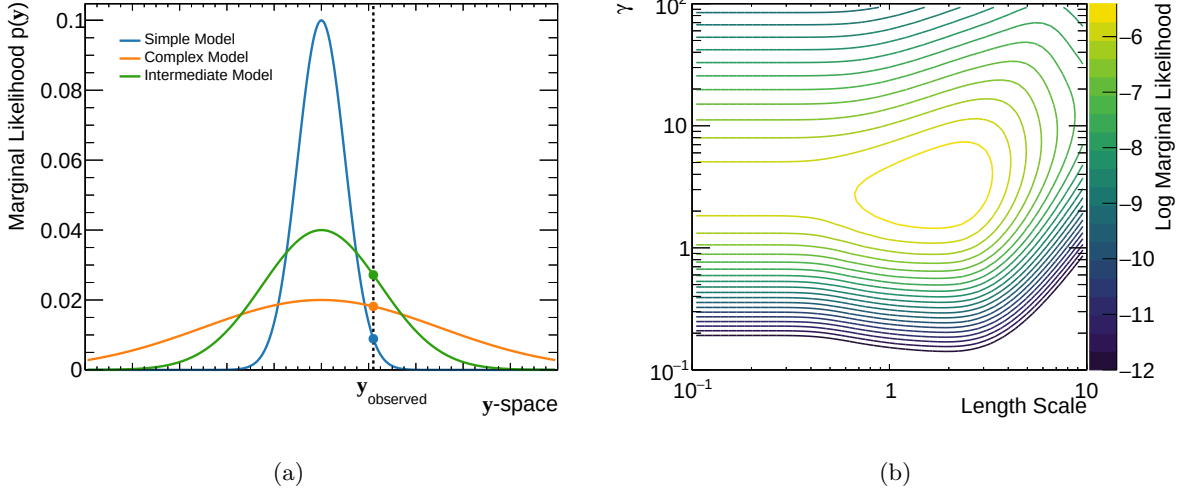


Figure 7.15: (a) Schematic explanation of why the marginal likelihood possesses a trade-off between complexity and accuracy, adapted from [116]; see text for more details. (b) Marginal likelihood contours using the setup from section 7.3.2 and the RBF kernel as the two hyperparameters are scanned.

7.4 Studies on Using Gaussian Process

Mirroring section 7.1, the performance of using Gaussian process regression was tested on W +jets MC and evaluated on several metrics including closure on background-only samples, statistical and systematic errors, and fit stability under configuration changes. The setup of the machinery follows the previous section, with the following prescriptions:

- The inputs $\mathbf{x}$ are for most of this section one-dimensional, consisting only of the fatjet mass $m(J)$. Since events are binned, the center of the bin is used as the input value. These one-dimensional fits are repeated in each bin of another variable of interest, such as $m(VV)$ or $p_T(J)$. In section 7.4.3 two-dimensional inputs are tested, where the variable of interest is added to the input vector and only one fit is performed.
- The targets y are the event counts in each bin divided by the bin width. They are normalized to unit variance and zero mean. The plots in this section show the targets rescaled back to their original values though.
- Bins with a MC statistical error greater than 40% are excluded from the fit.
- The target variances σ^2 are set to the target value, to mimic the variance of Poisson data. The MC statistical error is not used in the fit itself.
- Fits targeting a bin in $p_T(J)$ are restricted to an input space $m(J) \in [50, \min(120, p_T^{\min}/2)]$ GeV, where $p_T^{\min}$ is the lower end of the $p_T(J)$ bin, to avoid the jet-radius shoulder as mentioned in section 7.1.1.

Other fits use an input space $m(J) \in [50, 250]$ GeV. The signal region $m(J) \in [72, 102]$ GeV is excluded from the training data.

For plotting the posterior, we assign the test points $\mathbf{x}'$ to 200 equally-spaced points in the input space. When calculating the marginal posterior yield as in section 7.2.6, we use 200 equally-spaced points in the signal region $m(J) \in [72, 102]$ GeV.

7.4.1 Kernels, Binning, and ANN Warping

The difference between kernel forms was tested by fitting bins in $p_T(J)$ with both the RBF kernel, equation (7.18), and the Matern 5/2 kernel, equation (7.60). In both cases the kernels the two hyperparameters, the overall variance and the length scale, were optimized by gradient descent to the maximum marginal likelihood values. The effect of binning was also tested by using either a uniform 1 GeV bin size or a variable binning with bins between 3 and 5 GeV. Fit results are shown in figure 7.16 for a single $p_T(J)$ bin, and the complete set of fits are listed in appendix C.1.2.1.

A summary of all the fits is presented in figure 7.17. For simplicity, these results use the MMLE’s posterior, not the full marginal posterior as in equation (7.34). The choice of binning and kernel form is seen to have a very small effect. The differences are well explained by the posterior errors, which also do not vary much. The RBF kernel and a 3-5 GeV variable binning were chosen as the benchmark configuration, the former for no particular reason and the latter to better ensure that the Gaussian approximation remains valid. Note that even coarser binning is used for the highest energy bins as the amount of data becomes too small.

As with the functional fits, the strong warping caused by the ANN tagger causes the fits at low $p_T(J)$ to not perform well. This again motivates a cut on $p_T(J) > 270$ GeV or 290 GeV. Regression fits were performed against $m(VV)$ for these different $p_T(J)$ cuts, with the summary in figure 7.18. Relying only on the preselection cut of $p_T(J) > 200$ GeV reveals a strong bias to overestimate events in low $m(VV)$ bins. Both the 270 and 290 GeV cuts ameliorate this issue. A final cut of 300 GeV was chosen as a simpler and less fine-tuned threshold.

7.4.2 Marginal Likelihood Scans

The MMLE is useful for simple tests as in the previous section, but a correct treatment should use the marginal posterior. The procedure outlined in section 7.2.6 was tested on W +jets MC using an RBF kernel. This kernel has two hyperparameters, the length scale ℓ and an overall variance γ . The grid corresponding to the hyperprior was defined to be $\ell \in [10, 1000]$ GeV and $\gamma \in [10^{-1}, 10^3]$ with 50 points log-uniformly distributed in each variable.

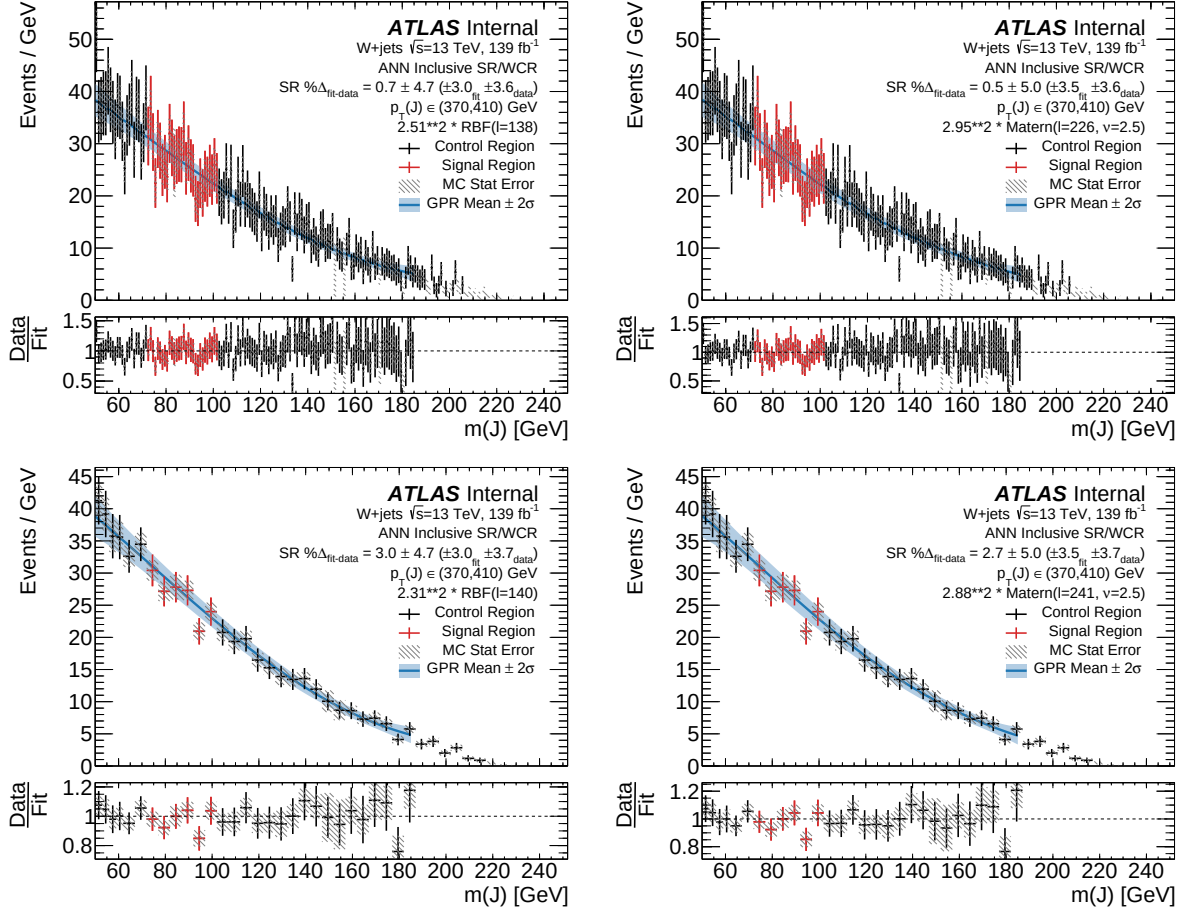


Figure 7.16: Gaussian process fit to W +jets MC with $p_T(J) \in [370, 410]$ GeV. The left plots use a RBF kernel and the right plots use a Matern 5/2 kernel. The top row uses a uniform 1 GeV binning and the bottom row uses a coarser variable binning. The black (red) points indicate the control (signal) region respectively, where the errors bars show $\sqrt{N}$ errors. The MC statistical error is shown in the gray hash instead. The blue band is the Gaussian process posterior mean $\pm 2\sigma$.

An example fit to a single $m(VV)$ bin is shown in figure 7.19. The first two plots are respectively the marginal likelihood and mean signal region yield I as functions of the hyperparameters, which are the inputs to the right hand side of equation (7.38). The marginal posterior yield $p(I|X, \mathbf{y})$ is shown in the third plot. Note that in general this distribution is not Gaussian, but appears so when the marginal likelihood is sharply peaked. The effect of estimating the θ integrals using a credibility interval, as mentioned at the end of section 7.2.6, is shown in figure 7.20. The MMLE yield is compared with the marginal posterior estimated using either a 1σ or 2σ credibility interval. As expected, convoluting the posterior over θ -space increases errors compared to the MMLE fit, though not by much. The difference between using a 1σ and 2σ threshold is quite small, indicating the approximation is robust. The 2σ threshold was chosen as the benchmark moving forward, as it suitably balances computational efficiency and accuracy.

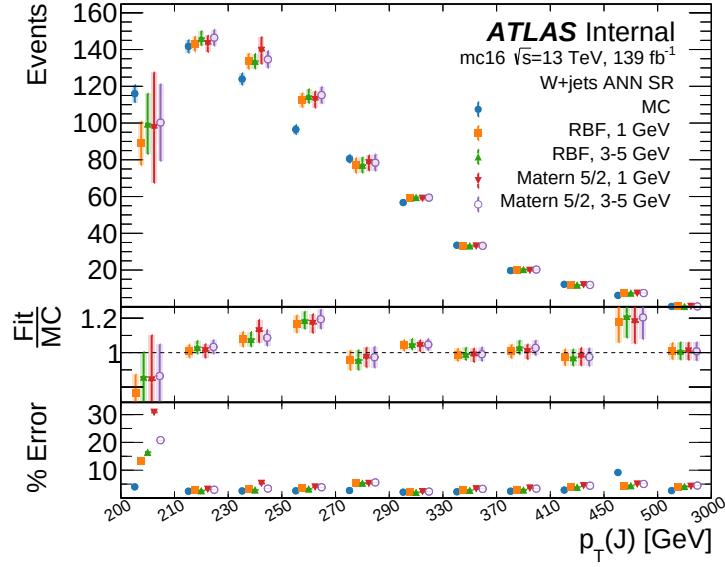


Figure 7.17: Interpolated signal region event counts of Gaussian process fits to W +jets with different kernels and $m(J)$ bin sizes, in bins of $p_T(J)$. The points show the prediction by the maximum marginal likelihood estimator, and are compared with the MC event counts. The first subplot shows the event ratio between the fits and the MC, and the second subplot shows the uncertainties as a percentage of the respective nominal values.

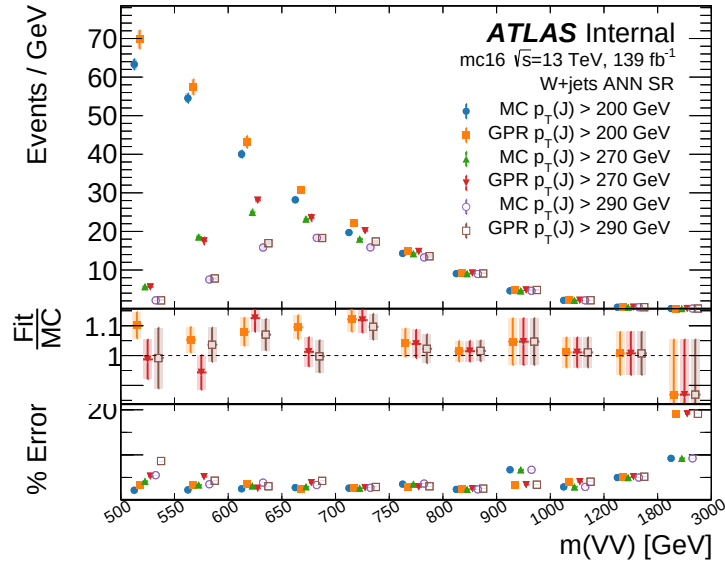


Figure 7.18: Interpolated signal region event counts of Gaussian process fits to W +jets with different cuts on $p_T(J)$, in bins of $m(VV)$. The points show the prediction by the maximum marginal likelihood estimator, and are compared with the MC event counts. The first subplot shows the event ratio between the fits and the MC, and the second subplot shows the uncertainties as a percentage of the respective nominal values.

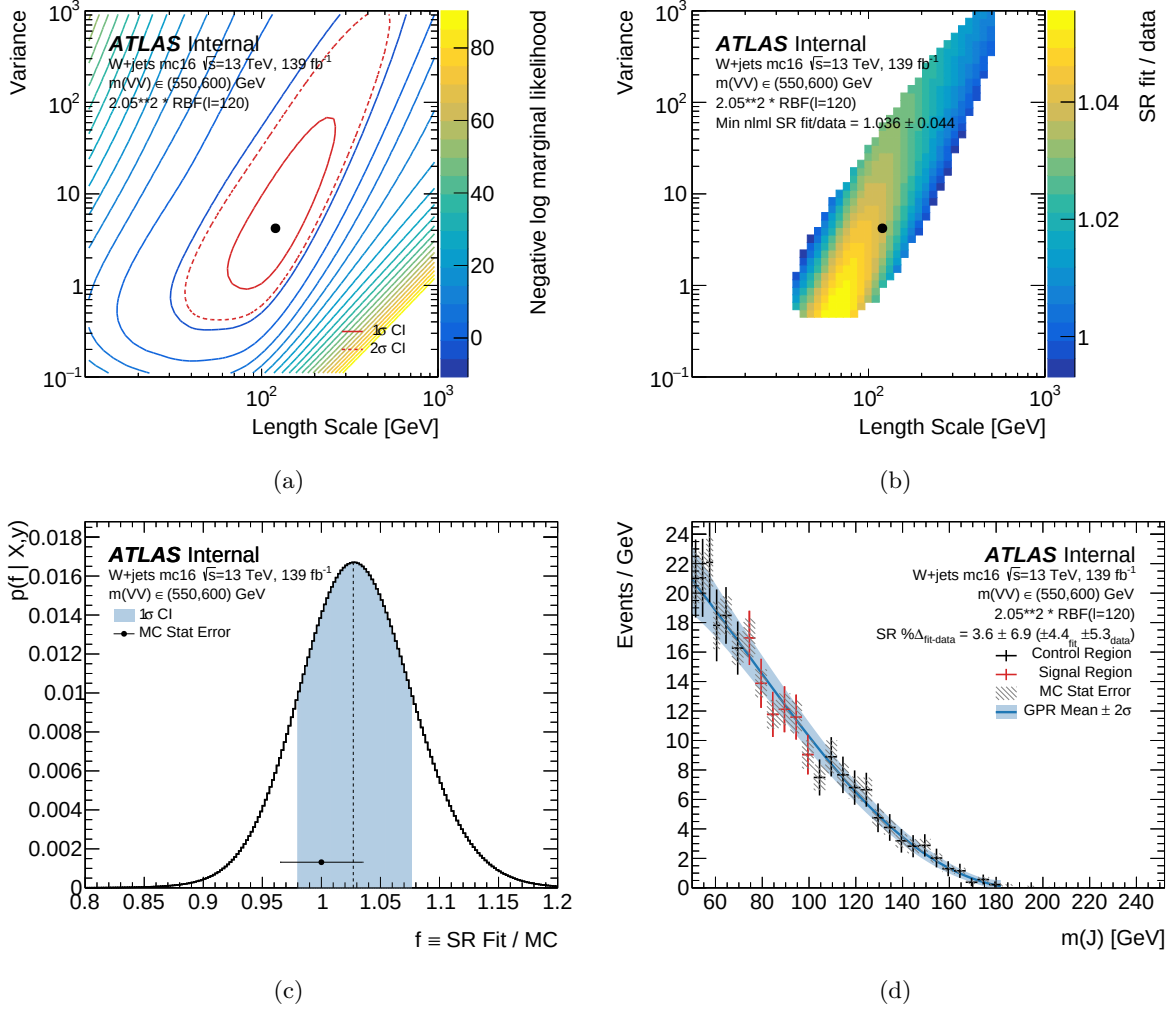


Figure 7.19: Gaussian process fit to the $m(VV) \in (550, 600)$ GeV bin, using an RBF kernel with an overall variance factor. (a) The marginal likelihood contours over the grid of hyperparameters, with a maximum at the black dot and 1σ and 2σ credibility intervals in the red lines. (b) The mean of the posterior signal region event yield along the same grid. Values are only shown in the 2σ credibility interval. (c) The marginal posterior yield distribution, with a 1σ credibility interval in blue and the MC stat error for comparison. (d) The MC data for this bin, with the MMLE fit in blue.

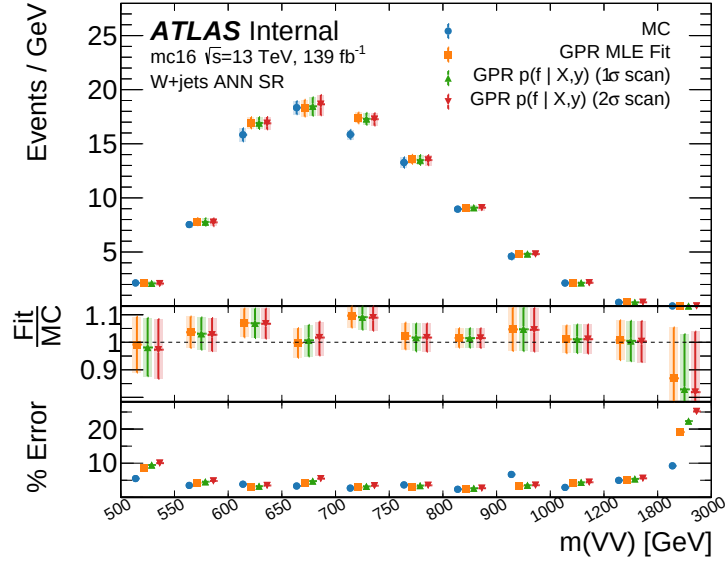


Figure 7.20: Comparison of different estimates of the Gaussian process marginal posterior yield, across bins of $m(VV)$. The orange series is the MMLE fit only, while the green and red series calculate the marginal posterior over the 1σ or 2σ credibility regions, respectively.

7.4.3 Two-dimensional Fits

In the previous sections, Gaussian process regression was applied to 1D inputs ($m(J)$) and repeated in several bins of another variable of interest, such as $m(VV)$. Instead, it may seem more intuitive to just supply the Gaussian process with both $m(J)$ and $m(VV)$ as inputs instead. However this introduces a few complications. The RBF kernel, as written in equation (7.18), couples the length scale between all input dimensions. But we expect that $m(J)$ and $m(VV)$ would scale differently. A simple adjustment is to have a separate length scale ℓ_i for each dimension,

$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp \left[\sum_{i=1}^2 -\frac{(\mathbf{x}_{1,i} - \mathbf{x}_{2,i})^2}{2\ell_i^2} \right]. \quad (7.62)$$

This however fixes the two length scales to be independent of each other, when we might expect otherwise. Indeed, in $m(VV)$ binned 1D fits, the MMLE's length scale varied between 50 to 200 GeV depending on the $m(VV)$ bin. Adding an extra hyperparameter also makes calculating the hyperprior harder, as the scan over marginal likelihoods has to be conducted over an extra dimension. Another complication is that the binning in $m(VV)$ must be finer than its expected length scale. This causes bins at larger $m(VV)$ to fall out of the Gaussian regime or even have no data, whereas the 1D fit can take advantage of wide, variable binning as in figure 7.20.

Nonetheless, a 2D fit against $m(J)$ and $m(VV)$ was studied using the modified RBF kernel in equation (7.62). The input space for $m(VV)$ was set to [500, 1500] GeV to avoid the non-Gaussian regime for simplicity, and 50 GeV bins were used. Other configurations were kept the same from the 1D fits, though

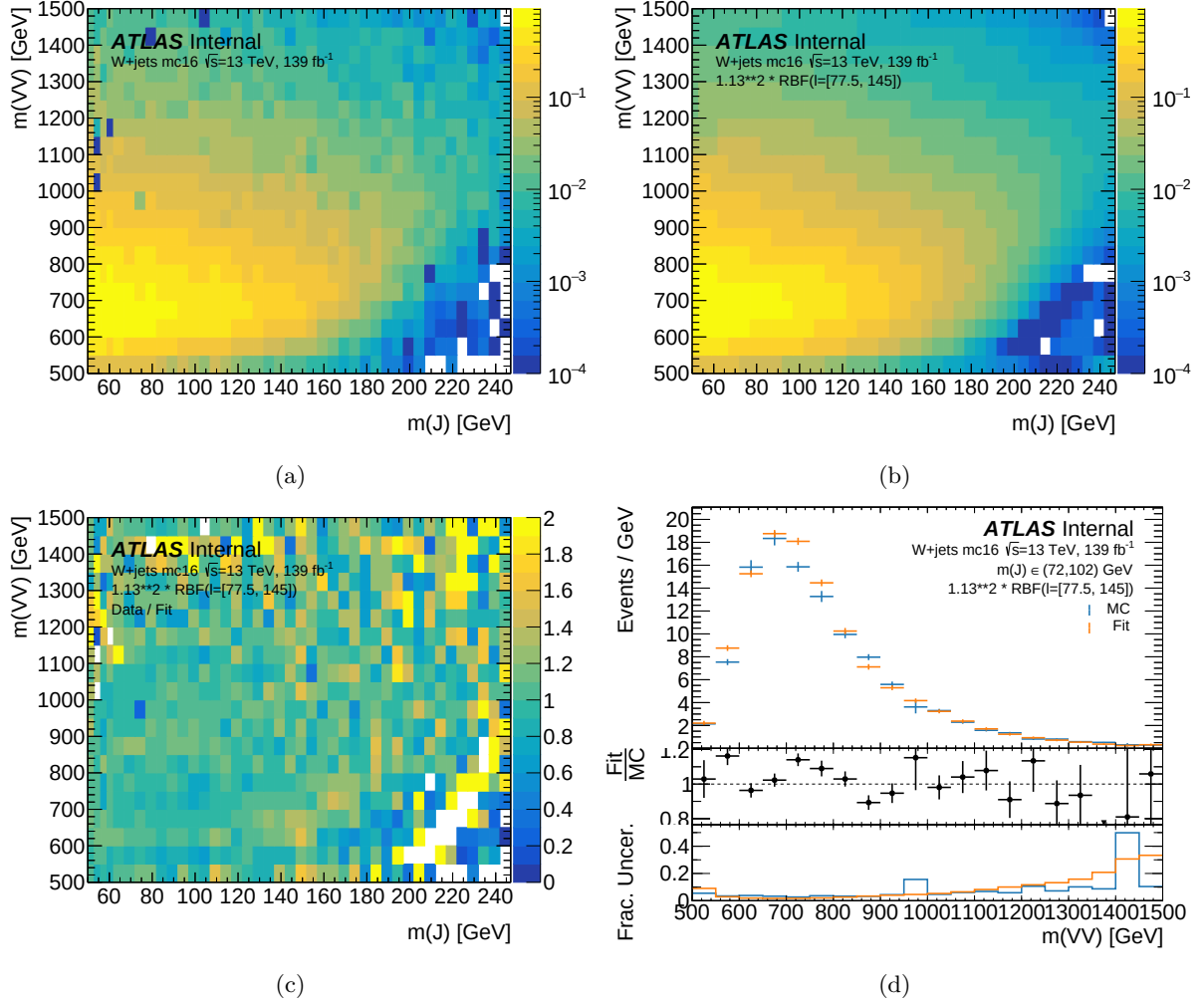
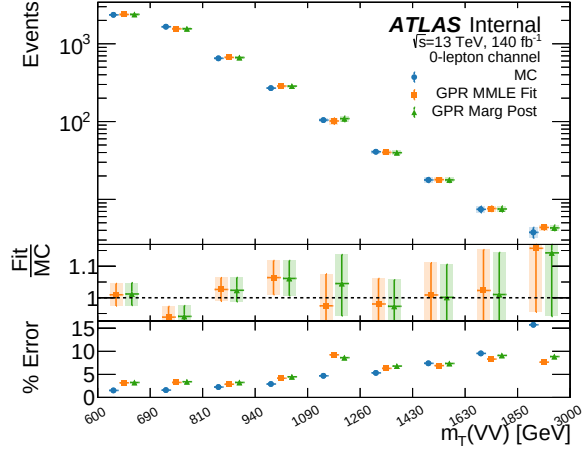
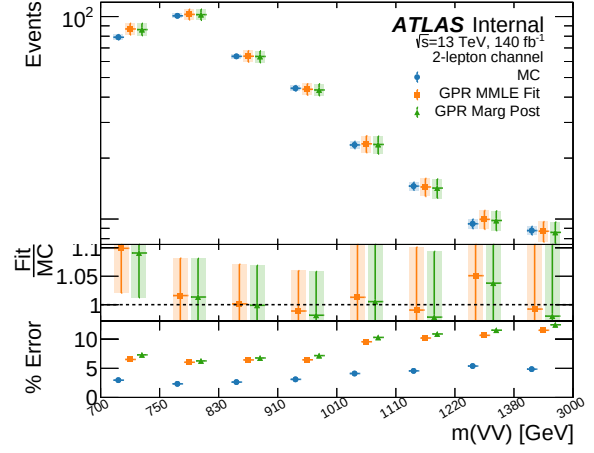


Figure 7.21: Two-dimensional Gaussian process fit against $m(VV)$ and $m(J)$ on the W +jets sideband. (a) The input MC data. (b) The Gaussian process posterior mean, evaluated at the center of each bin. (c) The ratio of MC over the posterior mean to demonstrate goodness of fit. (d) The MC and posterior's signal region yields per bin of $m(VV)$.

only the MMLE posterior was considered and not the marginal posterior. Results are summarized in figure 7.21, and a complete set of projections onto $m(J)$ are listed in appendix C.1.2.2. The predicted signal region yield generally agrees with the MC though not as well as the 1D fits, and a few $m(VV)$ bins have over a 2σ discrepancy. The fit also has similar errors compared to the 1D fits. One might have hoped that, since the 2D fit is given more data, it is better able to constrain the interpolated region; however, the effect is marginal as the posterior error is also driven by the data statistical error. Given the complications of a 2D fit and the lack of noticeable benefits, the method was dropped in favor of the 1D fits.



(a) 0-lepton channel $m_T(VV)$



(b) 2-lepton channel $m(VV)$

Figure 7.22: Summary of Gaussian process regression fits to V+jets MC in the (a) 0-lepton and (b) 2-lepton channels. Predicted signal region yields using the marginal posterior are compared to the MC yield, showing good closure. Note that the shown MC yield includes statistical errors only, and does not include systematic modeling errors.

7.4.4 Other Channels and Subtleties

The above sections focused on the 1-lepton channel and established a benchmark configuration: one-dimensional fits on $m(J)$ binned in another variable of interest, using the RBF kernel with the predicted yield taken from the marginal posterior estimated using a 2σ credibility interval. The $m(J)$ distribution is binned in variable 3–5 GeV bins, though coarser binning was optimized for bins with sparse data. The procedure was also tested in the other lepton channels on V+jets MC samples, and summary plots are shown in figure 7.22. As seen, the closure with MC is quite good. Fits to each bin are listed in appendix C.1.2.3.

An important subtlety that has been glossed over is that the real MCRs are not pure V+jets, but receives significant contributions from the top backgrounds in the 0- and 1-lepton channels, as seen in figure 6.3. There is also some more minor signal contamination, since the width of the boson $m(J)$ peak extends out of the signal region window. A method to deal with these is developed in chapter 9.

CHAPTER 8

Unfolding and EFT

The primary goal of the measurement side of the analysis is to produce a differential cross-section distribution in $m(VV)$ or $m_T(VV)$, which can then be used to constrain EFT operators. However, the detector and reconstruction algorithms do not have perfect precision and are not fully efficient in capturing every event, which will smear and distort the measured distribution. In order for this distribution to be readily comparable with theory and other experiments, or useful to scientists outside of ATLAS, these effects must be removed. This process is called *unfolding*, which takes the detector-level distribution and estimates the true underlying distribution. Note that to minimize the amount of extrapolation, the procedure usually targets a restricted truth phase space, called the *fiducial* region, which is designed by the analysis to match closely with the signal region phase space as detailed in table 6.5.

There are a plethora of unfolding methods that have been developed [122, 123]. The unfolding method chosen for this analysis is profile likelihood unfolding (PLU), which is described in section 8.1. This method can use the same likelihood fitting framework used in the resonance side of the analysis, which simplifies the software development. Studies of defining a proper fiducial phase space are then described in section 8.2. In particular, much effort was spent to ensure the fiducial region would be accurate for both SM and EFT processes. Finally, section 8.3 contains a short digression on using angular variables to maximize sensitivity to EFT effects, though these were not ultimately included in the final analysis.

8.1 Profile Likelihood Unfolding

PLU is an unfolding method most recently adopted by ATLAS $t\bar{t} + X$ measurements [124–126] and also chosen for VV+VH Semileptonic due to its simplicity and flexibility. It is a binned method that assumes a linear detector response,

$$N_i^{\text{Det}} \equiv R_{ij} N_j^{\text{Fid}}. \quad (8.1)$$

Here, N_i^{Det} is the number of events at detector-level in bin i , N_j^{Fid} is the number of events at fiducial-level in bin j , and R_{ij} is called the response matrix. In this chapter, the index i will always refer to detector-level bins, and j to fiducial-level bins. In principle the binning at detector and fiducial-level can be different, though in this analysis they are the same for simplicity.

The response matrix is determined from MC simulation of the diboson process and the detector response, and can be broken down into three heuristic components:

$$R_{ij} \equiv \frac{\epsilon_j}{\alpha_i} M_{ij} \quad (8.2)$$

$$M_{ij} = \frac{N_{ij}^{\text{Det} \cap \text{Fid}}}{N_j^{\text{Det} \cap \text{Fid}}}, \quad \epsilon_j = \frac{N_j^{\text{Det} \cap \text{Fid}}}{N_j^{\text{Fid}}}, \quad \alpha_i = \frac{N_i^{\text{Det} \cap \text{Fid}}}{N_i^{\text{Det}}}. \quad (8.3)$$

Here, $N^{\text{Det} \cap \text{Fid}}$ is the number of events that pass both detector-level and fiducial-level selections. The subscripts specify a specific detector and/or fiducial bin. M_{ij} is called the migration matrix and encodes the probability that an event passing both selections migrates from fiducial bin j to detector bin i . The efficiency ϵ_j and the fiducial accuracy α_i correct for the difference between the individual selections and their intersection. The fiducial region and binning are optimized with respect to these variables. Most importantly, for good unfolding behavior, a mostly diagonal migration matrix is desired.

A likelihood can then be defined on measured data $\vec{N}^{\text{Data}}$ via

$$L(\vec{N}^{\text{Data}} | \vec{N}^{\text{Fid}}) = \prod_i P(N_i^{\text{Data}} | R_{ij} N_j^{\text{Fid}}), \quad (8.4)$$

where the vector notation indicates the collection of all the individual bin counts, and P is the Poisson probability. Maximizing the likelihood in the usual manner results in the desired fiducial distribution $\vec{N}^{\text{Fid}}$. Effectively, PLU treats each fiducial bin as an independent signal sample, where the “signal strength” is the respective fiducial event count. The above has ignored uncertainties and backgrounds, but they can be easily included through nuisance parameters. A detailed discussion is deferred to chapter 9, but note that this is the exact same treatment as is in other maximum likelihood fits, like the one in the resonance search. Therefore the setup of the resonance fit and the PLU fit are almost identical, a major advantage of this method.

8.2 Fiducial Region Studies

The fiducial region was optimized to closely align with the detector-level selections in table 6.5. This limits the amount of extrapolation needed during the unfolding procedure, and generally produces a more precise result. However, there lacks a strong quantitative measure for determining the quality of a fiducial region. Instead, definitions were tuned qualitatively on the following criteria: kinematic distributions in detector-level and fiducial-level should be as similar as possible, the fiducial accuracy should generally be near unity, the efficiency should be as high as possible, and finally the migration matrix should be mostly diagonal. The studies below generally focus on the 1-lepton channel, as the other channels were investigated by other team members.

8.2.1 Particle versus Parton

It is important to note that fiducial regions can be defined in two separate ways. The first is by using quantities of particles that are metastable and measurable. That means defining properties on objects like jets and E_T^{miss} instead of bosons, quarks, or neutrinos. Note that pileup and detector material interactions, such as scattering, are not included, but the hadronization process is. This is called a particle-level definition. Meanwhile a definition using the direct products of the VV process, such as the bosons themselves or the two quarks of the hadronic decay, is called a parton-level definition.

Naively the parton-level definition appears simpler and more logical. However, one must remember that tree-level diagrams such as figure 6.1 are only simplistic representations, and not totally representative of the physical process. These diagrams receive a myriad of quantum corrections including the addition of loops and initial and final state radiation. Moreover, all Feynman diagrams actually interfere with each other. Therefore it is not entirely proper to refer to intermediate propagators, such as the bosons or the initial decay quarks, as independent physical objects. Only the final state particles that are measured have a precise quantum mechanical interpretation. Still, many MC simulators rely on generating some representation of the intermediate propagators, which are included in the output truth record. But the way these records are stored and their interpretation varies between each generator, as there is not a well-defined physical interpretation.

So while a parton-level definition is in principle possible to implement, most analyses favor a particle-level definition. A particle-level definition was also chosen for this analysis. This however introduces the potential for errors similar to detector-level reconstruction, including selection inefficiencies and parameter resolution. For example, a particle-level definition using $R = 1.0$ jets may not fully capture the showered decay of a hadronic boson, leading to a skewed p_T value. This is an additional subject of consideration when designing a fiducial region. Note that some analyses use both particle- and parton- level definitions, defining two fiducial regions, such as the ATLAS $t\bar{t}Z$ measurement [124].

8.2.2 Benchmark Definition

An initial benchmark fiducial definition in the 1-lepton channel was created by mirroring the subset of detector-level selections that could be easily applied to MC truth quantities. First, truth-level objects in the MC truth record are chosen and filtered as in section 6.2. Cuts on p_T , η , and m from tables 6.3 and 6.4 are copied directly to the truth-level objects. Similarly, the kinematic cuts from table 6.5 were also copied directly, and shown in table 8.1. In particular, this excludes the two neural network taggers, the ANN W/Z -tagger and the DL1r b -tagger, which are ATLAS specific and trained on detector-level quantities, not truth-level ones. Kinematic distributions for a wide swath of variables was compared between the fiducial and detector selections, and verified to be generally consistent. The 4-vector quantities for the VV system

Object	0-lepton Channel	1-lepton Channel	2-lepton Channel
Charged Lepton Multiplicity	None	Exactly one	Leading two and same flavour
V_ℓ Candidate	$E_T^{\text{miss}} > 250 \text{ GeV}$ $E_T^{\text{miss}}/H_T > 10$	$E_T^{\text{miss}} > 100 \text{ GeV}$ $p_T(\ell) > 30 \text{ GeV}$ $p_T(V_\ell) > 200 \text{ GeV}$	leading lepton $p_T > 27 \text{ GeV}$ m in Z window as in table 6.5
V_h Candidate	Leading large- R jet $p_T > 300 \text{ GeV}$ $72 \text{ GeV} < m < 102 \text{ GeV}$		
Topology	$\Delta\phi(V_h, E_T^{\text{miss}}) > 0.2$	$p_T(V) > 0.35m(VV)$ $E_T^{\text{miss}} > 0.2p_T(V_\ell)$	$p_T(V) > 0.35m(VV)$

Table 8.1: Fiducial region event selection cuts in all three lepton channels.

are shown in figure 8.1. Note that at high $p_T(VV)$ the fiducial distribution tends to bias much higher, which manifests in the lowest $m(VV)$ bin, which partially motivated moving the $m(VV)$ boundary of the unfolding to 600 or 700 GeV.

Next, the efficiency and fiducial accuracy from equation (8.3) were examined. These are plotted in figure 8.2 for $m(VV)$ and $p_T(VV)$. The efficiency averages around 0.3, reflecting the tightness of the analysis cuts used to reduce background, in particular the ANN tagger filter which has a set working point of 50% efficiency. Meanwhile the fiducial accuracy is at an acceptable average of around 0.8. Note that the fiducial accuracy always falls in the lowest $m(VV)$ bins because of the cut at 500 GeV and other correlated cuts, like those on $p_T(J)$ and $p_T(W)$. Events with fiducial $m(VV) < 500 \text{ GeV}$ will often get smeared to a detector $m(VV) > 500 \text{ GeV}$, but these events do not pass the fiducial selection and therefore decrease the accuracy. The inverse effect also causes the efficiency to drop at the lowest $m(VV)$. This is an unavoidable effect as long as a minimum or maximum cut is placed; notice that this falloff does not appear in $p_T(VV)$ since all values are accepted and the other selections are generally uncorrelated with $p_T(VV)$. Otherwise, the two quantities are generally independent of $m(VV)$, which is a good sign.

8.2.3 Efficiency Discrepancy

A somewhat more major problem is that the efficiency of the SM diboson process is noticeably lower than the efficiency of the EFT operator c_W . Note the latter is a MC simulation using only the pure quadratic EFT contribution, as discussed in section 2.3.2. This will cause the two samples to have different response matrices, but the analysis has no good method to distinguish the SM and EFT processes. If only the diboson response matrix were used, any EFT contributions, should they exist, would be unfolded incorrectly. This discrepancy was therefore thoroughly investigated in the hope of rectifying it.

After many refuted hypotheses and unsuccessful studies, it was noted that the efficiency in $p_T(VV)$ is actually identical between the two processes in the lowest bin, $p_T(VV) < 50 \text{ GeV}$, and the discrepancy

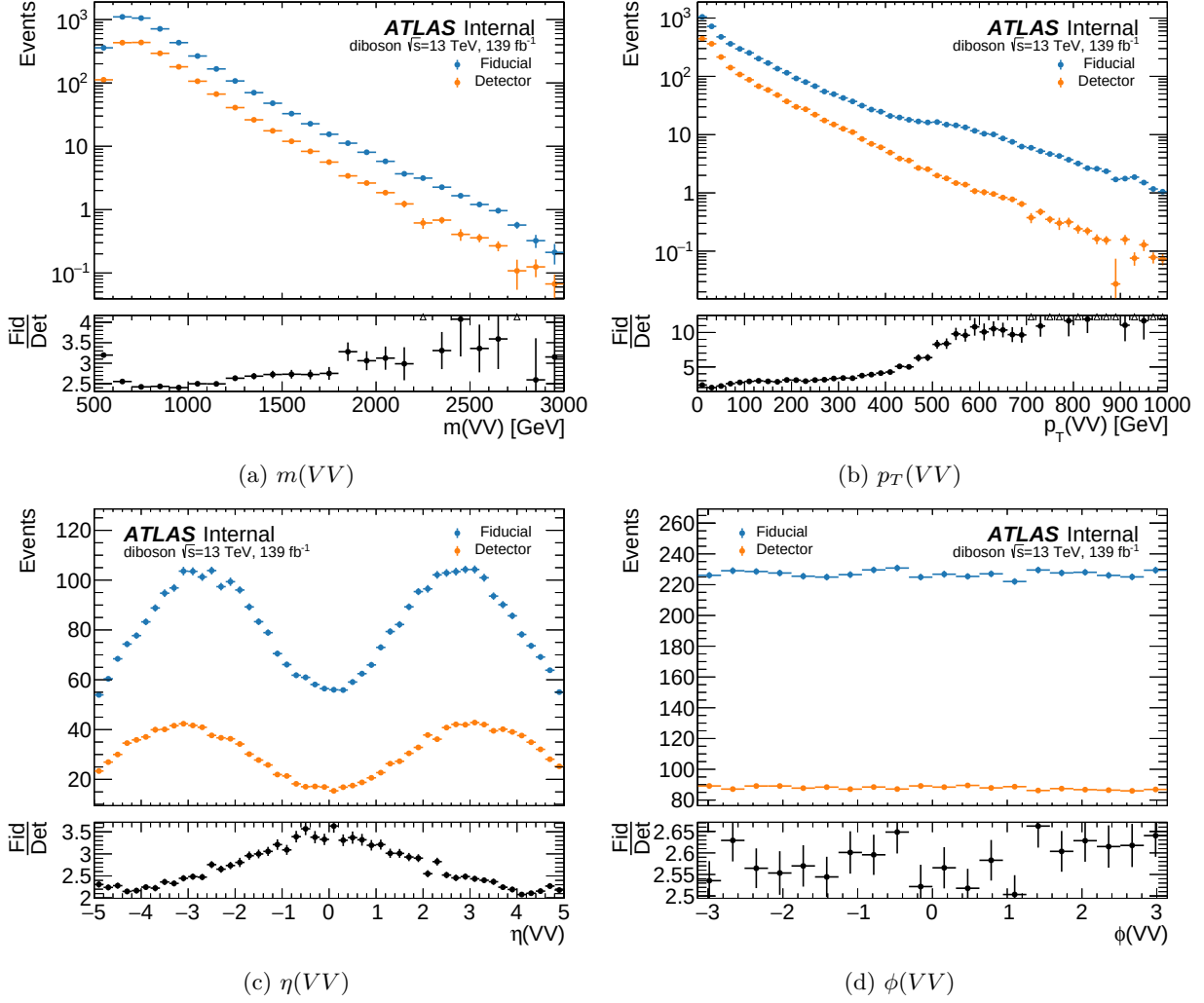


Figure 8.1: Comparison between fiducial and detector-level kinematic distributions of the VV system in the 1-lepton channel, using diboson MC samples.

increases with $p_T(VV)$, as seen in figure 8.2b. A cut-flow graph binned in the true, parton-level $p_T(VV)$, shown in figure 8.3, was created to examine the cause. One clearly sees that the diboson sample tends to fail the fatjet selection more. That is, many events do not contain any fatjet with the selections listed in table 6.4. This happens more frequently at the higher $p_T(VV)$ bins. Intuitively, this may not be too surprising. If the hadronic boson is produced in the direction opposite the boost of the VV system, $p_T(V_h)$ will be proportionally reduced by $p_T(VV)$. As $p_T(VV)$ becomes large relative to $p_T(V_h)$ in the VV rest frame, the decay products of the jet may no longer be collimated enough to be reconstructed in a fatjet. On the other hand, if the bosons are produced orthogonally to the VV boost direction, high $p_T(VV)$ may cause the two bosons to become collimated and their products to overlap. The jet may then be discarded via an overlap removal algorithm that favors the lepton. And even if not, resolutions of the respective elements

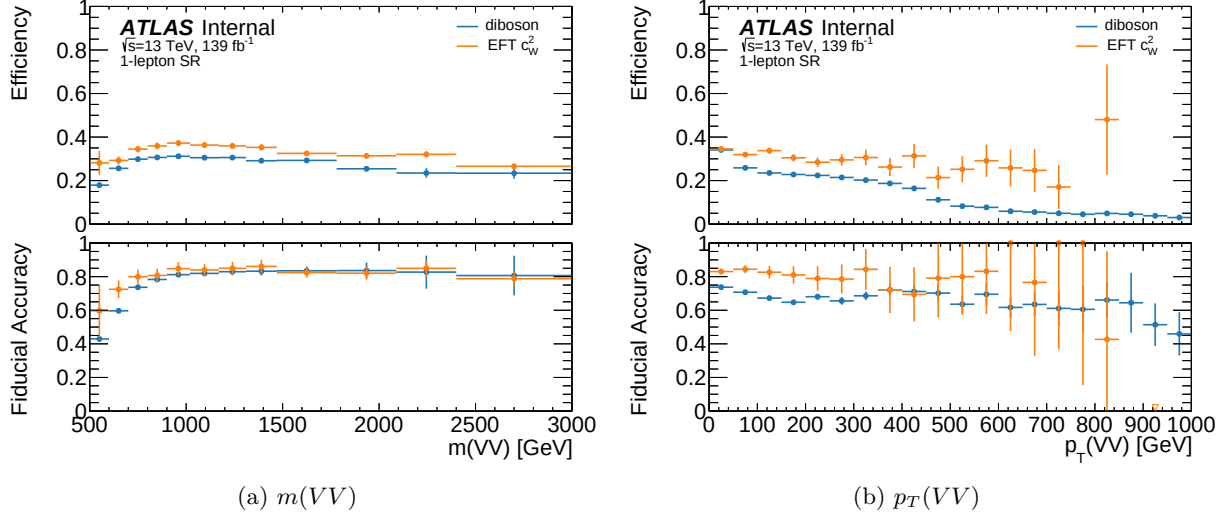


Figure 8.2: Efficiency and fiducial accuracy in the 1-lepton channel as functions of (a) $m(VV)$ and (b) $p_T(VV)$. These are compared between the SM diboson process and the pure quadratic EFT contribution of c_W .

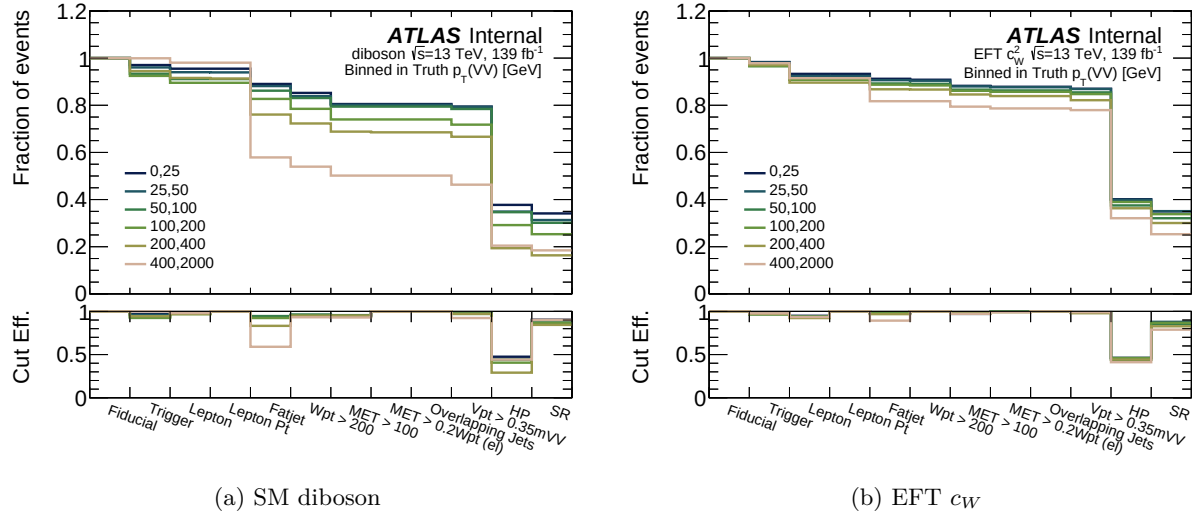


Figure 8.3: Cut-flow plot of the efficiency for (a) diboson and (b) c_W binned in truth (parton-level) $p_T(VV)$. The plot should be read starting from the leftmost bin, which represents all events passing the fiducial selection. Moving right, the cuts from table 6.5 are sequentially applied. The y-axis shows the fractional number of events passing each respective cut. The efficiency can be read off the value in the last bin.

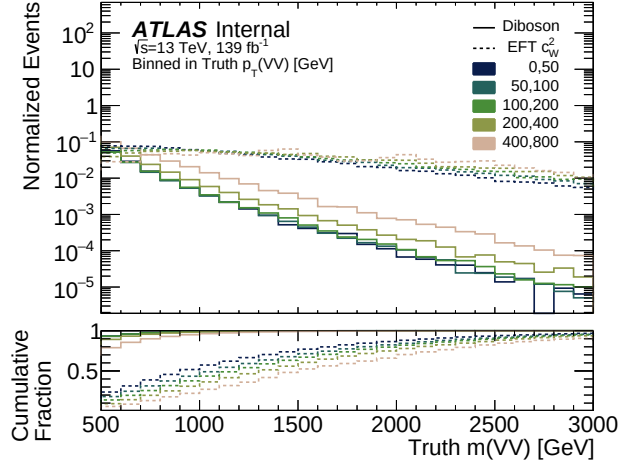


Figure 8.4: A comparison of the $m(VV)$ distribution in bins of $p_T(VV)$ between the SM diboson process and the EFT c_W operator. The main plot series are normalized to unit integral. The subplot shows the cumulative fraction of events moving from left to right. Quantities are at parton-level, taken from the MC truth record.

would be degraded, increasing the likelihood that the jet fails the mass or p_T cuts of table 6.4.

It was hypothesized that the EFT process avoids this effect because it scales with energy, so $m(VV)$ is generally much larger, as seen in figure 8.4. This directly corresponds to a larger $p_T(V_h)$ in the VV rest frame, perhaps enough to avoid the effect of $p_T(VV)$ in the range relevant to the analysis. But for a given $p_T(VV)$ and $m(VV)$, it is expected that the SM and EFT processes should have the same efficiency. This was verified by examining the efficiency as a function $p_T(VV)/m(VV)$, as seen in figure 8.5a. As desired, the efficiency agrees quite well between the two processes, and as expected falls dramatically as $p_T(VV)$ becomes comparable to $m(VV)$. A second verification examined the ΔR between the two bosons at parton-level, shown in figure 8.6. At low $p_T(VV)$, both SM and EFT samples are sharply peaked around $\Delta R = \pi$, indicating the two bosons are back-to-back. However, as $p_T(VV)$ increases, the bosons of the SM process begin to be very collimated, with ΔR shifting towards 0. A much less pronounced effect is seen in the EFT sample, which still clusters near $\Delta R = \pi$.

As the relative strengths of the EFT and SM processes are unknown, correcting for this effect is not straightforward. In addition, because the analysis is targeting a particle-level fiducial definition, one cannot cut on parton-level quantities. Otherwise, one could cut on the parton-level $p_T(VV)/m(VV)$ to a region of mostly uniform efficiency, discarding the events that tend to be mismeasured. However, the particle-level equivalent is not as effective, as seen in figure 8.5b. Since a fatjet needs to be identified, the particle-level selection suffers the same problems as the detector-level ones, where the troublesome events do not produce a suitable fatjet from the hadronic decay. Some of these events still end up passing the fiducial selection though by picking up an unrelated extraneous fatjet, and produces a mostly random $p_T(VV)/m(VV)$.

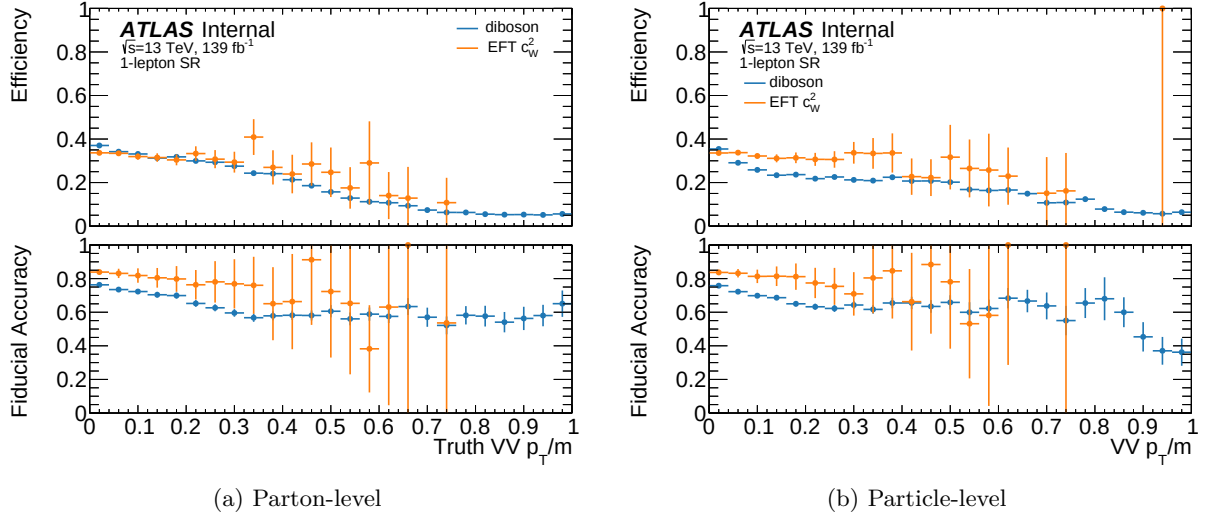


Figure 8.5: Efficiency and fiducial accuracy as functions of $p_T(VV)/m(VV)$ at (a) parton-level and (b) particle-level. The parton-level quantities represent the “true” event, ignoring quantum subtleties, while the particle-level uses the selections in table 8.1. In particular, the latter can falsely identify an extraneous fatjet as V_h , leading to the efficiency loss shown.

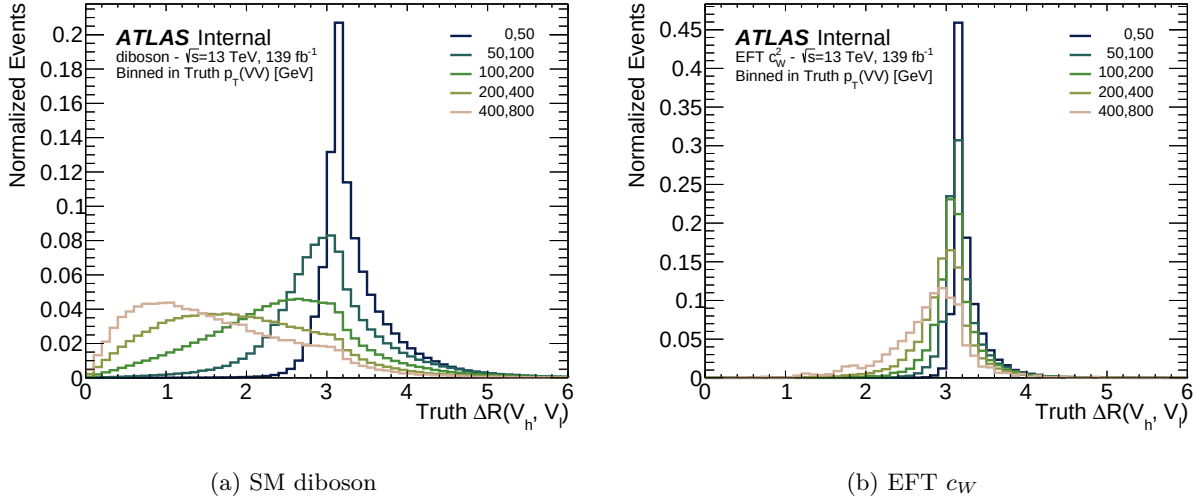


Figure 8.6: The ΔR between the two bosons in bins of $p_T(VV)$ for (a) diboson and (b) c_W . These values are all parton-level quantities taken from the MC truth record.

Most will still fail the detector selections though, leading to decreased efficiency. However, the smallest bins of $p_T(VV)$ or $p_T(VV)/m(VV)$ seem to be unaffected by the mis-selection mentioned above, producing consistent efficiencies still. A cut of $p_T(VV) < 50$ GeV in both detector and fiducial selections was considered. This however is a very tight cut, and would discard around half the data in all $m(VV)$ bins. A few other ideas were tested but did not produce satisfactory results. Ultimately, the analysis team decided to leave the efficiency discrepancy as is. It represents a legitimate difference between the two processes, and should be a minimal effect given that the EFT contribution is expected to be small. EFT coefficient fits will be conducted using the detector-level distribution to avoid this uncertainty.

8.2.4 Optimized Binning

The final component of the response matrix is the migration matrix. This is the most important component during the actual likelihood fit, as the matrix is effectively being inverted. A migration matrix with large off-diagonal components will cause significant problems with correlations between neighboring bins. Therefore, the binning was algorithmically optimized to maintain a mostly diagonal shape. First, a migration matrix is created using very fine uniform binning, and the first bin is taken as a seed. Subsequent fine bins are then iteratively merged into the seed until all the following criteria are met:

- The diagonal component is greater than 70% or 80%.
- The statistical error in the diagonal bin is less than 20%.
- The number of detector-level events is greater than 10.
- The current bin is larger than any previously merged bin.

The next fine bin is taken as a seed for the next merged bin, and the process is repeated. This process was tested and used on all three lepton channels, and the final migration matrices are shown in figure 8.7. The algorithm produces suitable results with sufficient numbers of bins that are intuitively spaced.

8.3 Interference Resurrection

As noted in section 2.3.2, the leading order EFT effect is actually an interference term between the SM and EFT operators. It turns out that in the diboson process, this interference term is actually greatly suppressed due helicity selections [22]. The SM mostly produces dibosons in opposite-sign transverse polarizations, while EFT operators tend to favor longitudinal or same-sign transverse polarizations. The energy scaling of each polarization mode in the W^+W^- process is shown in table 8.2. However, by focusing on specific polarization states, it is possible to resurrect the interference somewhat [127].

Experimentally, this means measuring differential distributions in specific angular variables. As described in reference [127], in the 1-lepton channel the variable of greatest interest is the azimuthal angle of the lepton

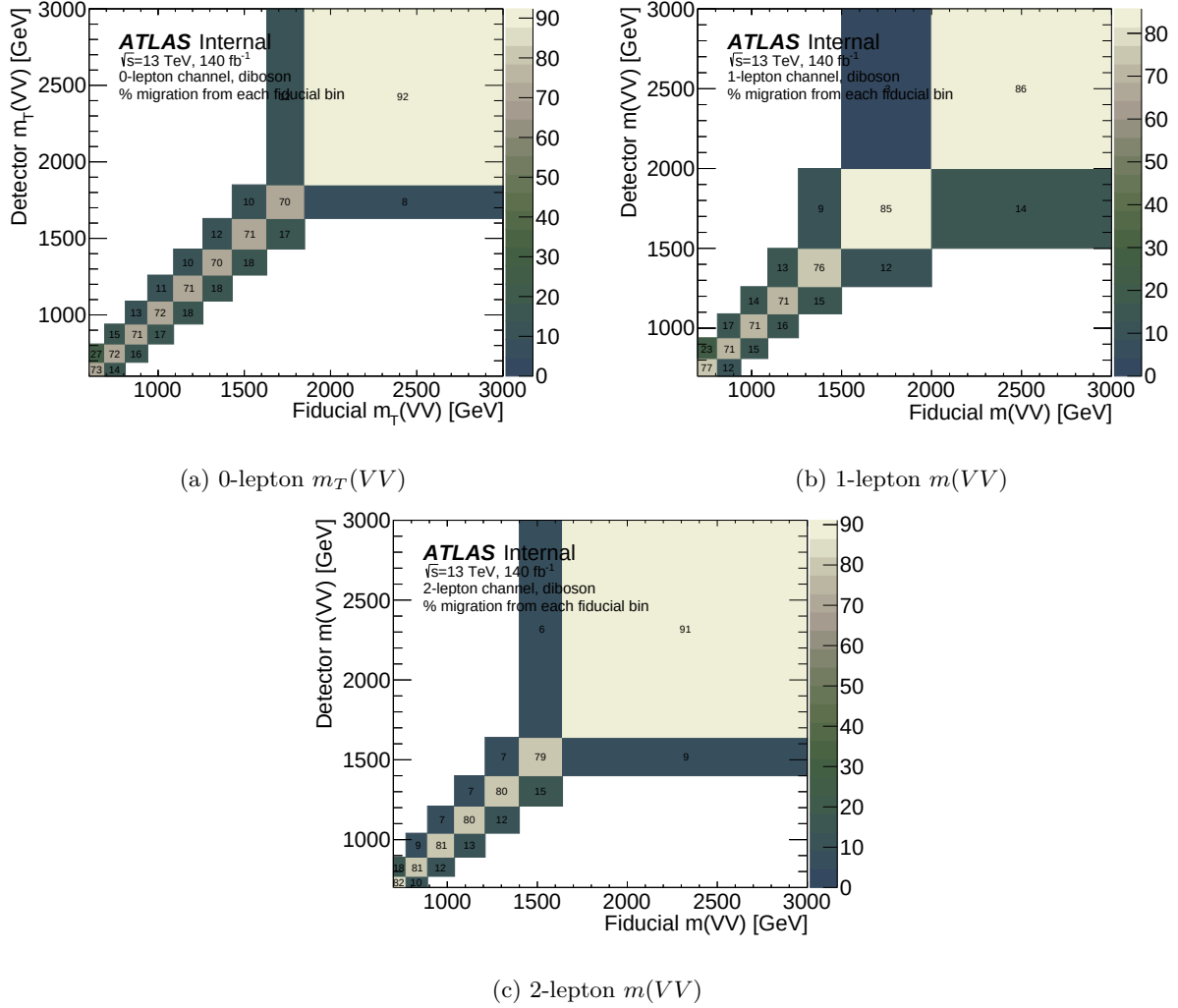


Figure 8.7: Migration matrices after using the optimized binning procedure in all three lepton channels.

(h_{W^+}, h_{W^-})	SM	O_W	O_{HW}	O_B	O_{HB}	O_{3W}
$(\pm, \mp)$	1	0	0	0	0	0
$(0, 0)$	1	$\frac{E^2}{\Lambda^2}$	$\frac{E^2}{\Lambda^2}$	$\frac{E^2}{\Lambda^2}$	$\frac{E^2}{\Lambda^2}$	0
$(0, \pm), (\pm, 0)$	$\frac{m_W}{E^2}$	$\frac{Em_W}{\Lambda^2}$	$\frac{Em_W}{\Lambda^2}$	$\frac{Em_W}{\Lambda^2}$	$\frac{Em_W}{\Lambda^2}$	$\frac{Em_W}{\Lambda^2}$
$(\pm, \pm)$	$\frac{m_W}{E^2}$	$\frac{m_W}{\Lambda^2}$	$\frac{m_W}{\Lambda^2}$	$\frac{m_W}{\Lambda^2}$	0	$\frac{E^2}{\Lambda^2}$

Table 8.2: Energy scaling in different helicity modes for the $q_L \bar{q}_R \rightarrow W^+ W^-$ process in the SM and for several EFT operators [90]. Note these operators use the SILH basis [128, 129], which are slightly different than the Warsaw basis [20] listed in appendix A. However O_{3W} above is directly comparable with O_W in the Warsaw basis. See references [130, 131] for a comparison of the other operators.

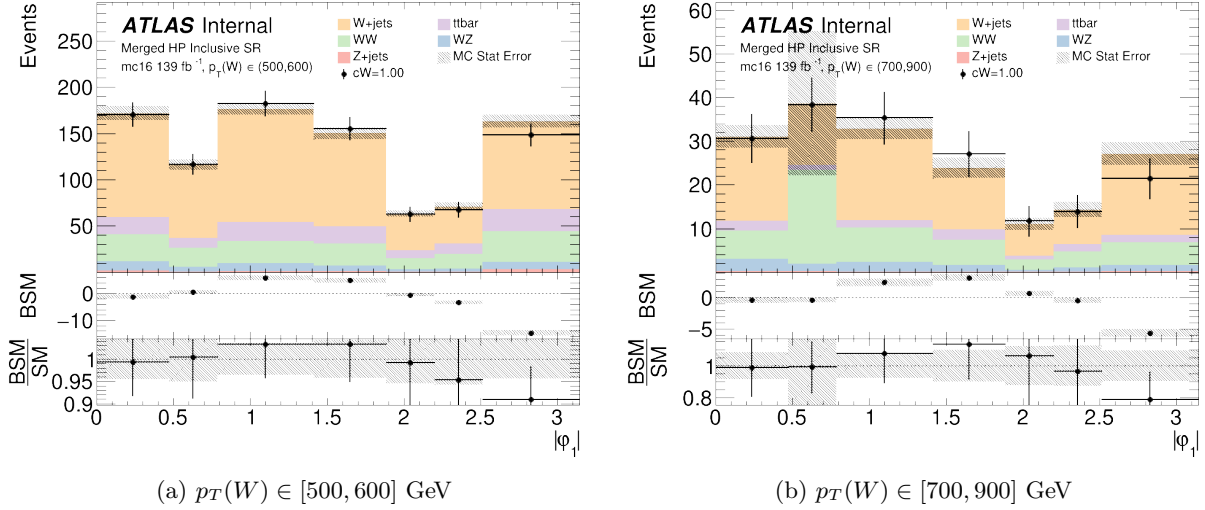


Figure 8.8: Distributions of $|\varphi_1|$ in two bins of $p_T(W)$, showing the expected effect of the c_W operator. The colored series are the SM processes, while the c_W operator is stacked on top as the points, with error set to $\sqrt{N}$ to represent asimov data. Here, c_W is set to $1/\text{TeV}^2$ and only the effect of the linear interference component is included. The first subplot shows the unstacked effect of the c_W operator, while the final subplot shows a ratio versus the SM-only distribution. Note that in the second bin of figure (b), an MC event with an anomalously large weight causes the large spike in WW .

in the VV rest frame. Here, the z -axis is defined in the direction of the leptonic W 's momentum, so in the rest frame the two bosons are back-to-back along the z -axis. The x -axis is then defined such that the lab frame VV momentum lies completely in the $x - z$ plane. This angle is denoted φ_1 . Detector-level distributions of $|\varphi_1|$ are shown in figure 8.8 for two $p_T(W)$ bins. The effect of the c_W operator is included by using only the linear interference component and fixing $c_W = 1/\text{TeV}^2$. Note that binning in $p_T(W)$ or some other energy-related variable is required as the SM processes dominate the lower-energy bins and drown out the interference effect, but also because the EFT effect tends to grow with energy. Note also that the absolute value of φ_1 is taken to increase sensitivity since the distribution is symmetric.

Unfortunately, the interference effect is still seen to be quite small relative to the MC statistical errors, mostly due to the large V +jets background. For this level of precision, a fully leptonic search may be more optimal. During the development of this analysis, CMS published a result also measuring φ_1 , using instead the $W\gamma$ channel (with $W \rightarrow \ell\nu$) [102], which suffers a similar interference suppression. Their result demonstrates the tremendous benefit of using polarization measurements, but also likely exceeds what could be achieved with this semileptonic VV analysis, at least for the linear interference term. Given then the lack of sensitivity seen here, it was decided not to pursue a measurement in $|\varphi_1|$.

CHAPTER 9

Fit Procedure and Results

Putting everything from the previous chapters together, a full fitting pipeline can finally be implemented. Section 9.1 discusses the procedure used, including a description of the likelihood fit and a method to incorporate the Gaussian process regression. Section 9.2 then tests the full pipeline using simulated data, often called *Asimov data*, to validate the fitting methods and determine expected limits. The use of Asimov data blinds the analysis, preventing the analyzers from accidentally introducing personal biases. Once the fit methods have been validated and frozen, the analysis can unblind, and real data can finally be examined. Unfortunately, at the time of publication of this dissertation, the analysis had yet to receive unblinding approval from ATLAS. Therefore, final results have to be omitted. However, the Asimov tests are fairly decent proxies for what the expected results will look like, assuming no BSM effects are present, and demonstrate the tremendous sensitivity of this analysis to the signal process and certain SMEFT operators.

9.1 Fit Procedure

All the main analysis components, the resonance search, the unfolding, the SM diboson cross-section measurement, and the EFT coefficient measurements, use the same profile likelihood framework commonly used in high-energy physics [132]. This method defines a likelihood of observing binned data $\vec{n}$ as a function of various parameters. These parameters are split into parameters of interest $\vec{\mu}$, usually signal strengths, and nuisance parameters $\vec{\theta}$, which are included to reflect uncertainties in the analysis. By maximizing the likelihood over the parameter space, the best-fit values of the parameters of interest can be determined, along with confidence intervals on the result. These are the final results of the analysis.

Common to all analysis components is using MC to estimate the $t\bar{t}$ and single top backgrounds. However, for the V+jets background, the unfolding and diboson cross-section measurement will use the data-driven Gaussian process regression method, while the resonance search will use MC, as it splits into dozens of signal channels, many with very low statistics, that would be impractical to fit with Gaussian process. Before discussing the likelihood method though, an important subtlety of the Gaussian process regression must be addressed. The Gaussian process studies from chapter 7 assumed that the MCR was purely V+jets, but the real MCR contains significant contributions from the top backgrounds as well as some small signal contamination, as seen in figure 9.1. The sideband-based method cannot be used with the top backgrounds

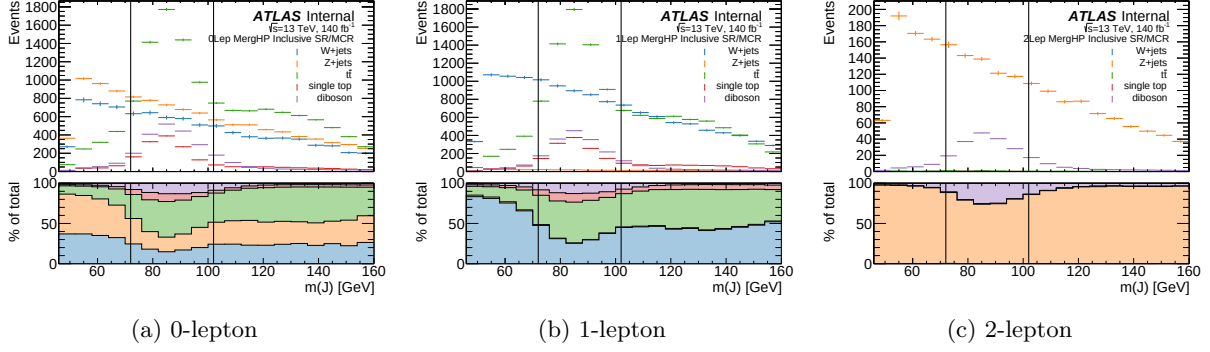


Figure 9.1: MC distributions of $m(J)$ for all three lepton channels in the signal region and MCR. The vertical bars represent the signal region window, and the subplot shows the percent contribution of each sample.

contaminating the region because they produce a resonance at the W mass.

9.1.1 Integrating the Gaussian Process

In order to isolate the V +jets contribution, a background subtraction method was implemented. Given data in the MCR, the V +jets distribution is obtained by subtracting out the expected $t\bar{t}$, single top, and diboson contributions using MC simulation,

$$N_{V+jets} = N_{\text{data}} - \mu_{t\bar{t}} N_{t\bar{t}}^{\text{MC}} - \mu_t N_t^{\text{MC}} - \mu_{VV} N_{VV}^{\text{MC}}. \quad (9.1)$$

Here, the μ parameters represent respective sample strengths that scale the nominal MC prediction. Note that this method relies on the MC predictions for the three other samples, so the method is no longer purely data driven. However, the MC for these samples are generally much more accurate than that of V +jets, and are used as the background templates in the likelihood fit anyways. They mostly err on their normalizations, hence the inclusion of the μ scaling parameters. First, the single top sample strength is derived from previous ATLAS measurements [133, 134]. Then the $t\bar{t}$ sample strength is fixed by a single-bin fit to the 1-lepton channel's TCR. Notably the TCR is almost purely top backgrounds as seen in figure 6.3, so the small V +jets and diboson contamination are left at their nominal MC value. Finally special care must be taken to handle the signal contamination. This is detailed momentarily, but for now the procedure progresses using the nominal value, with μ_{VV} set to unity.

Now given an presumably isolated V +jets distribution, the Gaussian process regression can be used following the final procedures in chapter 7. However, uncertainties from the above procedure need to be propagated to the Gaussian process result. These include theory uncertainties and detector reconstruction uncertainties, which affect the MC templates and are often referred to as *systematic uncertainties*, but also importantly the uncertainties on μ_t and $\mu_{t\bar{t}}$. The systematic uncertainties are commonly propagated by using $\pm 1\sigma$ variations. That is, given a parameter α with a nominal value and uncertainty $\alpha_0 \pm \sigma_\alpha$, the

entire analysis workflow is repeated for $\alpha = \alpha_0 - \sigma_\alpha$ and $\alpha = \alpha_0 + \sigma_\alpha$. This produces two new histograms for every source of systematic error, both for the MC samples as well as the Gaussian process result. The uncertainties on μ_t and $\mu_{t\bar{t}}$ are similarly handled. Both the TCR fit and the Gaussian process regression are repeated for the $\pm 1\sigma$ variations on μ_t , and then the Gaussian process is repeated for the $\pm 1\sigma$ variations on $\mu_{t\bar{t}}$. These variation histograms are included in the final likelihood fit via nuisance parameters, discussed further in section 9.1.2.

To handle the correlation with μ_{VV} , which is a freely floating parameter of interest, a different approach must be used. Given that in equation (9.1) a nominal value of $\mu_{VV} = 1$ is assumed, signal contamination occurs when $\mu_{VV}^{\text{truth}} > 1$, as not all of the diboson contribution is subtracted out. Inversely, a signal deficit is introduced if $\mu_{VV}^{\text{truth}} < 1$. The Gaussian process regression procedure was therefore tested again on V+jets MC, but now added with diboson MC scaled by $\mu_{VV}^{\text{truth}} - 1$ for various test values of μ_{VV}^{truth} . The effect on the predicted yield was determined as a function of μ_{VV}^{truth} , shown in figure 9.2. It was observed that the Gaussian process yield had an almost perfectly linear dependence to μ_{VV}^{truth} ,

$$N_{\text{GPR}}^{\mu_{VV}} = N_{\text{GPR}}^{\mu_{VV}=1} [1 + c \cdot (\mu_{VV}^{\text{truth}} - 1)], \quad (9.2)$$

where the coefficient c is determined for each bin, and can be read off the subplots of figure 9.2. To reemphasize, $N_{\text{GPR}}^{\mu_{VV}=1}$ is the fit result when no signal contamination is introduced, and represents the true V+jets contribution. Meanwhile $N_{\text{GPR}}^{\mu_{VV}}$ is the fit result with signal contamination, and represents the erroneous result that would occur when μ_{VV}^{truth} is not equal to 1. Therefore by inverting equation (9.2), a correction factor can be obtained to rectify the naive yield obtained from assuming $\mu_{VV} = 1$ in equation (9.1),

$$N_{\text{GPR}}^{\text{corrected}} = N_{\text{GPR}}^{\text{naive}} / [1 + c \cdot (\mu_{VV} - 1)]. \quad (9.3)$$

This correction factor scales the naive yield as μ_{VV} is scanned in the final likelihood fit.

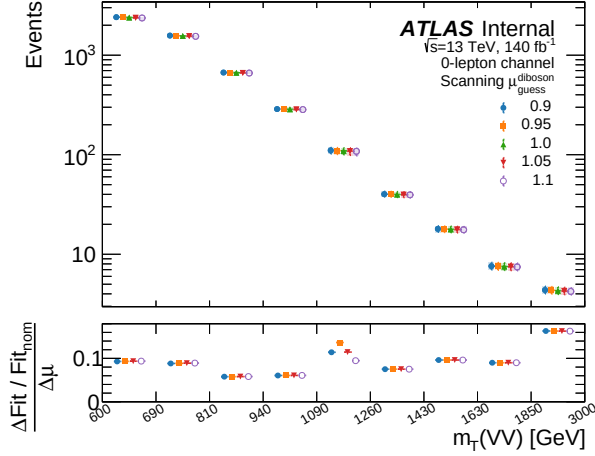
9.1.2 Likelihood Construction

With the Gaussian process method set, the likelihood function for the analysis can now be constructed. This section follows the procedure commonly used in high-energy physics [132], and initially focuses on the diboson cross section measurement for simplicity. The main building block of the likelihood is the Poisson probability of measuring n events given an expected mean of λ ,

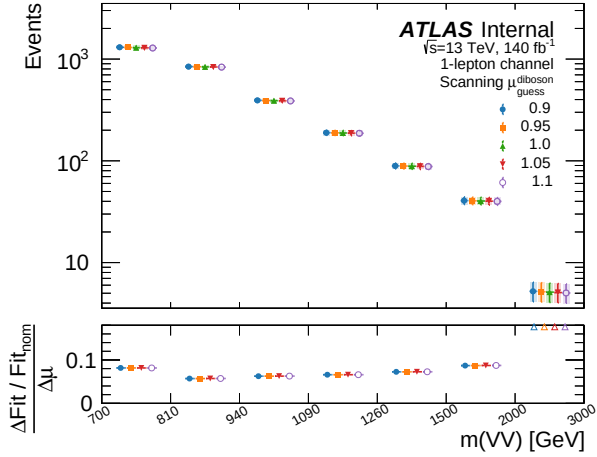
$$\mathcal{P}(n|\lambda) = \frac{e^{-\lambda} \lambda^n}{n!}. \quad (9.4)$$

By instead treating the probability as a function of λ with fixed n , one obtains a likelihood

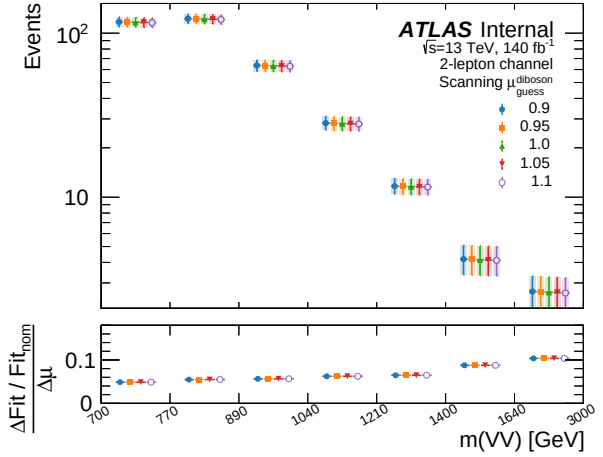
$$L(\lambda|n) = \mathcal{P}(n|\lambda). \quad (9.5)$$



(a) 0-lepton $m_T(VV)$



(b) 1-lepton $m(VV)$



(c) 2-lepton $m(VV)$

Figure 9.2: Gaussian process regression signal contamination tests in all three lepton channels. Here, the Gaussian process is fitted to $N_{V+\text{jets}}^{\text{MC}} + (1 - \mu_{VV})N_{VV}^{\text{MC}}$ as μ_{VV} is scanned. The ratio plot shows the relative difference in the predicted yield, $N_{\text{GPR}}^{\mu_{VV}} / N_{\text{GPR}}^{\mu_{VV}=1} - 1$, divided by $\mu_{VV} - 1$. A linear relationship between N_{GPR} as a function of the amount of signal contamination $\mu_{VV} - 1$ is evidenced by how, for each bin, the points in the subplot generally have the same value.

This includes the effect of only a single measurement. To include multiple bins and channels, the individual Poisson probabilities are multiplied together,

$$L(\vec{\lambda}|\vec{n}) = \prod_i \mathcal{P}(n_i|\lambda_i), \quad (9.6)$$

where the vector notation collects the measurements across the bins and channels, and the index i runs over them.

Next, the expected number of events $\vec{\lambda}$ needs to be written in terms of the signal and background processes. As mentioned previously, the contribution of $t\bar{t}$ and single top are taken from MC templates, while the V+jets contribution is taken from the Gaussian process. Together with the diboson signal, these can be written

$$\vec{\lambda} = \mu_{t\bar{t}}\vec{N}_{t\bar{t}}^{\text{MC}} + \mu_t\vec{N}_t^{\text{MC}} + \vec{N}_{\text{GPR}}^{\text{corrected}} + \mu_{VV}\vec{N}_{VV}^{\text{MC}}, \quad (9.7)$$

where $\vec{N}_{\text{GPR}}^{\text{corrected}}$ is from equation (9.3). The likelihood is now identified as a function of the three μ parameters instead of $\vec{\lambda}$. Note that therefore it is no longer a probability distribution since it does not integrate to unity; this is sometimes referred to as an extended likelihood [135].

However, only μ_{VV} is a free parameter; μ_t and $\mu_{t\bar{t}}$ are constrained, from previous measurements or the TCR fit respectively, as mentioned previously. These constraints are included by writing

$$\mu_t = \hat{\mu}_t + \sigma_t\alpha_t, \quad (9.8)$$

where $\hat{\mu}_t$ is the nominal value, σ_t is its error, and α_t is a nuisance parameter that represents the number of standard deviations μ_t is deviated from the nominal value. A similar procedure is done for $t\bar{t}$. Then Gaussian constraint terms can be multiplied into the likelihood,

$$L(\mu_{VV}; \alpha_t, \alpha_{t\bar{t}}|\vec{n}) = \prod_i \mathcal{P}(n_i|\lambda_i) \times \mathcal{N}(\alpha_t)\mathcal{N}(\alpha_{t\bar{t}}), \quad (9.9)$$

where $\mathcal{N}$ is the standard normal distribution

$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}. \quad (9.10)$$

Once again the likelihood is reinterpreted, written in terms of α_t and $\alpha_{t\bar{t}}$ which replace the respective μ parameters. The α parameters are allowed to freely float, but the constraint terms will cause the likelihood to vanish if either α deviates too much from 0, the nominal value. The semicolon indicates that the α parameters are nuisance parameters, and separates them from the parameter of interest μ_{VV} . A similar treatment can be done for the theory and experimental systematic uncertainties. Each uncertainty is given a nuisance parameter α_k , and a corresponding Gaussian constraint is multiplied into the likelihood.

Finally, the last nuance is to include the bin-by-bin uncertainties, which are the statistical uncertainties in the MC templates as well as the interpolation uncertainties of the Gaussian process regression. These are

added in a very similar manner, sometimes known as the Beeston-Barlow method [136]. Given a bin with an expected value of $\hat{\lambda}_i$ and a total uncertainty on that expectation of σ_i , one can introduce another nuisance parameter γ_i that scales the bin via

$$\lambda_i = \hat{\lambda}_i + \sigma_i \gamma_i. \quad (9.11)$$

These all introduce another set of Gaussian constrain terms into the likelihood. Finally, one obtains

$$L(\mu_{VV}; \vec{\alpha}, \vec{\gamma} | \vec{n}) = \prod_i \mathcal{P}(n_i | \lambda_i) \mathcal{N}(\gamma_i) \times \prod_k \mathcal{N}(\alpha_k). \quad (9.12)$$

Here, $\vec{\alpha}$ collects α_t , $\alpha_{t\bar{t}}$, and the nuisance parameters for all the systematic uncertainties, and the index k runs over all of these. Similarly $\vec{\gamma}$ collects all the statistical nuisance parameters in each bin. Note that $\vec{\lambda}$ is determined now from equation (9.11), where $\hat{\lambda}_i$ follows the form of equation (9.7). However, each of the sample templates $\vec{N}$ are now functions of $\vec{\alpha}$ determined by extrapolating the respective variation histograms mentioned in the previous section.

Equation (9.12) is the final likelihood used for the diboson cross-section measurement. For PLU, the likelihood can be easily adjusted by replacing the diboson sample with the columns of the response matrix as in equation (8.4). That is, instead of $\mu_{VV} \vec{N}_{VV}^{\text{MC}}$, one uses $R \cdot \vec{N}^{\text{Fid}}$. In this case there are multiple parameters of interest, $\vec{N}^{\text{Fid}}$, instead of a single μ_{VV} . This produces a slight complication with the Gaussian process signal contamination. The μ_{VV} in equation (9.3) must now be determined as $\sum_j R_{ij} N_j^{\text{Fid}} / N_i^{\text{VV,MC}}$ for each bin i , but otherwise the procedure is unchanged. Finally for EFT coefficient measurements, the diboson template is treated as a background and fixed to its nominal MC value, while extra terms representing the linear and quadratic effects of an operator are added into equation (9.7). These are scaled by the Wilson coefficients, which become the parameters of interest.

With likelihoods defined for all components of the analysis, predicted values for the parameters of interest are found by maximizing the likelihood. Confidence intervals can be determined using the profile likelihood approach [132]. A test statistic is defined as

$$q(\mu) = -2 \ln \frac{L(\mu, \hat{\theta}_\mu)}{L(\hat{\mu}, \hat{\theta})}, \quad (9.13)$$

where $\theta = (\vec{\alpha}, \vec{\gamma})$ represents all the nuisance parameters, and $\hat{\mu}$ and $\hat{\theta}$ are the values that maximize the likelihood. Note $\hat{\theta}_\mu$ are the values that maximize the likelihood when μ is fixed to a given value. The test statistic $q(\mu)$ follows a χ^2 distribution with one degree of freedom in the asymptotic limit, so a 1σ confidence interval can be determined by finding the values of μ where $q(\mu) = 1$.

9.2 Asimov Results

Validation tests on the fitting procedure were conducted using simulated data, often referred to as *Asimov data*, created by summing the nominal MC prediction for single top, $t\bar{t}$, V+jets, and diboson. These tests are

Channels	SM diboson μ	SMEFT c_W [TeV^{-2}]	
		linear only	linear + quadratic
0+1+2	1.020 ± 0.084	0.050 ± 0.416	0.001 ± 0.028
0	1.008 ± 0.099	0.214 ± 0.668	0.002 ± 0.035
1	1.082 ± 0.118	0.071 ± 0.616	0.000 ± 0.034
2	1.001 ± 0.212	0.153 ± 1.435	0.008 ± 0.142

Table 9.1: Measured SM and EFT signal strengths from fits on Asimov data. The first row fits all three lepton channels together, while the others only fit a single channel. These fits were done against $m_T(VV)$ or $m(VV)$ using the PLU bins as in figure 9.5.

used to debug the software, validate that the fit produces the expected results, and determine the expected sensitivity of the analysis. Asimov tests were run in all three lepton channels. The results are compared with the nominal MC diboson yields, the injected signal, as a closure test.

Pre-fit and post-fit distributions for the diboson signal-strength fit are displayed in figure 9.3 and the measured signal strengths are summarized in table 9.1. The post-fit distributions of the PLU fits are shown in figure 9.4, and the output fiducial distributions are shown in figure 9.5. Additional plots for $m_T(VV)$ in the 1-lepton channel are displayed in figure 9.6. Fits to the c_W operator are also summarized in table 9.1, for both a linear-only fit and a linear+quadratic fit, as detailed in section 2.3.2.

Do note that, because of several delays, systematic variations for some channels were not yet available at the time of this publication. But for those channels where systematic variations were available, the fits were tested both including and excluding the systematics. These tests verified the expected behavior; the fits converge to the same central values, and only the uncertainties are affected. The systematics mostly affect only the lower energy bins, increasing their relative uncertainties by a few percentage points. Meanwhile statistical uncertainties dominate the higher energy bins, and the inclusion of systematics is barely noticeable. Therefore statistics-only fits are fairly accurate representations of the expected final performance, especially for the limits on c_W , for which the sensitivity comes almost entirely from the higher energy bins. Since not all channels had systematics available, for consistency, the results presented in this section use only statistical uncertainties across all channels.

Additional tests were also conducted to validate certain parts of the fit procedure. First, the Gaussian process signal contamination correction described in section 9.1.1 was tested. The entire fit procedure was repeated but with the diboson MC sample scaled by different values of μ_{VV}^{truth} when constructing the Asimov dataset. Fits were conducted both with and without the correction to verify that the correct signal strengths are returned. Next, the PLU procedure was validated by repeating the fit but with the Asimov dataset varied based on the expected Poisson fluctuations in each bin. This was repeated 100 times in the 1-lepton channel, and summarized in figure 9.7. The uncertainties produced by the nominal PLU fit are seen to accurately cover the range of values obtained from the injected statistical fluctuations.

Overall, the full analysis framework performs exceptionally well. The PLU fit accurately recovers the

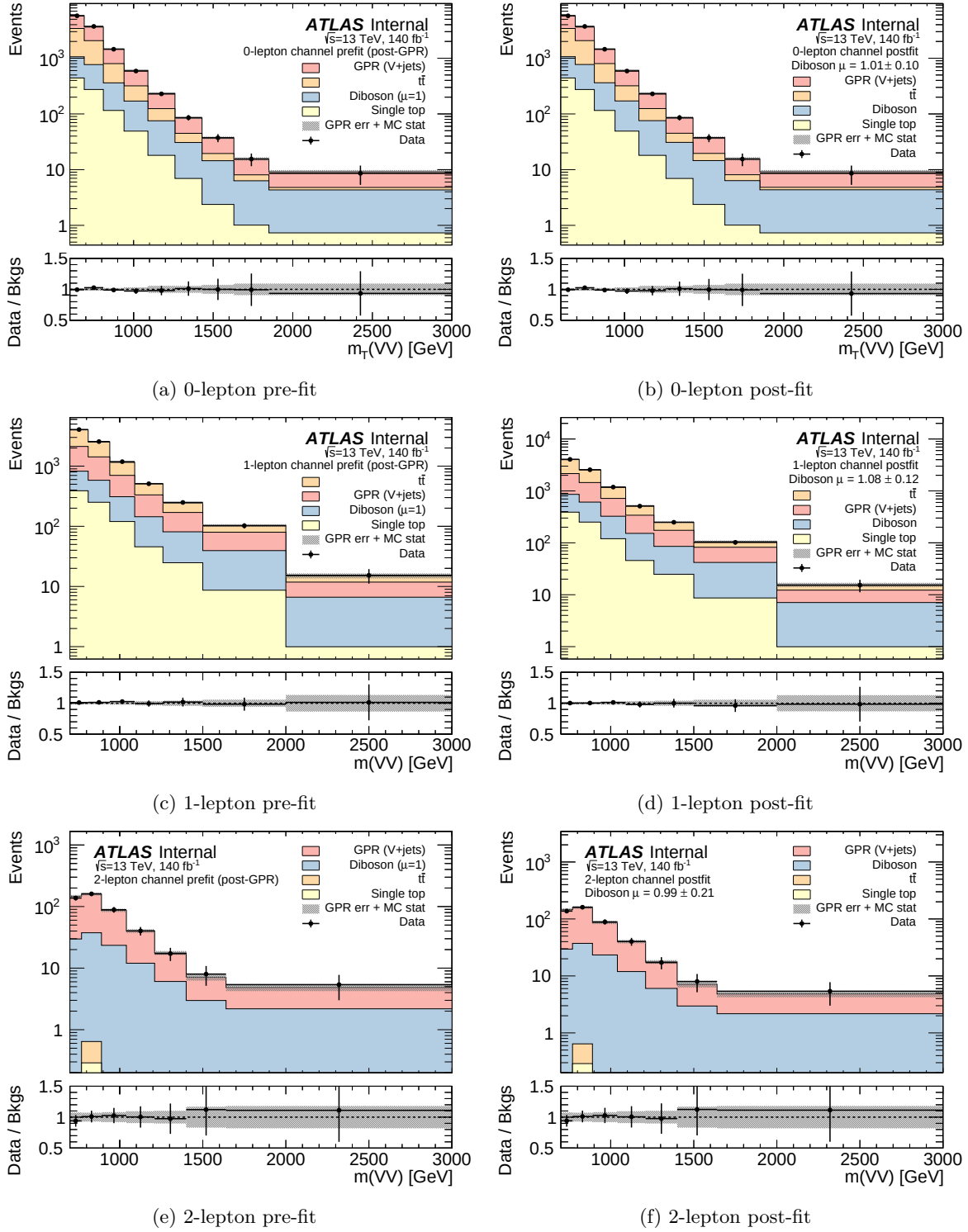
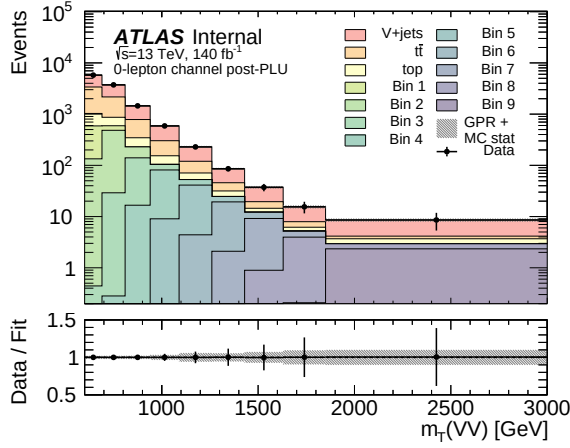
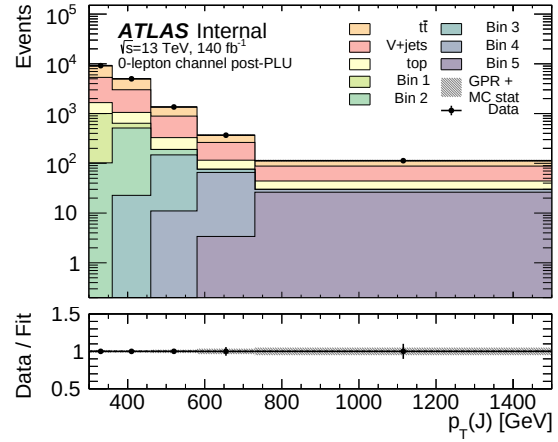


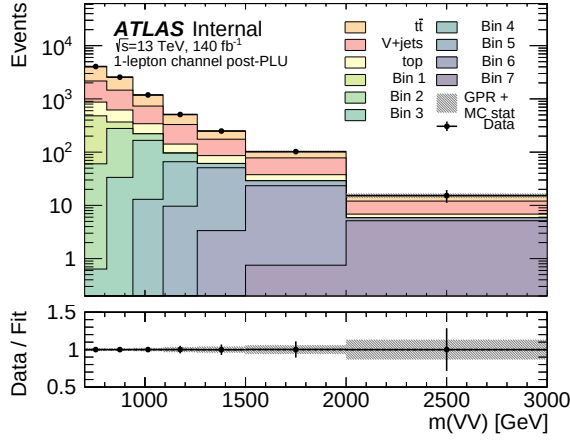
Figure 9.3: Pre-fit and post-fit distributions of the SM diboson signal strength fit in $m_T(VV)$ in the 0-lepton channel and $m(VV)$ in the 1- and 2-lepton channels. The last bin includes overflows. The shaded band includes MC statistical errors and the Gaussian process total error.



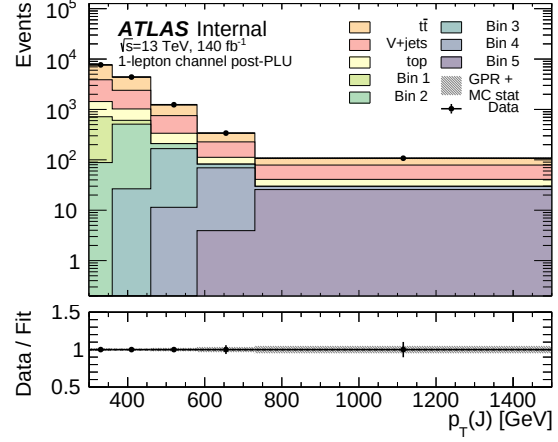
(a) 0-lepton $m_T(VV)$



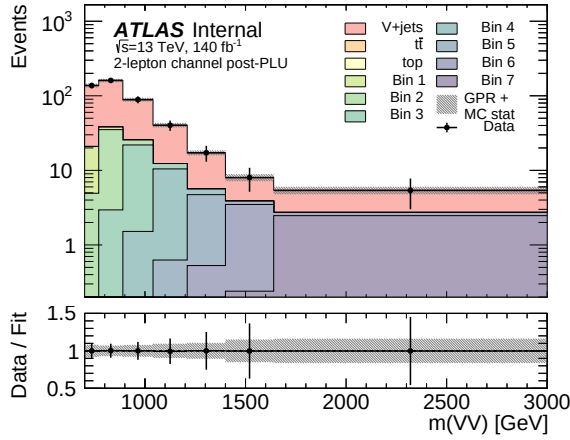
(b) 0-lepton $p_T(J)$



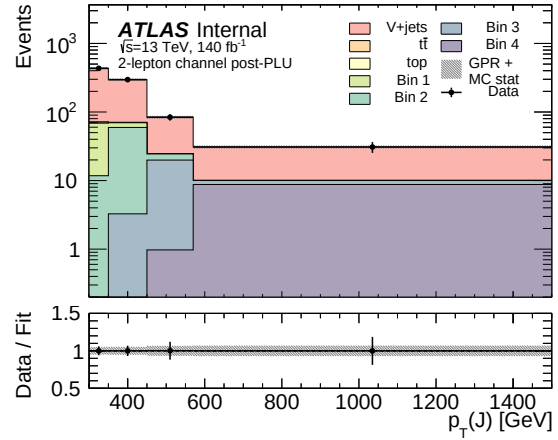
(c) 1-lepton $m(VV)$



(d) 1-lepton $p_T(J)$



(e) 2-lepton $m(VV)$



(f) 2-lepton $p_T(J)$

Figure 9.4: PLU post-fit distributions in $m_T(VV)$, $m(VV)$, or $p_T(J)$ for the three lepton channels. The last bin includes overflows. The shaded band includes MC statistical errors and the Gaussian process total error.

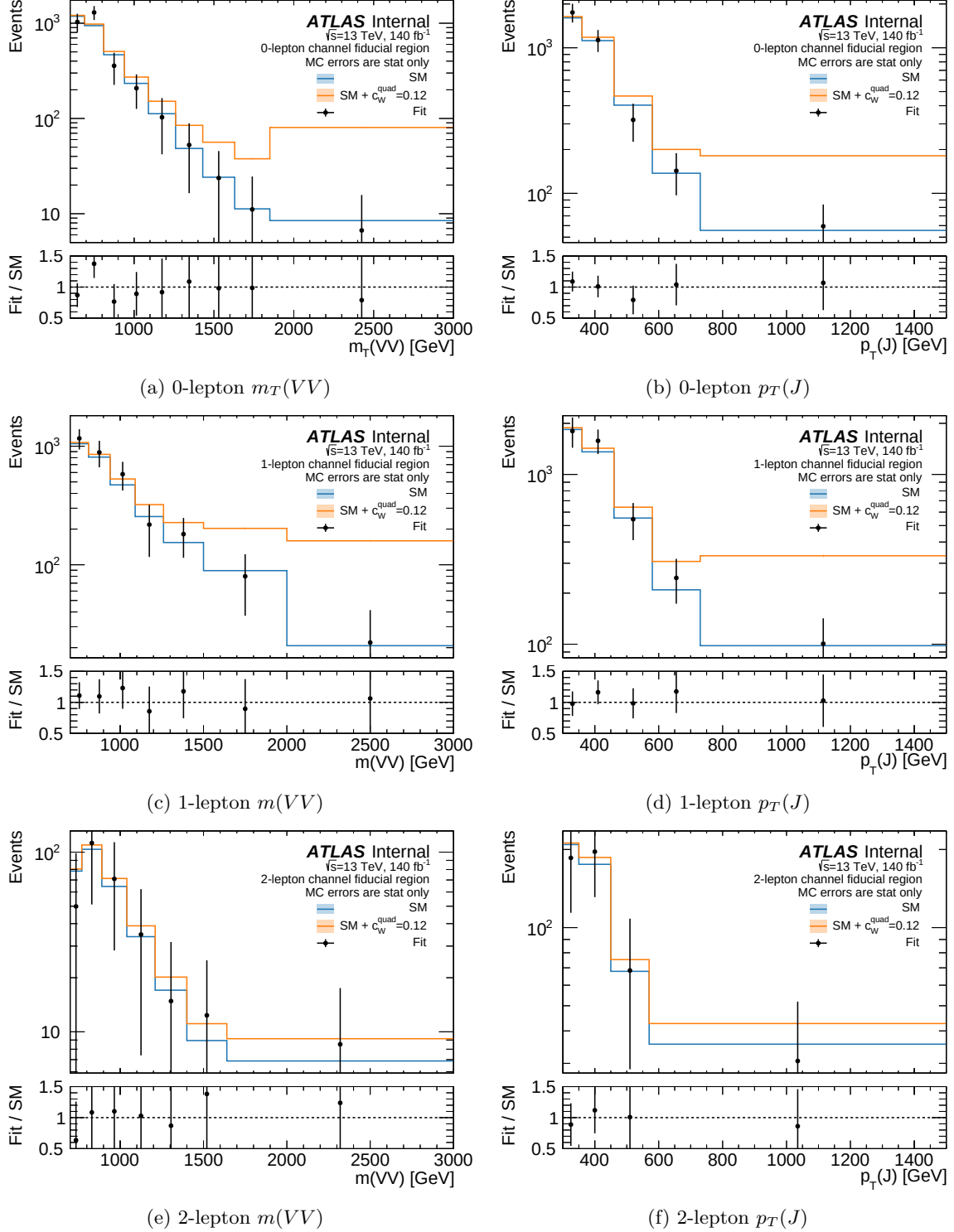


Figure 9.5: Measured fiducial distributions in all lepton channels from Asimov tests, in $m_T(VV)$, $m(VV)$ and $p_T(J)$. The last bin includes overflows. The effect of the c_W EFT operator set to 0.12 TeV^{-2} , its 2σ limit from the most recent ATLAS global EFT fit [23], including both the linear and quadratic components, is also shown to illustrate the sensitivity of the analysis to EFT.

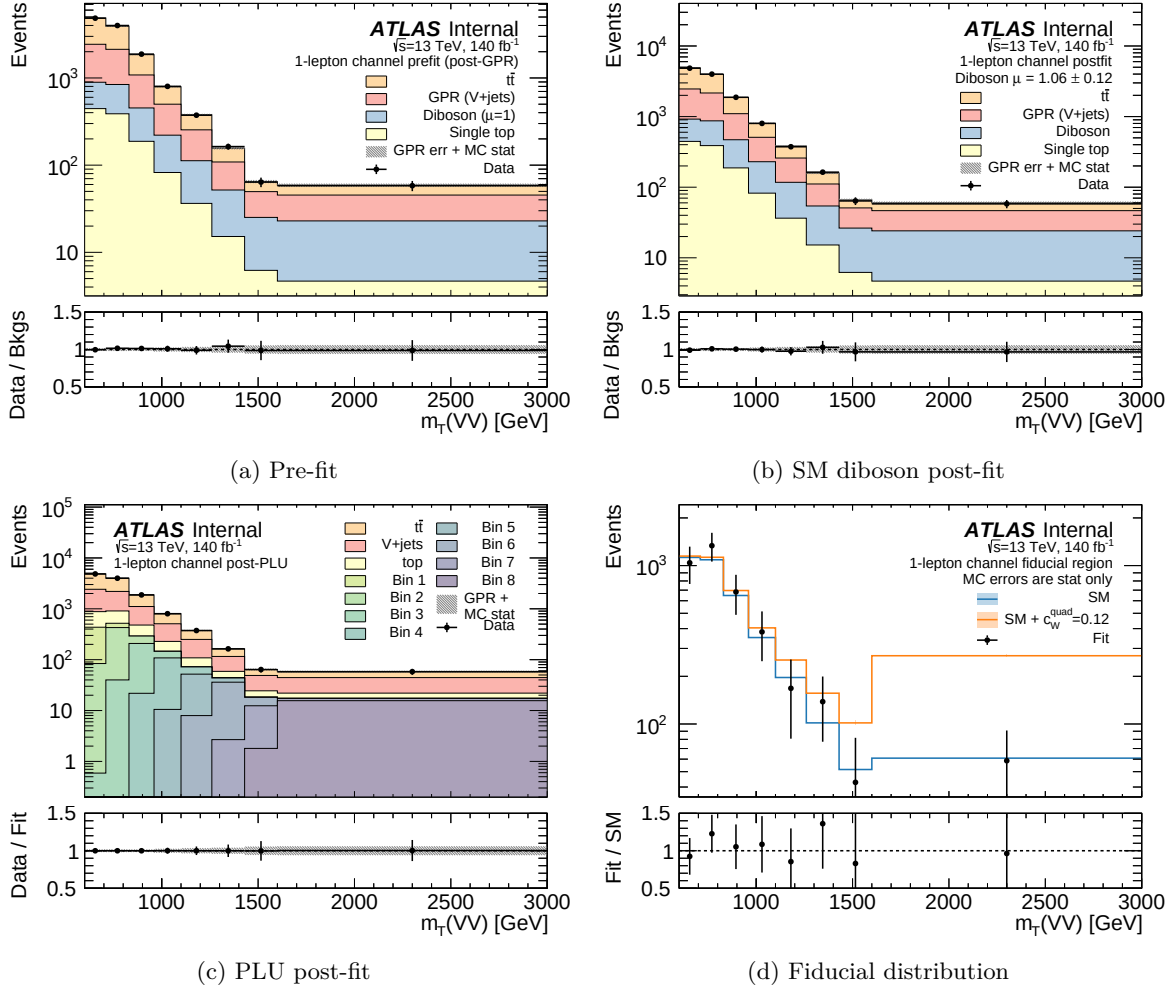


Figure 9.6: Additional fit plots for $m_T(VV)$ in the 1-lepton channel. See the previous figures for descriptions.

SM signal in every channel; the small non-closures seen arise entirely from the Gaussian process fit and not the likelihood unfolding step, and are well within uncertainties. Tremendous sensitivities to both the SM process and the effect of the c_W EFT operator can be observed. The 0- and 1-lepton channels are the most sensitive at approximately equal strengths, while the 2-lepton channel has larger uncertainties. Note also the significant difference in sensitivity to the c_W operator when considering the linear-only versus the linear+quadratic fits, caused by the interference suppression mentioned in section 8.3. The latter can be compared with the results from the latest ATLAS global fit [23], which measures a value of around $c_W/\Lambda^2 = -0.05 \pm 0.08 \text{ TeV}^{-2}$. This Asimov test obtains an uncertainty of $\pm 0.03 \text{ TeV}^{-2}$ in the combined lepton channel fit, a vast improvement to the global fit.

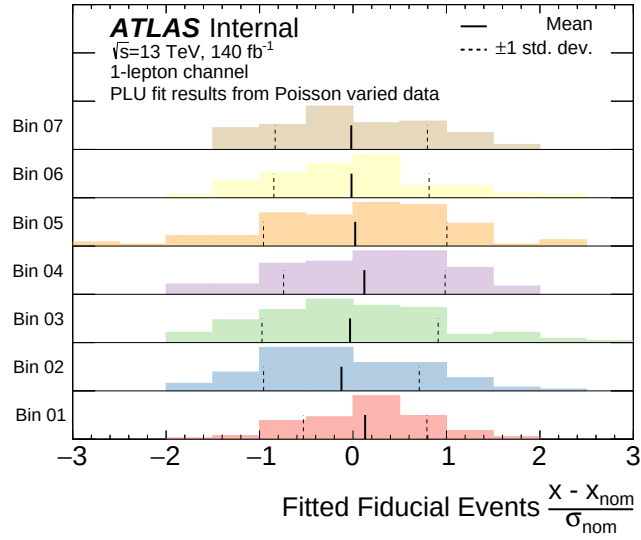


Figure 9.7: Validation test of the PLU uncertainties in the 1-lepton channel $m(VV)$ unfolding. The PLU procedure is repeated 100 times with Poisson-fluctuated Asimov data, and the yields are compared with the nominal Asimov result. Each row shows a normalized histogram of these variation results for each bin in the unfolding, plotted as the number of deviations from the nominal fit yield. The mean and standard deviation of the variations are shown as the solid and dashed lines respectively, which are seen to match with the nominal results and uncertainties.

CHAPTER 10

Conclusion

This dissertation has presented two projects conducted on behalf of the ATLAS experiment at the LHC. The first, presented in part II, helped implement tracking in the trigger as part of the Phase-II upgrade program of the ATLAS detector, so that the trigger could handle the much increased rate of the HL-LHC without losing efficiency. The project was originally designed around a pattern matching method utilizing custom hardware components, but due to requirement changes, an alternative method called the Hough transform was also evaluated. Firmware was developed for the original hardware project, while a robust software framework was implemented for both stages of the project. This framework was used to optimize the respective tracking algorithms, producing a wide range of suitable working points.

The second project, presented in part III, focused on a measurement of the $pp \rightarrow WW, WZ$, and ZZ processes in the semileptonic decay channel. This analysis targeted the full ATLAS Run-2 dataset totaling an integrated luminosity of 139 fb^{-1} at a collision energy of $\sqrt{s} = 13 \text{ TeV}$. However, as the analysis has yet to receive unblinding approval, results using simulated data were presented instead. Differential cross sections in a fiducial region were derived for multiple variables, showing good closure with the SM signal. A sensitivity of 12σ to the SM process is expected when not accounting for systematic uncertainties. Expected limits on the c_W SMEFT operator were also determined at approximately $|c_W/\Lambda^2| < 0.42 \text{ TeV}^{-2}$ at the linear order, and $|c_W/\Lambda^2| < 0.03 \text{ TeV}^{-2}$ at the quadratic order. These uncertainties will expand after accounting for systematics, but nonetheless represent very significant sensitivity to the diboson process.

While actual results have not been unblinded yet, the analysis workflow has been well validated, and the techniques developed can hopefully be brought to other analyses. In particular, a novel method called Gaussian process regression was investigated and found to perform extremely well for background modeling. It is significantly more robust and well-motivated than traditional functional-form fits. As it has seen scarce usage in high-energy physics so far, many analyses could potentially be improved by adopting such techniques. Of course, almost all analyses could also benefit from more data. As of spring 2024, the LHC is midway through Run-3, which by its conclusion in 2025 will approximately double the amount of data available. Much anticipated though is the HL-LHC, which will hopefully begin 2029 and collect the equivalent of the entire Run-2 dataset every year. This will greatly enhance sensitivities across the board, and with some luck, a new discovery may finally come to light.

APPENDIX A

SMEFT Operators

X^3		φ^6 and $\varphi^4 D^2$		$\psi^2 \varphi^3$	
Q_G	$f^{ABC} G_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$	Q_φ	$(\varphi^\dagger \varphi)^3$	$Q_{e\varphi}$	$(\varphi^\dagger \varphi)(\bar{l}_p e_r \varphi)$
$Q_{\tilde{G}}$	$f^{ABC} \tilde{G}_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$	$Q_{\varphi\Box}$	$(\varphi^\dagger \varphi)\Box(\varphi^\dagger \varphi)$	$Q_{u\varphi}$	$(\varphi^\dagger \varphi)(\bar{q}_p u_r \tilde{\varphi})$
Q_W	$\varepsilon^{IJK} W_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$	$Q_{\varphi D}$	$(\varphi^\dagger D^\mu \varphi)^* (\varphi^\dagger D_\mu \varphi)$	$Q_{d\varphi}$	$(\varphi^\dagger \varphi)(\bar{q}_p d_r \varphi)$
$Q_{\tilde{W}}$	$\varepsilon^{IJK} \tilde{W}_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$				
$X^2 \varphi^2$		$\psi^2 X \varphi$		$\psi^2 \varphi^2 D$	
$Q_{\varphi G}$	$\varphi^\dagger \varphi G_{\mu\nu}^A G^{A\mu\nu}$	Q_{eW}	$(\bar{l}_p \sigma^{\mu\nu} e_r) \tau^I \varphi W_{\mu\nu}^I$	$Q_{\varphi l}^{(1)}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi)(\bar{l}_p \gamma^\mu l_r)$
$Q_{\varphi \tilde{G}}$	$\varphi^\dagger \varphi \tilde{G}_{\mu\nu}^A G^{A\mu\nu}$	Q_{eB}	$(\bar{l}_p \sigma^{\mu\nu} e_r) \varphi B_{\mu\nu}$	$Q_{\varphi l}^{(3)}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu^I \varphi)(\bar{l}_p \tau^I \gamma^\mu l_r)$
$Q_{\varphi W}$	$\varphi^\dagger \varphi W_{\mu\nu}^I W^{I\mu\nu}$	Q_{uG}	$(\bar{q}_p \sigma^{\mu\nu} T^A u_r) \tilde{\varphi} G_{\mu\nu}^A$	$Q_{\varphi e}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi)(\bar{e}_p \gamma^\mu e_r)$
$Q_{\varphi \tilde{W}}$	$\varphi^\dagger \varphi \tilde{W}_{\mu\nu}^I W^{I\mu\nu}$	Q_{uW}	$(\bar{q}_p \sigma^{\mu\nu} u_r) \tau^I \tilde{\varphi} W_{\mu\nu}^I$	$Q_{\varphi q}^{(1)}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi)(\bar{q}_p \gamma^\mu q_r)$
$Q_{\varphi B}$	$\varphi^\dagger \varphi B_{\mu\nu} B^{\mu\nu}$	Q_{uB}	$(\bar{q}_p \sigma^{\mu\nu} u_r) \tilde{\varphi} B_{\mu\nu}$	$Q_{\varphi q}^{(3)}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu^I \varphi)(\bar{q}_p \tau^I \gamma^\mu q_r)$
$Q_{\varphi \tilde{B}}$	$\varphi^\dagger \varphi \tilde{B}_{\mu\nu} B^{\mu\nu}$	Q_{dG}	$(\bar{q}_p \sigma^{\mu\nu} T^A d_r) \varphi G_{\mu\nu}^A$	$Q_{\varphi u}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi)(\bar{u}_p \gamma^\mu u_r)$
$Q_{\varphi WB}$	$\varphi^\dagger \tau^I \varphi W_{\mu\nu}^I B^{\mu\nu}$	Q_{dW}	$(\bar{q}_p \sigma^{\mu\nu} d_r) \tau^I \varphi W_{\mu\nu}^I$	$Q_{\varphi d}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi)(\bar{d}_p \gamma^\mu d_r)$
$Q_{\varphi \tilde{W}B}$	$\varphi^\dagger \tau^I \varphi \tilde{W}_{\mu\nu}^I B^{\mu\nu}$	Q_{dB}	$(\bar{q}_p \sigma^{\mu\nu} d_r) \varphi B_{\mu\nu}$	$Q_{\varphi ud}$	$i(\tilde{\varphi}^\dagger D_\mu \varphi)(\bar{u}_p \gamma^\mu d_r)$

Table A.1: Bosonic dimension-6 SMEFT operators, copied from [20], often known as the Warsaw basis.

$(\bar{L}L)(\bar{L}L)$		$(\bar{R}R)(\bar{R}R)$		$(\bar{L}L)(\bar{R}R)$	
Q_{ll}	$(\bar{l}_p \gamma_\mu l_r)(\bar{l}_s \gamma^\mu l_t)$	Q_{ee}	$(\bar{e}_p \gamma_\mu e_r)(\bar{e}_s \gamma^\mu e_t)$	Q_{le}	$(\bar{l}_p \gamma_\mu l_r)(\bar{e}_s \gamma^\mu e_t)$
$Q_{qq}^{(1)}$	$(\bar{q}_p \gamma_\mu q_r)(\bar{q}_s \gamma^\mu q_t)$	Q_{uu}	$(\bar{u}_p \gamma_\mu u_r)(\bar{u}_s \gamma^\mu u_t)$	Q_{lu}	$(\bar{l}_p \gamma_\mu l_r)(\bar{u}_s \gamma^\mu u_t)$
$Q_{qq}^{(3)}$	$(\bar{q}_p \gamma_\mu \tau^I q_r)(\bar{q}_s \gamma^\mu \tau^I q_t)$	Q_{dd}	$(\bar{d}_p \gamma_\mu d_r)(\bar{d}_s \gamma^\mu d_t)$	Q_{ld}	$(\bar{l}_p \gamma_\mu l_r)(\bar{d}_s \gamma^\mu d_t)$
$Q_{lq}^{(1)}$	$(\bar{l}_p \gamma_\mu l_r)(\bar{q}_s \gamma^\mu q_t)$	Q_{eu}	$(\bar{e}_p \gamma_\mu e_r)(\bar{u}_s \gamma^\mu u_t)$	Q_{qe}	$(\bar{q}_p \gamma_\mu q_r)(\bar{e}_s \gamma^\mu e_t)$
$Q_{lq}^{(3)}$	$(\bar{l}_p \gamma_\mu \tau^I l_r)(\bar{q}_s \gamma^\mu \tau^I q_t)$	Q_{ed}	$(\bar{e}_p \gamma_\mu e_r)(\bar{d}_s \gamma^\mu d_t)$	$Q_{qu}^{(1)}$	$(\bar{q}_p \gamma_\mu q_r)(\bar{u}_s \gamma^\mu u_t)$
		$Q_{ud}^{(1)}$	$(\bar{u}_p \gamma_\mu u_r)(\bar{d}_s \gamma^\mu d_t)$	$Q_{qu}^{(8)}$	$(\bar{q}_p \gamma_\mu T^A q_r)(\bar{u}_s \gamma^\mu T^A u_t)$
		$Q_{ud}^{(8)}$	$(\bar{u}_p \gamma_\mu T^A u_r)(\bar{d}_s \gamma^\mu T^A d_t)$	$Q_{qd}^{(1)}$	$(\bar{q}_p \gamma_\mu q_r)(\bar{d}_s \gamma^\mu d_t)$
				$Q_{qd}^{(8)}$	$(\bar{q}_p \gamma_\mu T^A q_r)(\bar{d}_s \gamma^\mu T^A d_t)$
$(\bar{L}R)(\bar{R}L)$ and $(\bar{L}R)(\bar{L}R)$		B -violating			
Q_{ledq}	$(\bar{l}_p^j e_r)(\bar{d}_s q_t^j)$	Q_{duq}	$\varepsilon^{\alpha\beta\gamma} \varepsilon_{jk} [(d_p^\alpha)^T C u_r^\beta] [(q_s^\gamma)^T C l_t^k]$		
$Q_{quqd}^{(1)}$	$(\bar{q}_p^j u_r) \varepsilon_{jk} (\bar{q}_s^k d_t)$	Q_{qqqu}	$\varepsilon^{\alpha\beta\gamma} \varepsilon_{jk} [(q_p^{\alpha j})^T C q_r^{\beta k}] [(u_s^\gamma)^T C e_t]$		
$Q_{quqd}^{(8)}$	$(\bar{q}_p^j T^A u_r) \varepsilon_{jk} (\bar{q}_s^k T^A d_t)$	Q_{qqq}	$\varepsilon^{\alpha\beta\gamma} \varepsilon_{jn} \varepsilon_{km} [(q_p^{\alpha j})^T C q_r^{\beta k}] [(q_s^{\gamma m})^T C l_t^n]$		
$Q_{lequ}^{(1)}$	$(\bar{l}_p^j e_r) \varepsilon_{jk} (\bar{q}_s^k u_t)$	Q_{duu}	$\varepsilon^{\alpha\beta\gamma} [(d_p^\alpha)^T C u_r^\beta] [(u_s^\gamma)^T C e_t]$		
$Q_{lequ}^{(3)}$	$(\bar{l}_p^j \sigma_{\mu\nu} e_r) \varepsilon_{jk} (\bar{q}_s^k \sigma^{\mu\nu} u_t)$				

Table A.2: Four fermion dimension-6 SMEFT operators, copied from [20], often known as the Warsaw basis.

APPENDIX B

Resonance Event Selections

These are the full event selection cuts for the resonance side of the VV+VH Semileptonic analysis, separated by the three lepton channels.

Selection	VV merged	VH merged	VV resolved (not explored)	VH merged
0-lepton Selection				
Trigger	MET Trigger			
Lepton Multiplicity	0 “loose” Leptons			
$E_{\text{T}}^{\text{miss}}$	> 200 GeV		> 150 GeV	
S	> 10			
$\min[\Delta\phi(\text{jets}, E_{\text{T}}^{\text{miss}})]$	> 0.2			
Jet Cleaning	Tight			
Jet Selection				
Number of Jets	1 large- R jet		2 small- R jet	
Leading jet p_{T}	> 200 GeV	> 250 GeV	> 45 GeV	
$W/Z/h$ candidate mass	Tagger dependent mass cut	$75 < m(J) < 145$ GeV	0/1-lepton: $110 < m(jj) < 140$ 2-lepton: $100 < m(jj) < 145$	W : $68 < m(jj) < 98$ Z : $78 < m(jj) < 105$

Table B.1: Cuts in the 0-lepton channel of the resonance side of the VV+VH Semileptonic analysis.

Selection	VV merged	VH merged	VV resolved	VH merged
1-lepton Selection				
Trigger	Single lepton or MET Trigger			
Lepton Multiplicity	1 “Tight” lepton and 0 “loose” leptons			
lepton p_T	> 30 GeV			
E_T^{miss}	> 100 GeV		> 60 GeV	
$p_T(W)$	> 200 GeV		> 75 GeV	
Jet Selection				
Number of Jets	1 large- R jet		2 small- R jet	
Leading jet p_T	> 200 GeV	> 250 GeV	> 45 GeV	
$W/Z/h$ candidate mass	Tagger dependent mass cut	$75 < m(J) < 145$ GeV	0/1-lepton: $110 < m(jj) < 140$ 2-lepton: $100 < m(jj) < 145$	W : $68 < m(jj) < 98$ Z : $78 < m(jj) < 105$
Topology Requirements				
$E_T^{\text{miss}}/p_T(W)$	> 0.2 electron-only			
b -veto	No additional b -jet in event		No b -jet in $\Delta R(J, b) < 1.0$	
R	GGF: > 0.35 VBF: > 0.25			
$\Delta\phi(\ell, E_T^{\text{miss}})$	-		< 1.5	
$\Delta\phi(j_1, j_2)$	-		< 1.5	
$\Delta\phi(\ell, j_1/j_2)$	-		> 1.0	
$\Delta\phi(\ell, j_1/j_2)$	-		> 1.0	

Table B.2: Cuts in the 1-lepton channel of the resonance side of the VV+VH Semileptonic analysis.

Selection	VV merged	VH merged	VV resolved	VH merged
2-lepton Selection				
Trigger	Dilepton			
Lepton Multiplicity	2 “loose” lepton and no additional			
Leading lepton p_T	> 27 GeV			
Subleading lepton p_T	> 25 GeV		> 20 GeV	
$m(\ell\ell)$	$83 < m_{ee} < 99$ GeV $85.6 - 0.0117p_T(\ell\ell) < m_{\mu\mu} < 94.0 + 0.0185p_T(\ell\ell)$ GeV			
Jet Selection				
Number of Jets	1 large- R jet		2 small- R jet	
Leading jet p_T	> 200 GeV	> 250 GeV	> 45 GeV	
$W/Z/h$ candidate mass	Tagger dependent mass cut	$75 < m(J) < 145$ GeV	0/1-lepton: $110 < m(jj) < 140$ 2-lepton: $100 < m(jj) < 145$	W : $68 < m(jj) < 98$ Z : $78 < m(jj) < 105$
Topology Requirements				
R	GGF: > 0.35 VBF: > 0.25			

Table B.3: Cuts in the 2-lepton channel of the resonance side of the VV+VH Semileptonic analysis.

APPENDIX C

V+jets Fits

This appendix collects the bin-by-bin fits to the V+jets background. Appendix C.1 contains fits to MC V+jets, while appendix C.2 contains fits to actual data.

C.1 Fits to Monte Carlo

These plots are part of studies in chapter 7 for the VV+VH Semileptonic analysis. See the text there for more detailed description. Most of these plots focus on the 1-lepton channel, and therefore are fit against W +jets MC samples. They serve as closure and performance tests.

C.1.1 Functional Fits

C.1.1.1 Three-variable Tagger

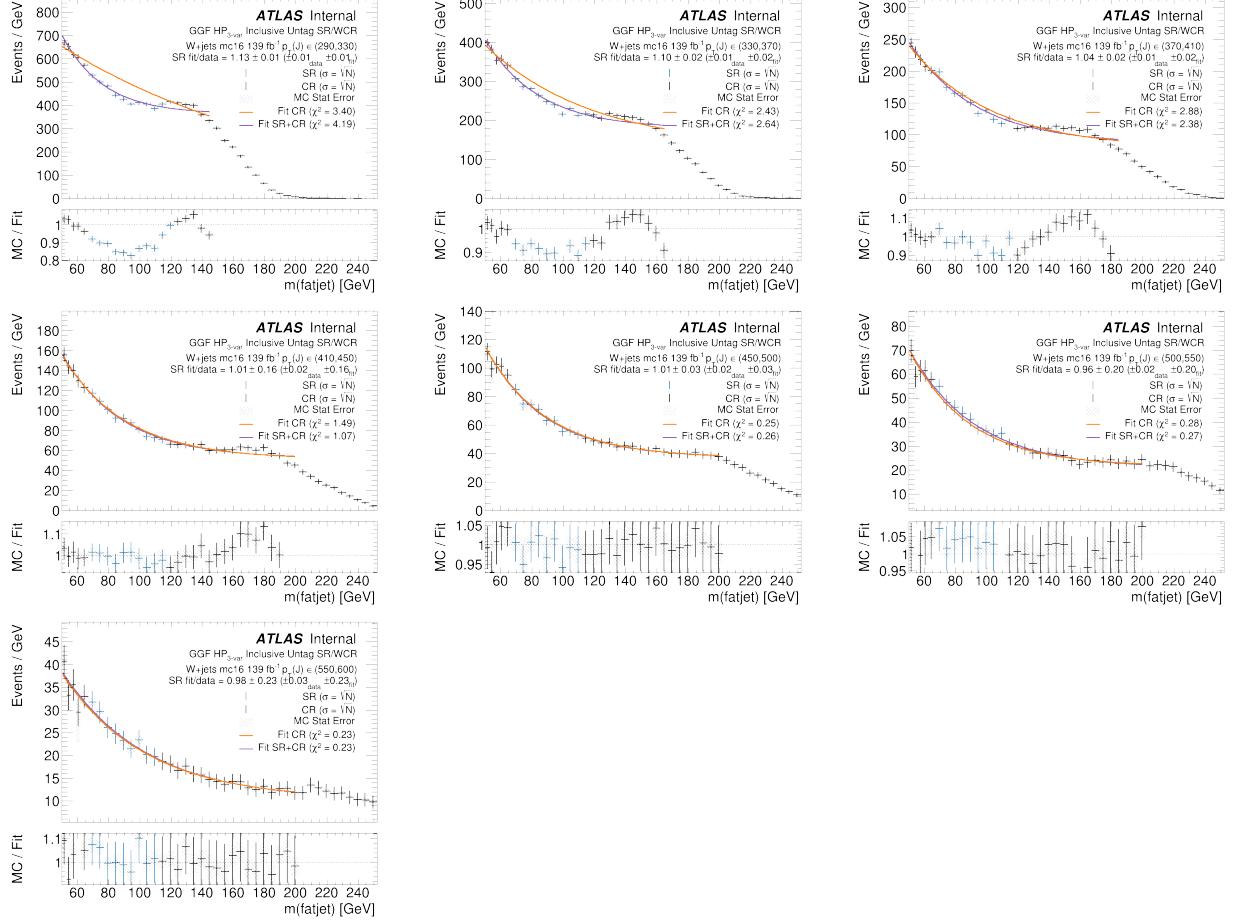


Figure C.1: Fits to W +jets MC in the three-variable tagger HP region in bins of $p_T(J)$, using a single exponential functional form, equation (7.2). The plots scan different $p_T(J)$ bins. The MC yields are shown in the black points for the control region and blue points for the signal region, where the error bars show $\sqrt{N}$ errors. The orange line shows the fit to the control region, and the purple line shows a fit to both the control and signal regions as a closure test.

C.1.1.2 ANN

C.1.1.2.1 Signal Region Scan

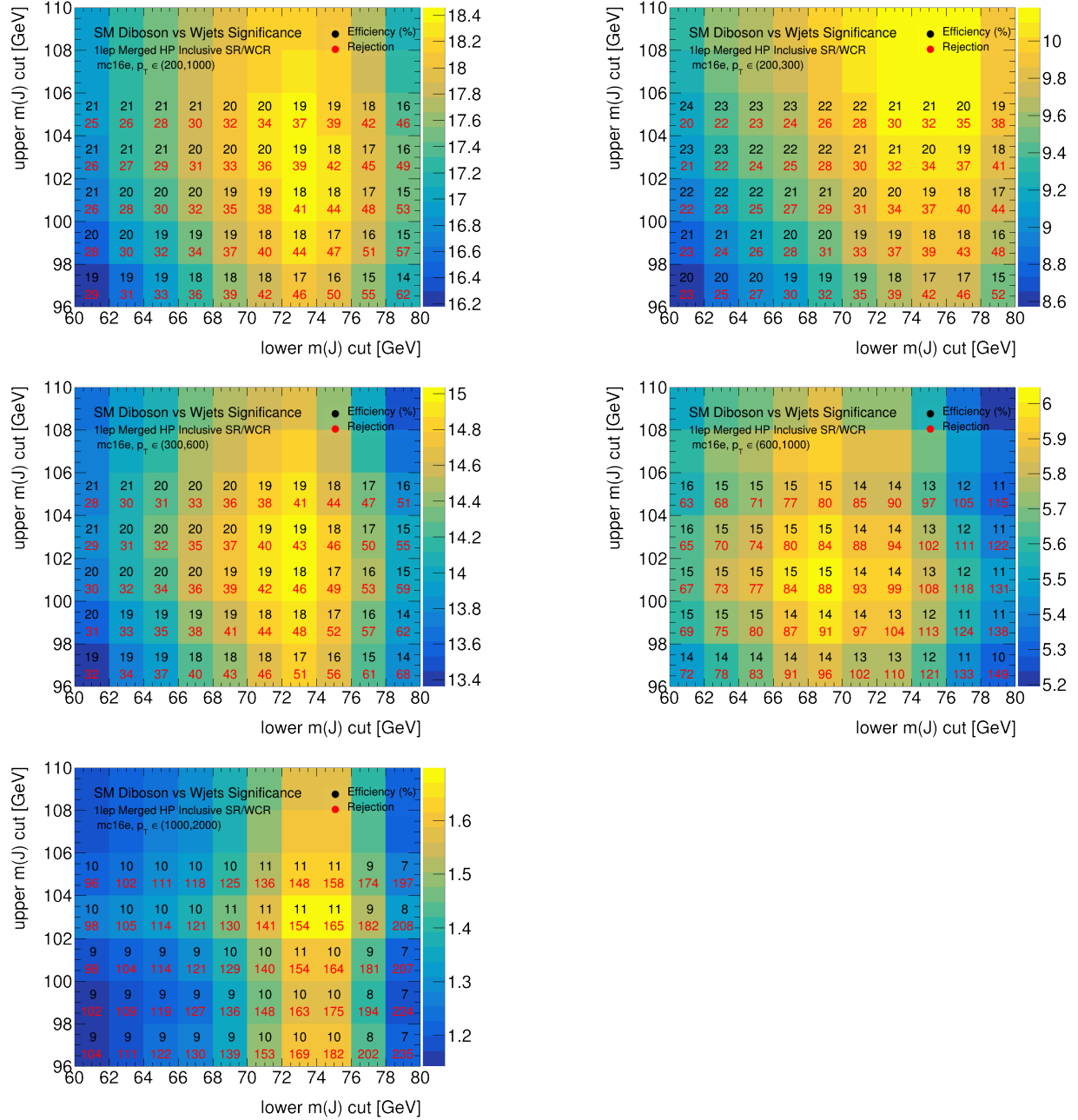


Figure C.2: A scan of the lower and upper $m(J)$ cuts to be used as the signal region definition with the ANN tagger. Each bin corresponds to cuts at the bottom-left corner of the bin, so for example the bottom-left bin corresponds to a signal region $m(J) \in [60, 96]$ GeV. The z axis shows the significance, and the efficiency (as a percentage) and background rejection are overlaid as text. Each plot is a different bin in p_T .

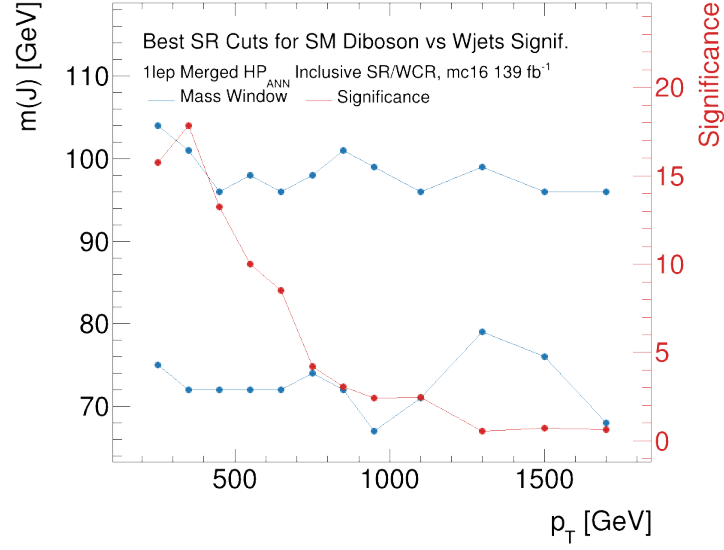


Figure C.3: A summary of the highest significance mass window in 100 GeV bins of $p_T(J)$. The blue points are the lower and upper cut values (left axis), and the red points are the significance of that cut (right axis).

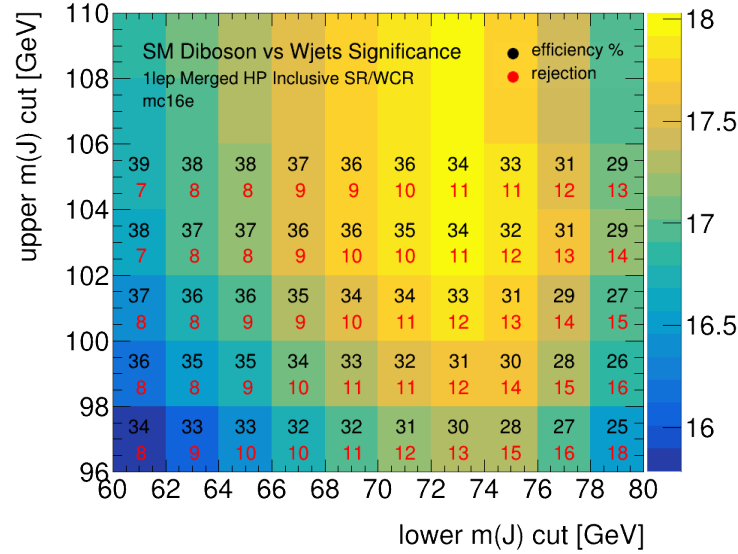


Figure C.4: A scan of the lower and upper $m(J)$ cuts to be used as the signal region definition with an alternative 80% working point ANN tagger. Each bin corresponds to cuts at the bottom-left corner of the bin, so for example the bottom-left bin corresponds to a signal region $m(J) \in [60, 96]$ GeV. The z axis shows the significance, and the efficiency (as a percentage) and background rejection are overlaid as text.

C.1.1.2.2 One-exponential Fits

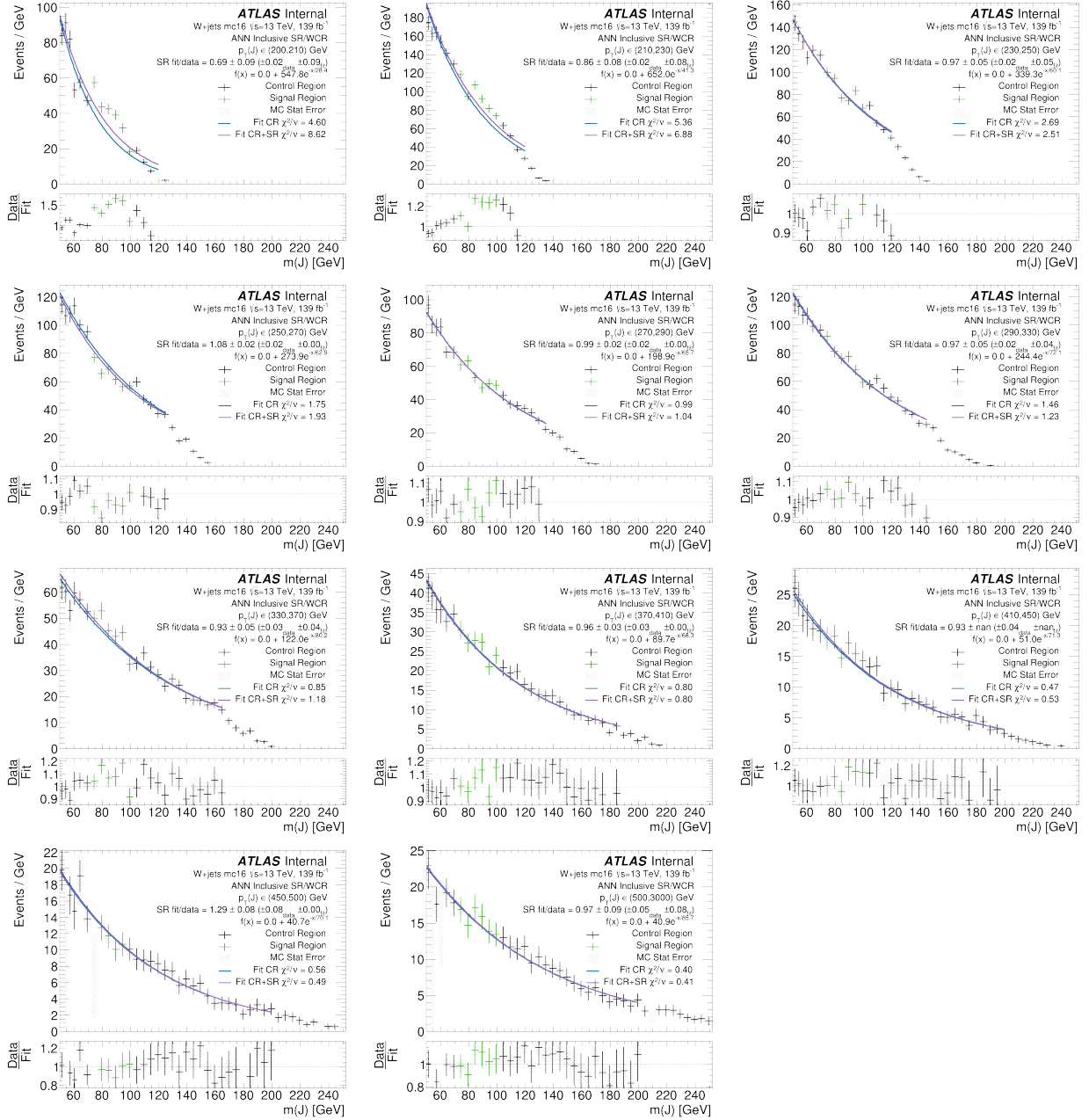


Figure C.5: Fits to the W +jets MC in the ANN HP region across bins of $p_T(J)$. These fits use equation (7.2) as the functional form. The MC yields are shown in the black points for the control region and green points for the signal region, where the error bars show $\sqrt{N}$ errors. The blue line shows the fit to the control region, and the purple line shows a fit to both the control and signal regions as a closure test.

C.1.1.2.3 Two-exponential Fits

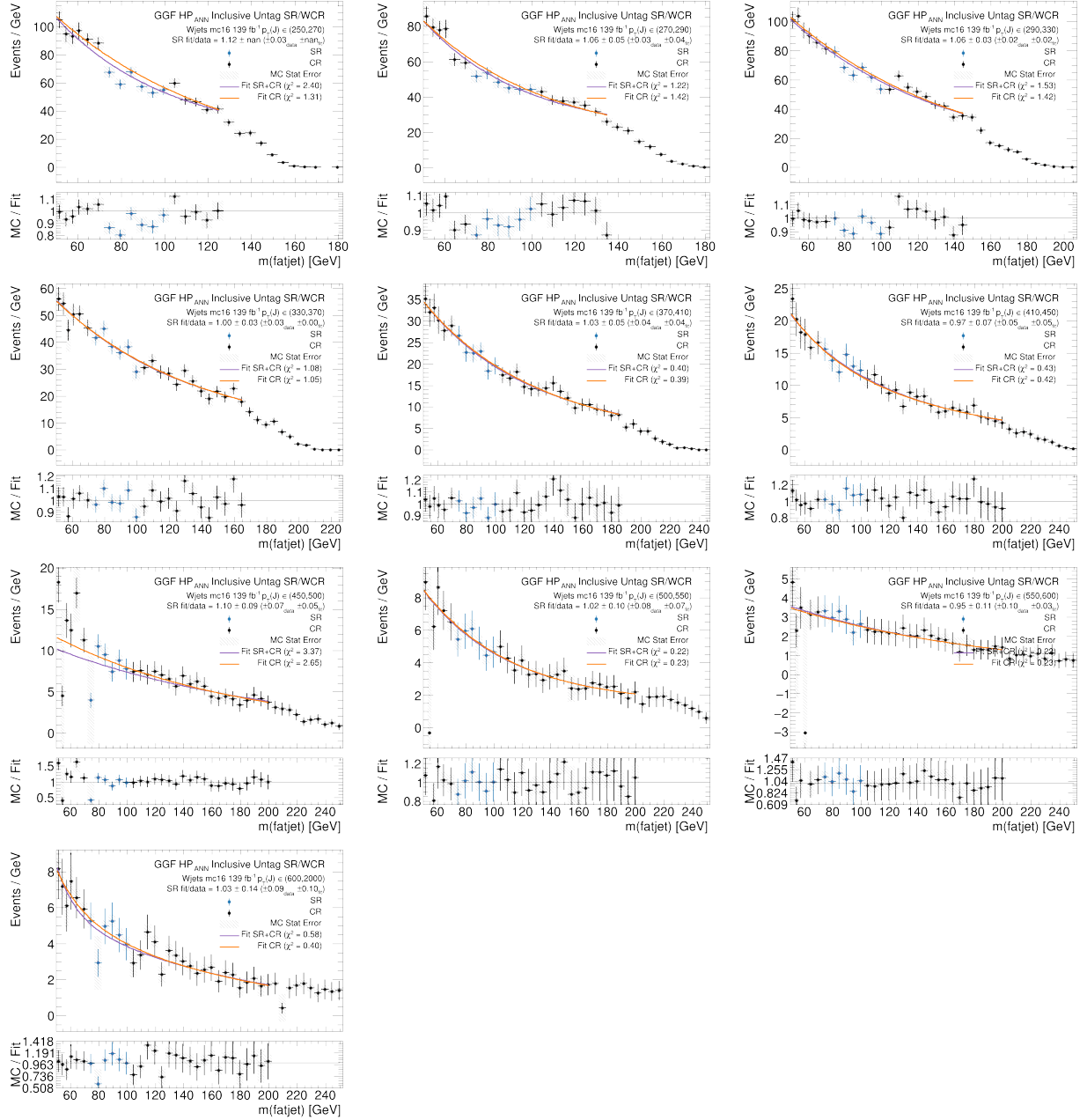


Figure C.6: Fits to the W +jets MC in the ANN HP region across bins of $p_T(J)$. These fits use equation (7.4) as the functional form. The MC yields are shown in the black points for the control region and blue points for the signal region, where the error bars show $\sqrt{N}$ errors. The orange line shows the fit to the control region, and the purple line shows a fit to both the control and signal regions as a closure test.

C.1.1.2.4 Third-order Chebyshev Fits

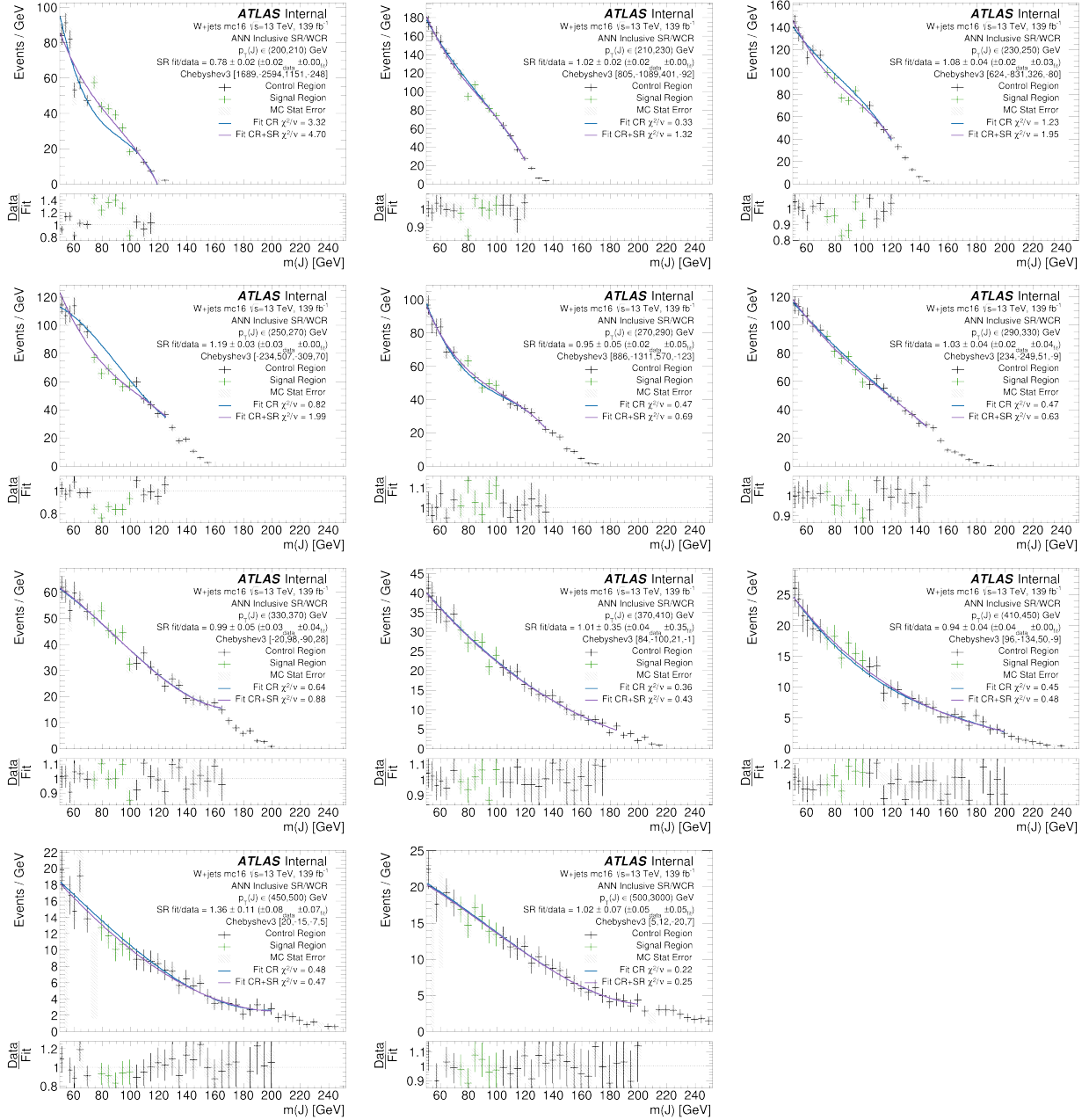


Figure C.7: Fits to the W +jets MC in the ANN HP region across bins of $p_T(J)$. These fits use equation (7.5) at order 3 as the functional form. The MC yields are shown in the black points for the control region and green points for the signal region, where the error bars show $\sqrt{N}$ errors. The blue line shows the fit to the control region, and the purple line shows a fit to both the control and signal regions as a closure test.

C.1.2 Gaussian Process

In these plots, the black (red) points indicate the control (signal) region respectively, where the errors bars show $\sqrt{N}$ errors used in the fit. The MC statistical error is shown in the gray hash instead. The blue band is the Gaussian process posterior mean $\pm 2\sigma$.

C.1.2.1 Binning and Kernel

These fits use Gaussian process regression in bins of $p_T(J)$ to test the effect of binning and kernel choice. See section 7.4.1 for details.

C.1.2.1.1 Radial Basis Function

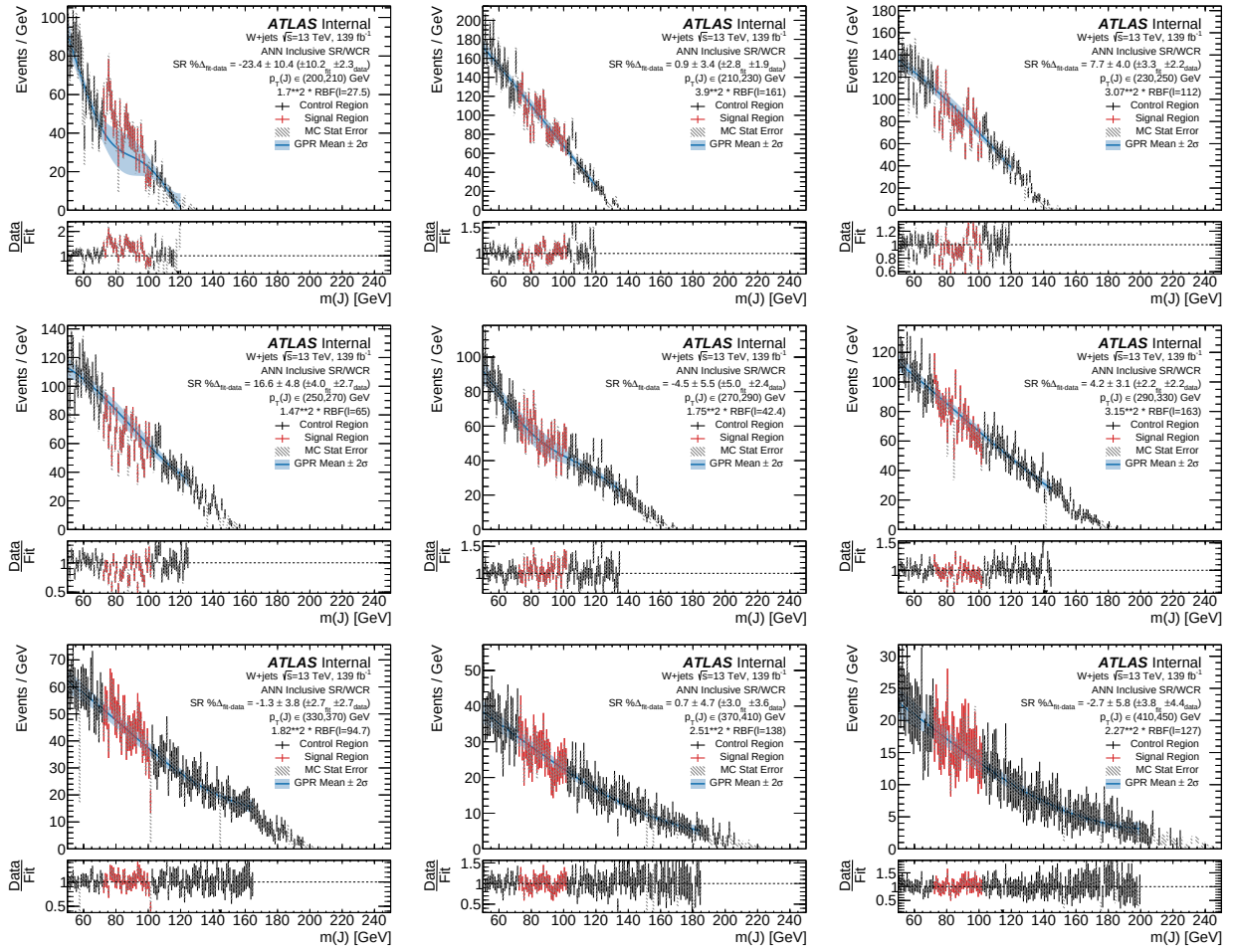


Figure C.8: Gaussian process fit to W +jets MC using the RBF kernel and 1 GeV bins.

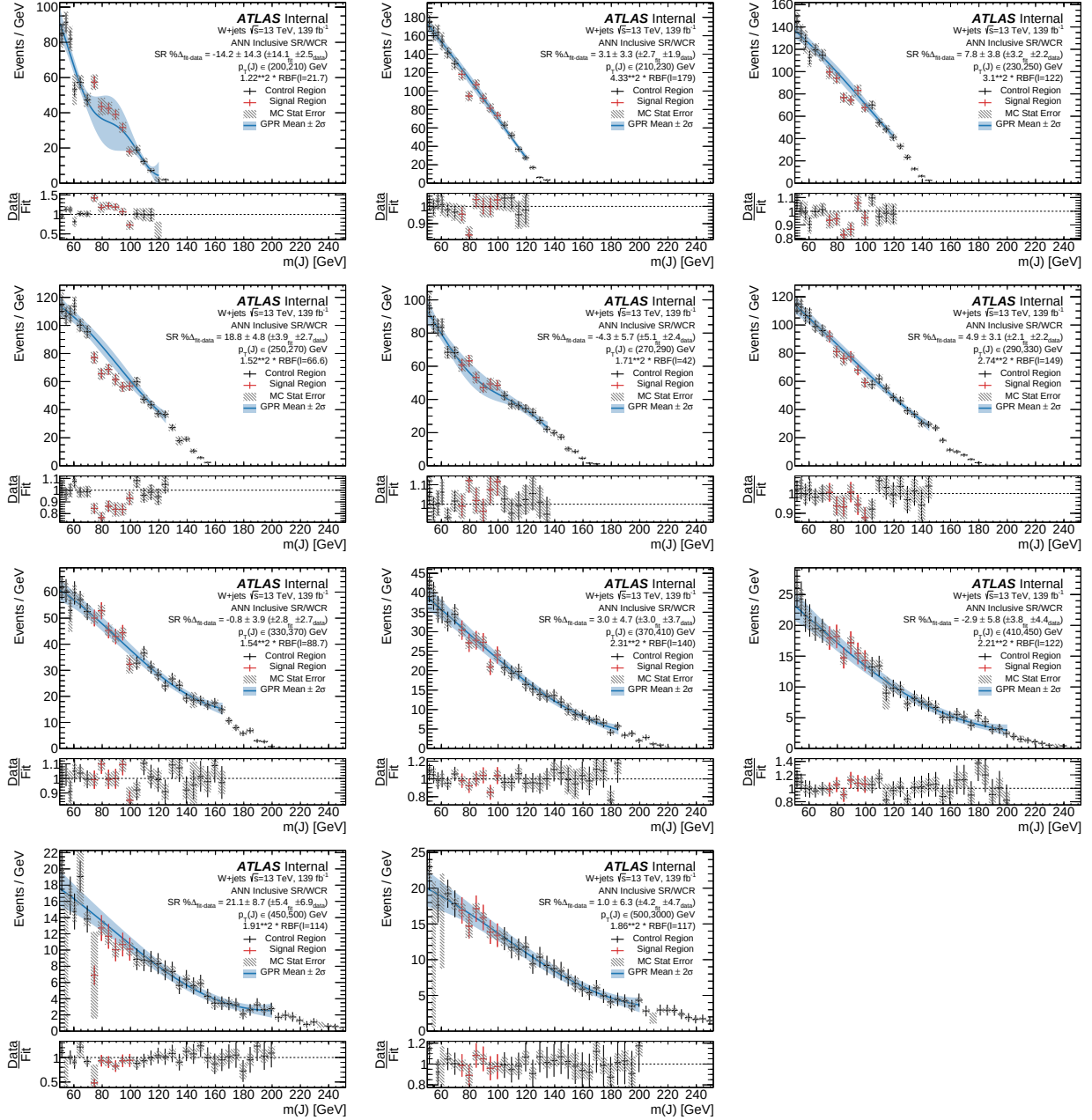


Figure C.9: Gaussian process fit to W +jets MC using the RBF kernel and variable 3–5 GeV bins.

C.1.2.1.2 Matern 5/2

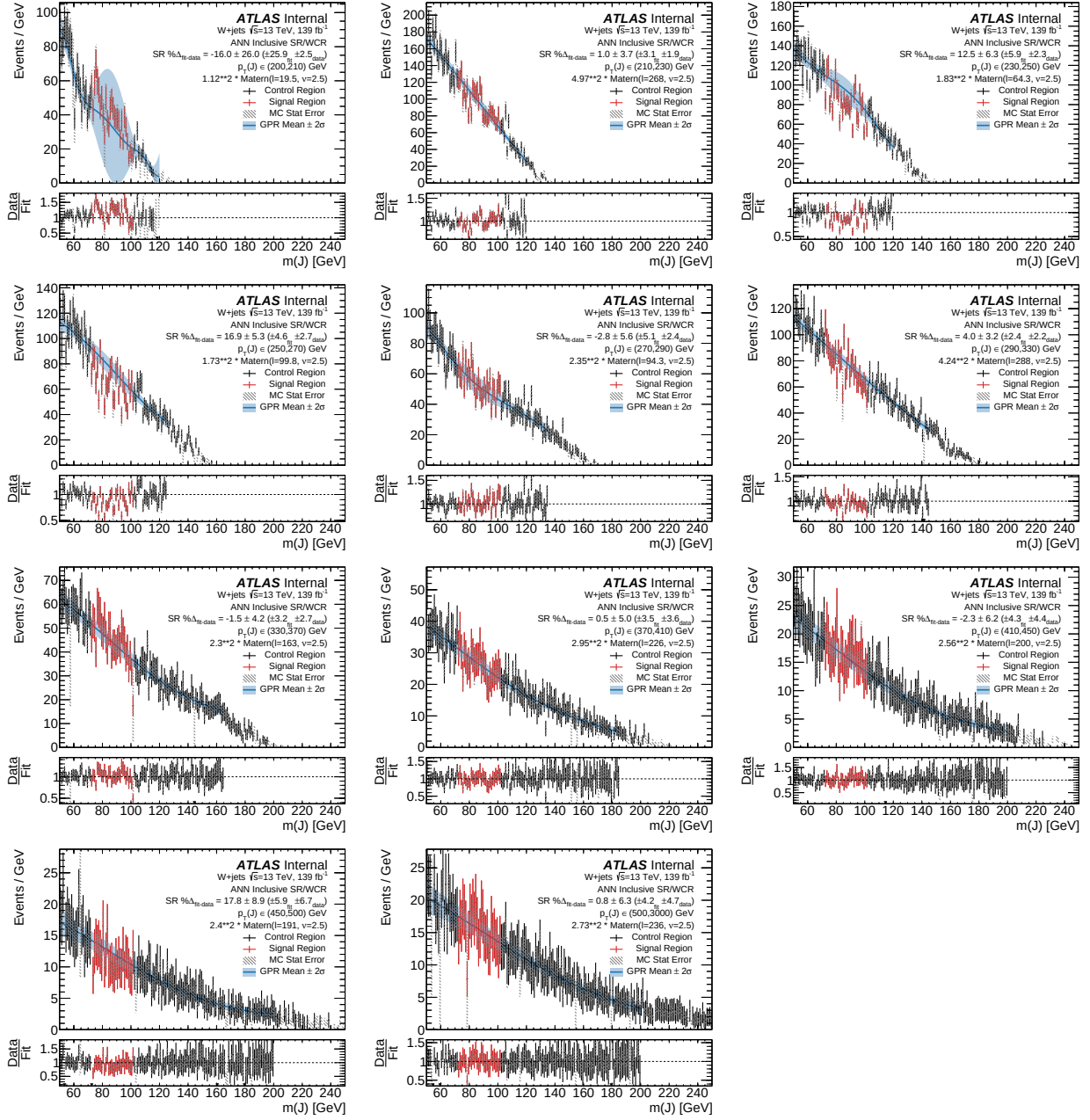


Figure C.10: Gaussian process fit to W +jets MC using the Matern 5/2 kernel and 1 GeV bins.

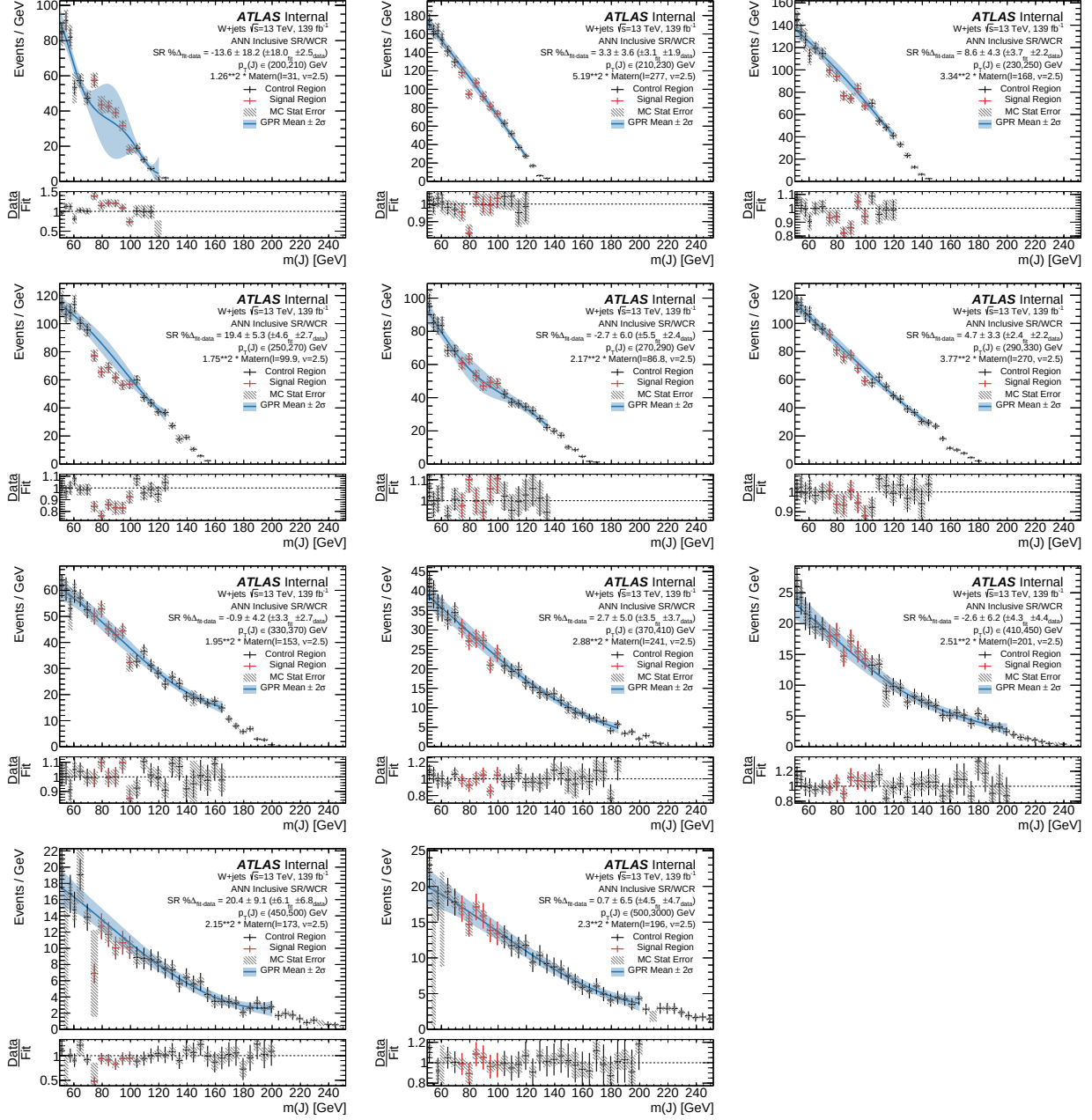


Figure C.11: Gaussian process fit to W +jets MC using the Matern 5/2 kernel and variable 3–5 GeV bins.

C.1.2.2 2D Fits

These use a 2-dimensional Gaussian process regression fit against $m(VV)$. See section 7.4.3 for details.

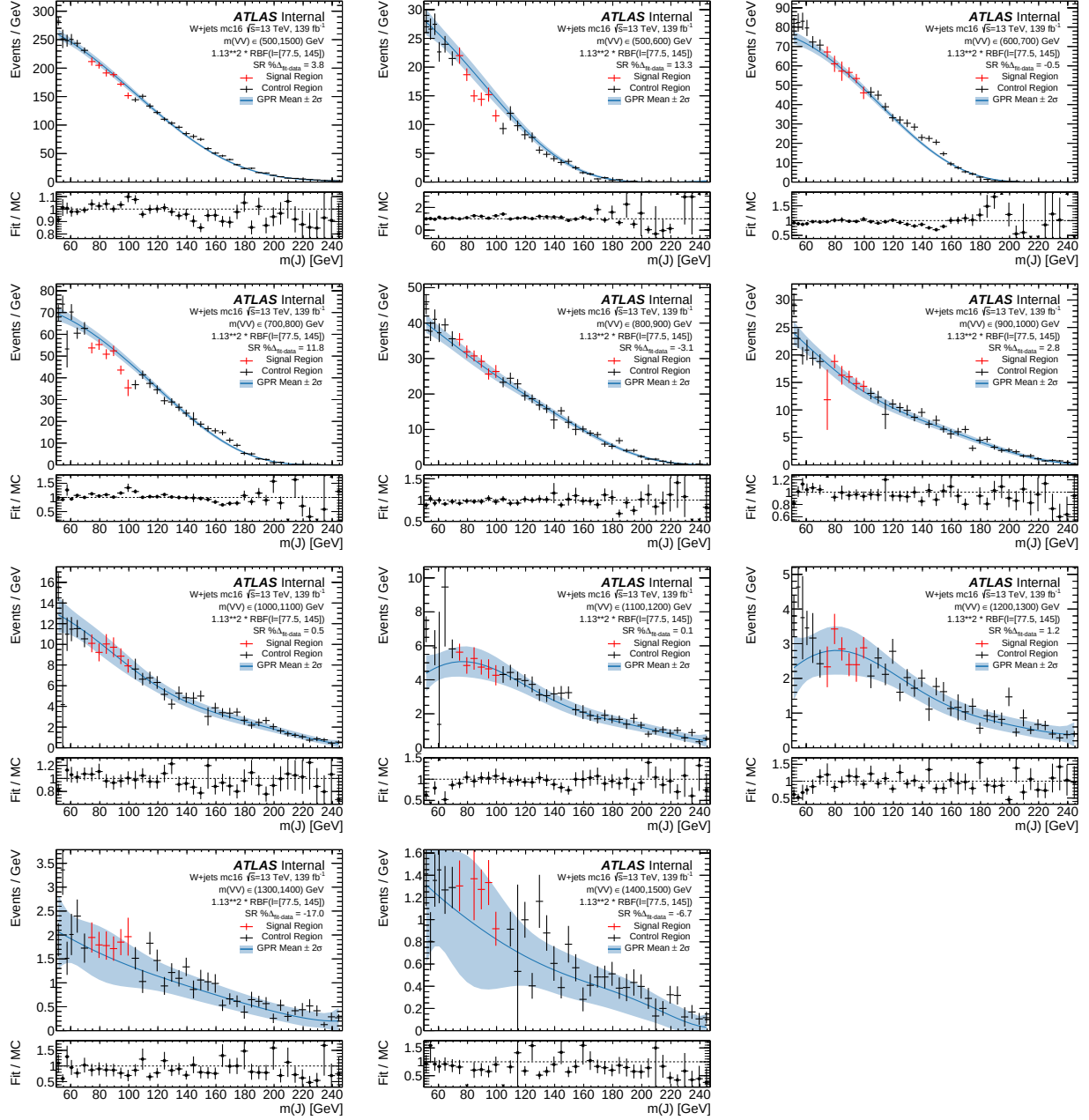


Figure C.12: Two-dimensional Gaussian process regression fits against $m(VV)$ and $m(J)$ using W +jets MC, projected onto $m(J)$ in 100 GeV bins of $m(VV)$. Note the first plot is the inclusive $m(VV)$ spectrum, while the rest scan through the $m(VV)$ bins.

C.1.2.3 0-lepton and 2-lepton Channels

Additional fits to V+jets in the other lepton channels. See section 7.4.4 for context. These plots show only a MMLE fit, and omit the contour scan plots.

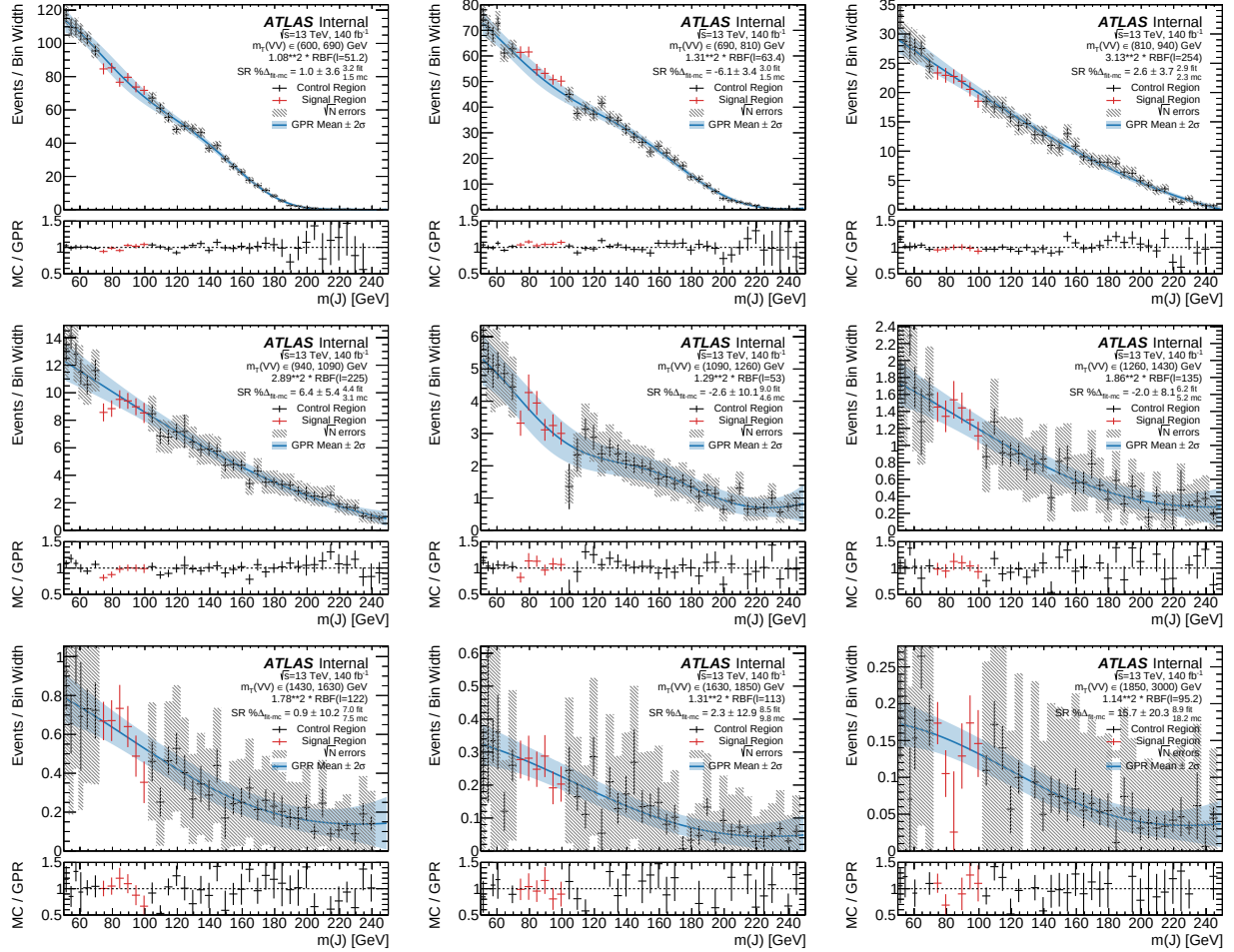


Figure C.13: Gaussian process fits to V+jets MC in the 0-lepton channel.

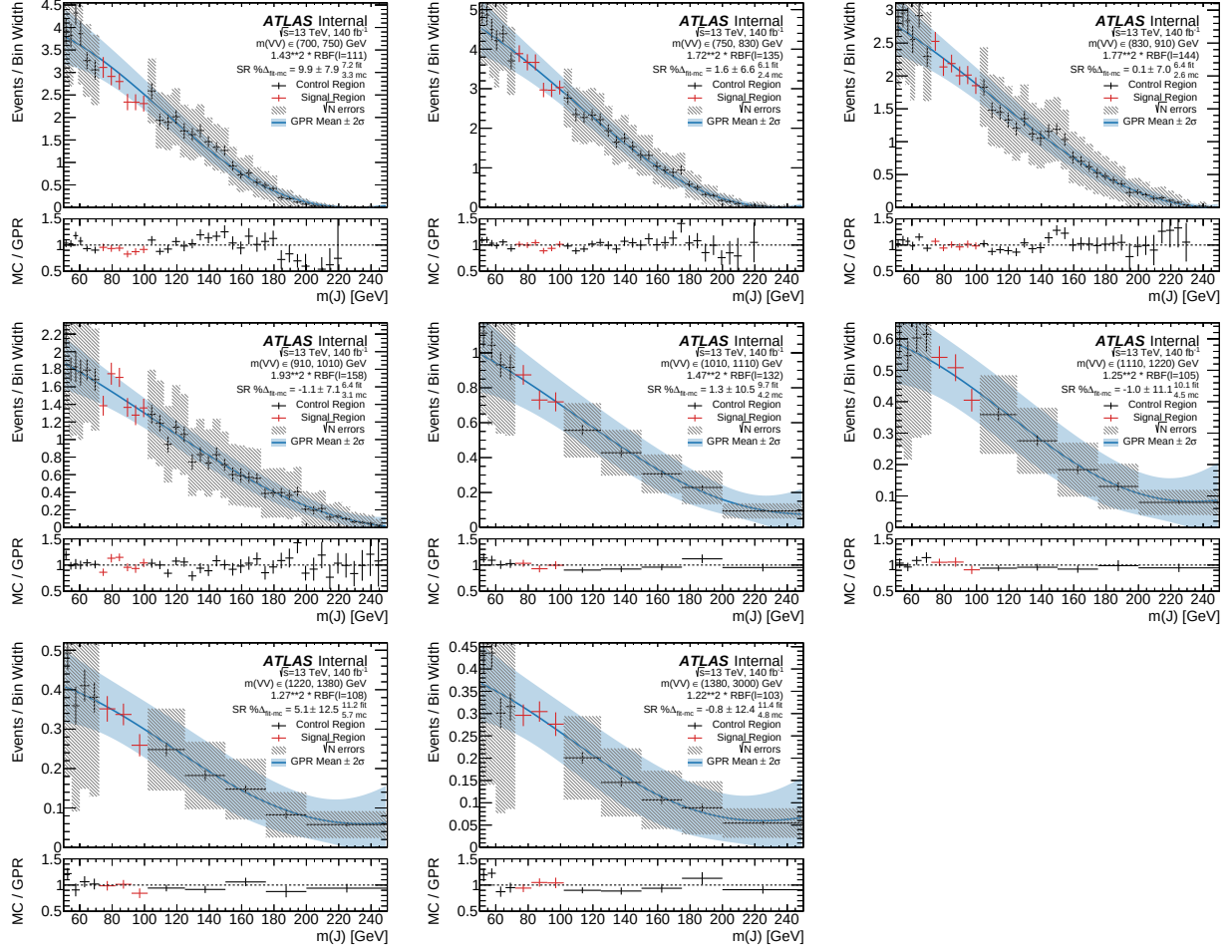


Figure C.14: Gaussian process fits to V+jets MC in the 2-lepton channel.

C.2 Fits to Data

These are Gaussian process regression fits to the event-subtracted MCR, as described in section 9.1.1. In these plots, the black points are the fitted data, and the blue band is the Gaussian process MMLE posterior mean $\pm 2\sigma$. Only fits to the nominal distribution are included; the fits to systematic variations are also omitted. Plots relating to the contour scan are also omitted.

C.2.1 0-lepton Channel

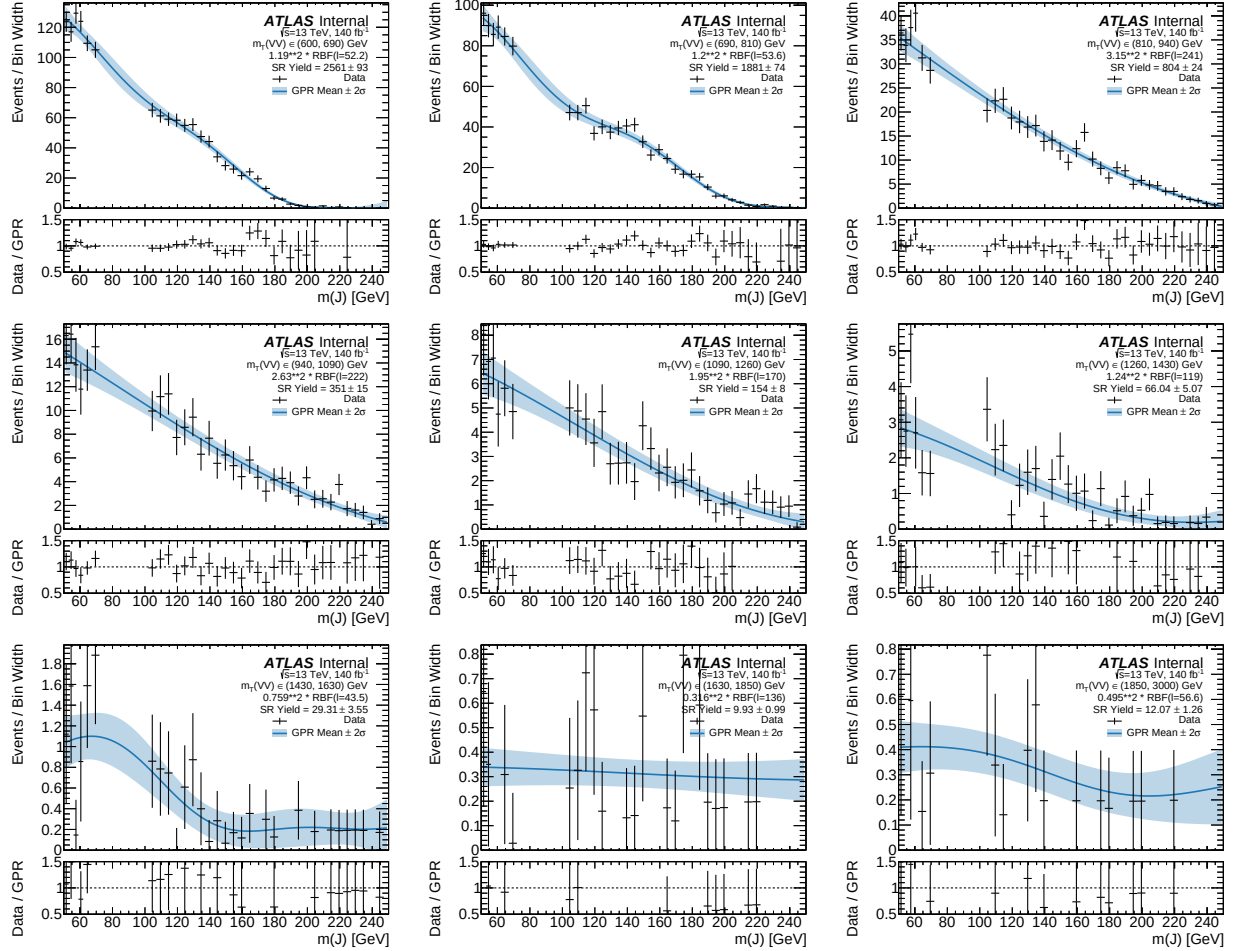


Figure C.15: Gaussian process fits to data in the 0-lepton channel, binned in $m_T(VV)$. These plots show the MMLE, but note that final values are taken from the marginal posterior.

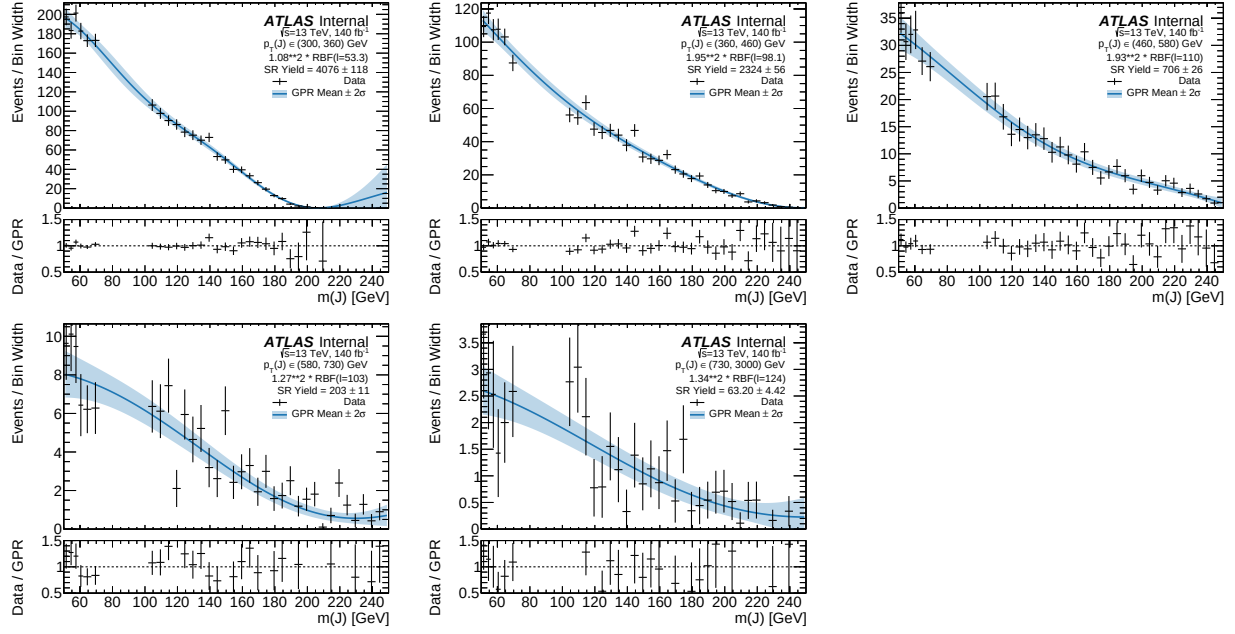


Figure C.16: Gaussian process fits to data in the 0-lepton channel, binned in $p_T(J)$. These plots show the MMLE, but note that final values are taken from the marginal posterior.

C.2.2 1-lepton Channel

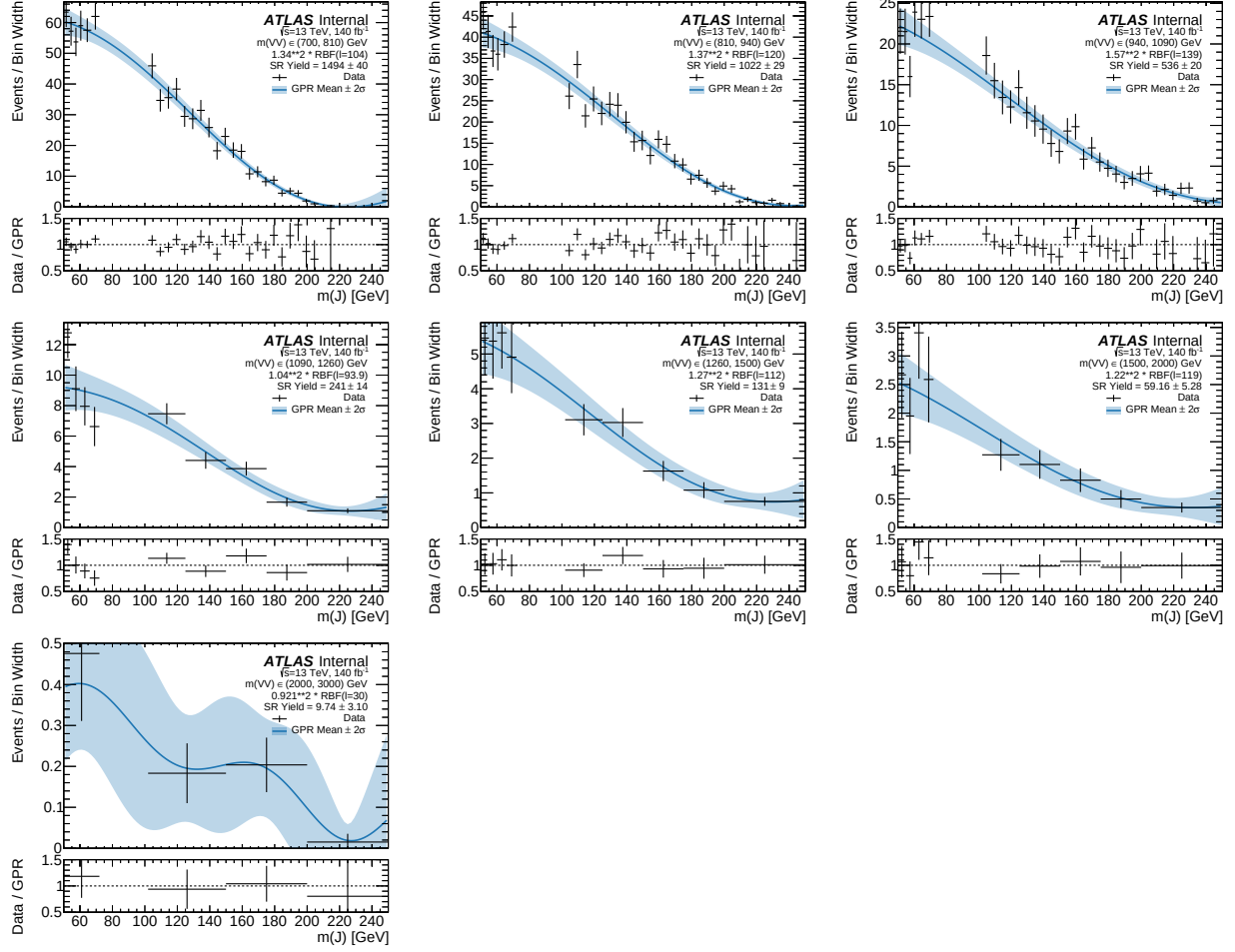


Figure C.17: Gaussian process fits to data in the 1-lepton channel, binned in $m(VV)$. These plots show the MMLE, but note that final values are taken from the marginal posterior.

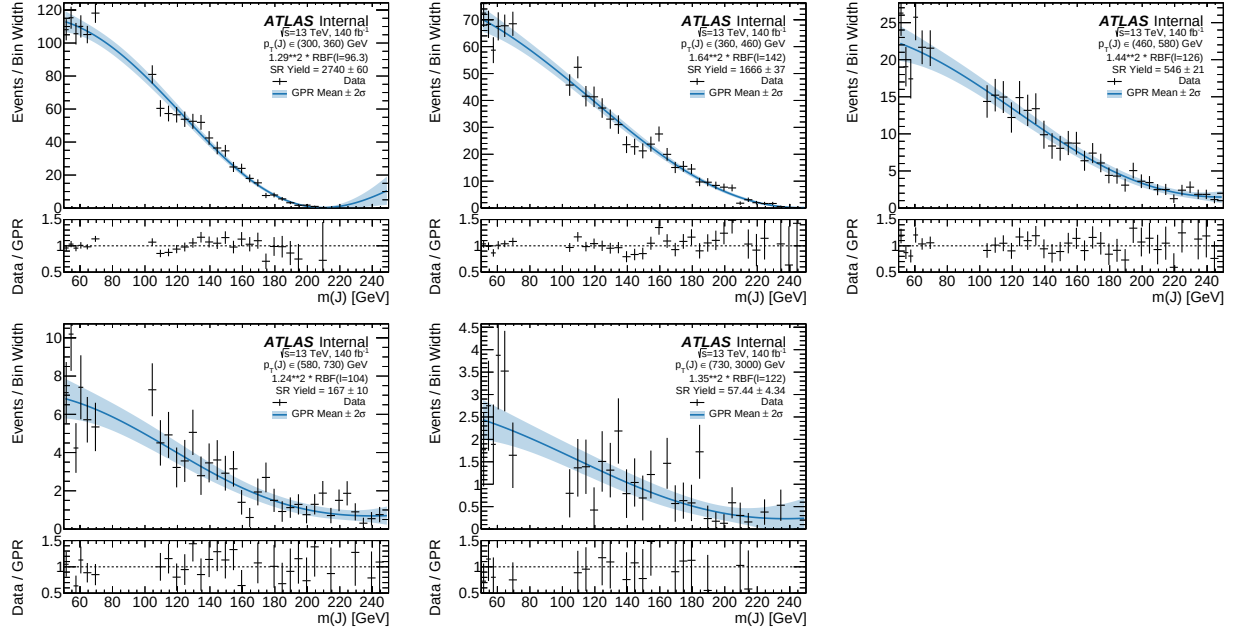


Figure C.18: Gaussian process fits to data in the 1-lepton channel, binned in $p_T(J)$. These plots show the MMLE, but note that final values are taken from the marginal posterior.

C.2.3 2-lepton Channel

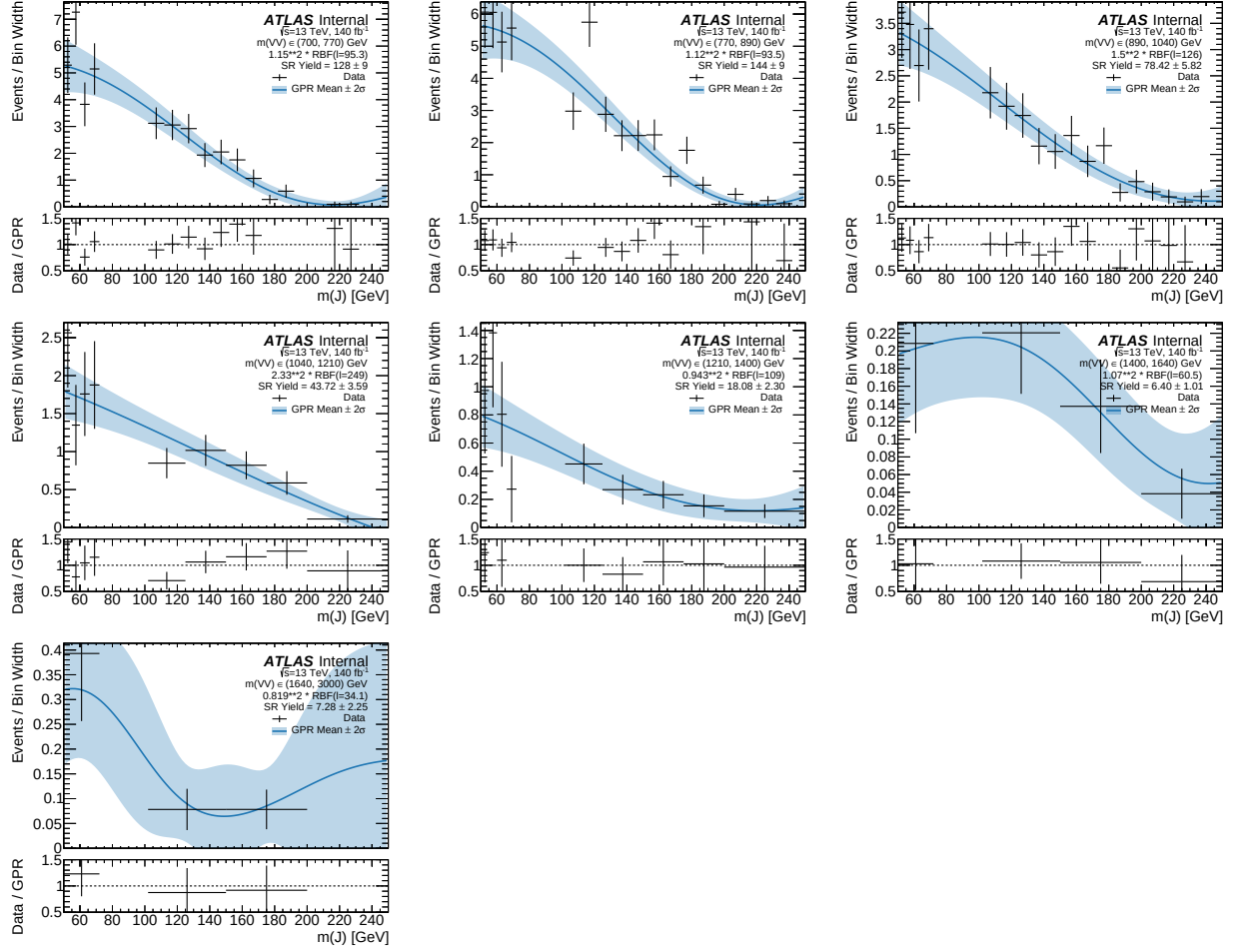


Figure C.19: Gaussian process fits to data in the 2-lepton channel, binned in $m(VV)$. These plots show the MMLE, but note that final values are taken from the marginal posterior.

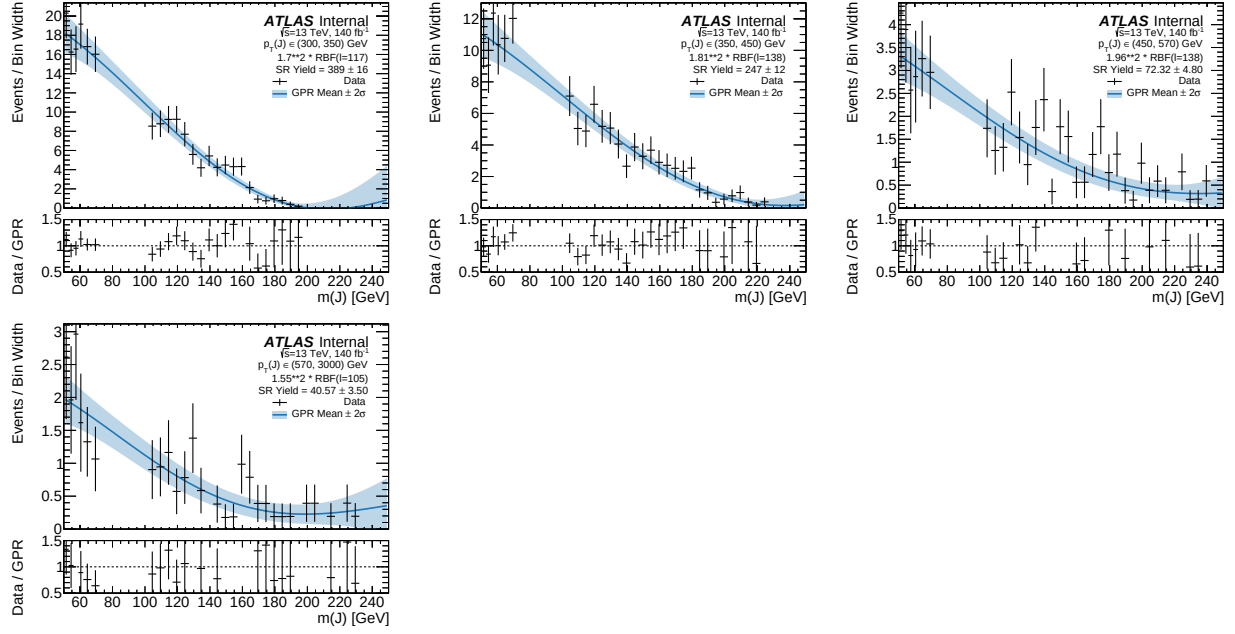


Figure C.20: Gaussian process fits to data in the 2-lepton channel, binned in $p_T(J)$. These plots show the MMLE, but note that final values are taken from the marginal posterior.

APPENDIX D

Glossary

- ALICE** A Large Ion Collider Experiment, one of the four collision-point detectors along the LHC ring, which focuses on ion collisions. 15
- AM** Associative Memory. 39–41, 43, 47, 49
- AMTP** Associative Memory Tracking Processor. 42
- ANN** adversarial neural network. xi, 69, 73, 74, 77–79, 97, 106, 107, 134–138
- aQGC** anomalous quartic gauge coupling. 67
- ASIC** application-specific integrated circuit. 39
- ATCA** Advanced Telecommunications Computing Architecture. 41–43, 45
- ATLAS** A general purpose particle detector on the LHC ring in Switzerland. ix, x, xiii, xv, 1, 2, 10, 11, 14–25, 27–36, 41, 47, 49, 55, 62, 65–68, 71–73, 77, 104, 106, 115, 116, 124, 125, 127, 153
- AXI** Advanced eXtensible Interface, an on-chip communication bus interface protocol. 45
- BSM** beyond the Standard Model. 1, 6, 10–12, 17, 35, 65, 68, 115
- CERN** The European Organization for Nuclear Research, originally known in French as the Conseil européen pour la Recherche nucléaire, an intergovernmental organization that manages the LHC and ATLAS. x, 15, 16, 18, 45
- CMS** The Compact Muon Solenoid, the sister detector to ATLAS. ix, 15, 66, 67, 72, 114
- coverage** In HTT, coverage is the number of training tracks that generated the same pattern. Given limited bank sizes, high coverage patterns were prioritized as they cover more cases. 50, 52, 53
- CSC** Cathode Strip Chamber. 29, 33
- DC** “don’t care”. x, 41, 50–54

EFT effective field theory. ix, x, xiv, xv, 11–14, 104, 107, 109–115, 120, 121, 124, 125

eMMC Embedded MultiMediaCard, a type of non-volatile memory. 44

Event Filter A trigger system in the proposed phase-II upgrade of the ATLAS detector. x, 36–39, 52, 55

fatjet A $R = 1.0$ jet, used to reconstruct hadronically decaying bosons. 34, 35, 72, 96, 108, 110, 111

FPGA field-programmable gate array. 31, 41–43, 55, 57–59

FTK Fast TracKer, a predecessor system to the HTT that was ultimately canceled. xiii, 31, 47

Gaussian process A Bayesian regression method that assumes Gaussian statistics. xi, xii, xv, 2, 71, 79–84, 86–92, 94–96, 98–103, 115–123, 125, 127, 139–151

gHTT global HTT. 37–39

HL-LHC High-Luminosity LHC. x, 2, 17–19, 36, 39, 127

HLT High-Level Trigger. 31, 36

HP Phase space in the VV semileptonic analysis for events passing jet substructure selections. xi, 68, 69, 73, 79, 133, 136–138

HTT Hardware-based Tracking for Trigger. x, xiii, xiv, 36, 37, 39–43, 47, 49–55, 57–60

HTTIF HTT InterFace. 42

HTTSim Simulation framework for the HTT project. x, xiii, xiv, 47–50, 59

hyperparameter In the context of Gaussian process regression, additional parameters that describe the kernel form. xi, 84–87, 92, 93, 95–98, 100, 101

hyperprior In the context of Gaussian process regression, the prior $p(\boldsymbol{\theta})$ on the hyperparameters $\boldsymbol{\theta}$. 85, 97, 101

I2C Inter-Integrated Circuit, a communication bus interface protocol. 46

IBL insertable B-layer. 18, 22

IP interaction point. 19, 20, 22, 33

IP Intellectual property. In the context of Xilinx Vivado, these can be thought of as prepackaged components. 45

IPMC IPMI Controller. 42, 45, 46

IPMI Intelligent Platform Management Interface. 45

ITk Inner Tracker. x, 39–41, 55

L0 Level-0. 36–38, 55

L1 Level-1. 31, 36, 38, 39, 55

LAr liquid argon. 26, 27

LEP Large Electron-Positron Collider. 15

LHC Large Hadron Collider. xiii, xiv, 1–3, 13, 15–21, 30, 127

LHCb One of the four collision-point detectors along the LHC ring, which focuses on b physics. 15

LP Phase space in the VV semileptonic analysis for events failing jet substructure selections. 68

LS2 Long Shutdown 2, an upgrade phase in the LHC program that took place between 2018 and 2022. xiii, xiv, 19

LS3 Long Shutdown 3, an upgrade phase in the LHC program that will commission the HL-LHC, scheduled to start 2026.. xiii, xiv, 19, 36, 39

marginal likelihood In the context of Gaussian process regression, $p(\mathbf{y}|X, \boldsymbol{\theta})$. xi, 83–86, 93, 95–101

marginal posterior In the context of Gaussian process regression, the posterior where the hyperparameters have been marginalized, $p(\mathbf{f}|X, \mathbf{y})$. xi, 85, 86, 93, 97, 98, 100–103, 146–151

MC Monte Carlo. xi, xii, 5, 35, 47, 49, 50, 52, 53, 62, 70–72, 74–79, 96–100, 102, 103, 105–108, 110, 111, 114–117, 119–123, 132, 133, 136–145

MCR mass-sideband control region. xi, 69, 71, 73–75, 77, 103, 115, 116, 146

MDT Monitored Drift Tube. 28, 29, 33

MMLE maximum marginal likelihood estimator. 85, 93, 97, 98, 100–102, 144, 146–151

NSW New Small Wheel. 29, 36

PCB printed circuit board. 42

PDF probability density function. xi, 81, 87, 89–91

PFlow Particle flow, an algorithm used by ATLAS for combining calorimeter and tracking measurements used in jet finding. 34, 35, 67

PL programmable logic. 43, 45

PLU profile likelihood unfolding. xi, 104, 105, 120, 121, 123, 125, 126

PRM Pattern Recognition Mezzanine. 42, 43

PS processing system. 43–45

QCD quantum chromodynamics. 5, 71

QFT quantum field theory. 3, 5, 8

QSPI Quad Serial Peripheral Interface, a communication protocol. 44

RBF radial basis function. ix, xi, xii, 83, 84, 87, 91–93, 95–98, 100, 101, 103, 139, 140

RDO Raw Data Object. 49

rHTT regional HTT. 37, 38

road In the context of tracking in the trigger, a road is a set of hits for each layer, corresponding to track candidates. Because each road can have multiple hits in the same layer, there are a combinatoric number of track candidates per road. In HTT, a road is a matched pattern. 40, 41, 47, 60

RPC Resistive Plate Chamber. 29, 33

SCT Semiconductor Tracker. 22, 24, 32, 39

SM Standard Model. ix–xi, 1–7, 9, 10, 12–14, 35, 71, 104, 107, 109–115, 121, 122, 125, 127

SMEFT Standard Model Effective Field Theory. ix, xi, 1–3, 12–14, 65, 66, 115, 121, 127–129

SoC system on a chip. ix, x, 42–46

SSH Secure Shell Protocol. 44

SSID superstrip identifier. 40, 41, 43, 49, 51

SSTP Second-Stage Tracking Processor. 42

TCR top control region. 69, 116, 117, 119

TDAQ Trigger and Data Acquisition. x, 20, 31, 36–38, 42, 55

TDR Technical Design Report. 55

TFM Track-Fitting Mezzanine. 42, 43, 45

TGC Thin Gap Chamber. 29, 33

three-variable tagger A W/Z fatjet tagger that cuts on D_2 , m , and the number of tracks. xi, 69, 73–75, 77, 133

TP Tracking Processor. ix, x, 42–46

TRT Transition Radiation Tracker. 22, 24, 28, 32, 39

TSP thin-strip. 50–53

UFO Unified Flow Object. 35, 67

V+jets Events with a W or Z boson in association with QCD jets. xi, xii, xv, 66, 69–71, 80, 81, 84, 86, 103, 114–117, 119, 120, 132, 144, 145

VBS vector boson scattering. 67–69

VV+VH Semileptonic An analysis investigating boosted dibosons decaying semileptonically at the ATLAS detector, the primary physics analysis of this thesis. ix, xi, xiv, 1, 2, 65, 73, 104, 130–132

Bibliography

- [1] M. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge University Press, 2014, ISBN: 9781107034730 (cit. on pp. 3, 5, 8).
- [2] T. Lancaster and S. Blundell, *Quantum Field Theory for the Gifted Amateur*, OUP Oxford, 2014, ISBN: 9780199699339 (cit. on pp. 3, 5, 8).
- [3] I. Brivio and M. Trott, *The standard model as an effective field theory*, Physics Reports **793** (2019) 1, ISSN: 0370-1573 (cit. on p. 3).
- [4] Cush, *Standard Model of Elementary Particles*, (2019),
URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg
(visited on 03/04/2024) (cit. on p. 4).
- [5] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on p. 5).
- [6] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics*, Phys. Rev. **76** (6 1949) 769 (cit. on p. 5).
- [7] *First Published Feynman Diagram, 1949*,
URL: https://history.fnal.gov/virtual/feynman_diagram.html (visited on 03/05/2024)
(cit. on p. 5).
- [8] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5: going beyond*, Journal of High Energy Physics **2011** (2011), ISSN: 1029-8479 (cit. on p. 5).
- [9] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Physics **7** (2019), ISSN: 2542-4653 (cit. on p. 5).
- [10] R. L. Workman et al., *Review of Particle Physics*, PTEP **2022** (2022) 083C01 (cit. on pp. 6, 67).
- [11] CDF Collaboration, *High-precision measurement of the W boson mass with the CDF II detector*, Science **376** (2022) 170 (cit. on p. 6).
- [12] D. P. Aguillard et al.,
Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm,
Phys. Rev. Lett. **131** (2023) 161802, arXiv: 2308.06230 [hep-ex] (cit. on p. 6).
- [13] J. C. Romao and J. P. Silva, *A resource for signs and Feynman diagrams of the Standard Model*, Int. J. Mod. Phys. A **27** (2012) 1230025, arXiv: 1209.6213 [hep-ph] (cit. on p. 10).
- [14] T. D. Gutierrez, *Standard Model Lagrangian (density)*, 1999,
URL: <http://nuclear.ucdavis.edu/~tgutierrr/files/stmL1.html> (visited on 03/07/2024)
(cit. on p. 10).
- [15] J. Woithe, G. J. Wiener, and F. F. Van der Veken,
Let's have a coffee with the Standard Model of particle physics!, Phys. Educ. **52** (2017) 034001
(cit. on p. 10).
- [16] S. P. Martin, *A Supersymmetry primer*,
Adv. Ser. Direct. High Energy Phys. **18** (1998) 1, ed. by G. L. Kane, arXiv: hep-ph/9709356
(cit. on p. 10).

- [17] J. N. Butler et al.,
Report of the 2021 U.S. Community Study on the Future of Particle Physics (Snowmass 2021), (2023) (cit. on p. 10).
- [18] ATLAS Collaboration, *SUSY August 2023 Summary Plot Update*, (2023),
URL: <http://cds.cern.ch/record/2871728> (cit. on p. 11).
- [19] E. Fermi, *Tentativo di una Teoria Dei Raggi β* , *Il Nuovo Cimento* **11** (1934) 1 (cit. on p. 12).
- [20] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek,
Dimension-Six Terms in the Standard Model Lagrangian, *JHEP* **10** (2010) 085,
arXiv: 1008.4884 [hep-ph] (cit. on pp. 13, 113, 128, 129).
- [21] C. W. Murphy, *Dimension-8 operators in the Standard Model Effective Field Theory*,
JHEP **10** (2020) 174, arXiv: 2005.00059 [hep-ph] (cit. on p. 13).
- [22] A. Azatov, R. Contino, C. S. Machado, and F. Riva,
Helicity selection rules and noninterference for BSM amplitudes, *Phys. Rev. D* **95** (2017) 065014,
arXiv: 1607.05236 [hep-ph] (cit. on pp. 14, 112).
- [23] ATLAS Collaboration, *Combined effective field theory interpretation of Higgs boson and weak boson production and decay with ATLAS data and electroweak precision observables*, (2022),
URL: <https://cds.cern.ch/record/2816369> (cit. on pp. 14, 124, 125).
- [24] L. R. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001, This report is an abridged version of the LHC Design Report (CERN-2004-003), URL: <https://cds.cern.ch/record/1129806> (cit. on p. 15).
- [25] *LEP design report*, Report, Geneva: CERN, 1984, URL: <https://cds.cern.ch/record/102083> (cit. on p. 15).
- [26] *Design Report Tevatron 1 project*, (1984) (cit. on p. 15).
- [27] *CERN's accelerator complex*, (2024), URL: <https://www.home.cern/science/accelerators/accelerator-complex> (visited on 03/11/2024) (cit. on p. 15).
- [28] *Accelerating: Radiofrequency cavities*, (2024),
URL: <https://home.cern/science/engineering/accelerating-radiofrequency-cavities> (visited on 03/11/2024) (cit. on p. 15).
- [29] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, (2022), General Photo, URL: <https://cds.cern.ch/record/2800984> (cit. on p. 16).
- [30] J. Gareyte, *LHC main parameters*, *Part. Accel.* **50** (1995) 61,
URL: <https://cds.cern.ch/record/304825> (cit. on p. 16).
- [31] ATLAS Collaboration,
Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC, *Eur. Phys. J. C* **83** (2023) 982, arXiv: 2212.09379 [hep-ex] (cit. on p. 17).
- [32] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-3 of the LHC*, 2024,
URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3> (visited on 03/11/2023) (cit. on pp. 17, 18).
- [33] I. Zurbano Fernandez et al.,
High-Luminosity Large Hadron Collider (HL-LHC): Technical design report, **10/2020** (2020), ed. by I. Béjar Alonso et al. (cit. on p. 18).
- [34] *The HL-LHC project*, 2024,
URL: <https://hilumilhc.web.cern.ch/content/hl-lhc-project> (visited on 03/12/2024) (cit. on p. 19).

- [35] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 19, 20, 27).
- [36] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3*, (2023), arXiv: 2305.16623 [physics.ins-det] (cit. on pp. 19, 25, 28).
- [37] J. Pequenaio and P. Schaffner,
“How ATLAS detects particles: diagram of particle paths in the detector”, 2013,
URL: <https://cds.cern.ch/record/1505342> (cit. on p. 21).
- [38] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*,
Technical design report. ATLAS, Geneva: CERN, 1997,
URL: <https://cds.cern.ch/record/331063> (cit. on p. 22).
- [39] ATLAS Collaboration,
Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2,
Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on p. 23).
- [40] ATLAS Collaboration, *Study of the material of the ATLAS inner detector for Run 2 of the LHC*,
JINST **12** (2017) P12009, arXiv: 1707.02826 [hep-ex] (cit. on p. 23).
- [41] M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, (2010),
URL: <https://cds.cern.ch/record/1291633> (cit. on p. 22).
- [42] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 689,
arXiv: 2007.02645 [hep-ex] (cit. on p. 25).
- [43] A. Artamonov et al., *The ATLAS forward calorimeters*, JINST **3** (2008) P02010 (cit. on p. 27).
- [44] J. Wotschack,
ATLAS Muon Chamber Construction Parameters for CSC, MDT, and RPC chambers, (2008)
(cit. on pp. 28, 29).
- [45] T. Kawamoto et al., *New Small Wheel Technical Design Report*, (2013),
URL: <https://cds.cern.ch/record/1552862> (cit. on p. 29).
- [46] K. Nagai, *Thin gap chambers in ATLAS*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **384** (1996) 219,
ISSN: 0168-9002 (cit. on p. 29).
- [47] ATLAS Outreach,
ATLAS Fact Sheet : To raise awareness of the ATLAS detector and collaboration on the LHC, 2010,
URL: <https://cds.cern.ch/record/1457044> (cit. on p. 30).
- [48] M. Aleksa et al., *Results of the ATLAS solenoid magnetic field map*,
J. Phys.: Conf. Ser. **110** (2008) 092018, URL: <https://cds.cern.ch/record/1178488>
(cit. on p. 30).
- [49] ATLAS Collaboration, *ApprovedPlotsDAQ*, 2020,
URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ> (visited on 03/15/2024) (cit. on p. 31).
- [50] ATLAS Collaboration, *Software Performance of the ATLAS Track Reconstruction for LHC Run 3*,
(2021), URL: <https://cds.cern.ch/record/2766886> (cit. on pp. 32, 33).
- [51] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,
Detectors and Associated Equipment **262** (1987) 444, ISSN: 0168-9002 (cit. on p. 32).
- [52] T. G. Cornelissen et al., *The global χ^2 track fitter in ATLAS*,
J. Phys. Conf. Ser. **119** (2008) 032013, ed. by R. Sobie, R. Tafirout, and J. Thomson (cit. on p. 32).

- [53] ATLAS Collaboration, *Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment*, (2023), arXiv: 2308.13362 [hep-ex] (cit. on pp. 32, 33).
- [54] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 32).
- [55] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on p. 33).
- [56] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2021) 578, arXiv: 2012.00578 [hep-ex] (cit. on p. 34).
- [57] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 34).
- [58] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 34).
- [59] ATLAS Collaboration, *Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions*, Eur. Phys. J. C **81** (2021) 334, arXiv: 2009.04986 [hep-ex] (cit. on p. 35).
- [60] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv: 1802.08168 [hep-ex] (cit. on p. 35).
- [61] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on p. 35).
- [62] S. Agostinelli et al., *GEANT4 - a simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250 (cit. on p. 35).
- [63] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*, tech. rep., CERN, 2017, URL: <https://cds.cern.ch/record/2285584> (cit. on pp. 36–38, 43, 55, 59).
- [64] ATLAS Collaboration, *ATLAS Phase-II Upgrade Scoping Document*, (2015) (cit. on p. 36).
- [65] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Strip Detector*, (2017), URL: <https://cds.cern.ch/record/2257755> (cit. on pp. 39, 40).
- [66] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, (2017) (cit. on p. 39).
- [67] A. Annovi et al., *AM06: the Associative Memory chip for the Fast TracKer in the upgraded ATLAS detector*, (2016), URL: <https://cds.cern.ch/record/2228284> (cit. on p. 39).
- [68] F. Gray, *Pulse Code Communication*, U.S. Patent 2,632,058. 1953 (cit. on p. 41).
- [69] *AdvancedTCA Base Specification PICMG 3.0 Revision 3.0*, 2015 (cit. on p. 41).
- [70] *Zynq UltraScale+ MPSoC Data Sheet: Overview*, 2021, URL: <https://xilinx.com/products/silicon-devices/soc/zynq-ultrascale-mpsoc.html> (cit. on p. 43).
- [71] Enclustra, *Mercury XU5*, 2024, URL: <https://www.enclustra.com/en/products/system-on-chip-modules/mercury-xu5/> (visited on 03/22/2024) (cit. on p. 43).

- [72] *PetaLinux Tools Documentation*, 2023, URL: <https://docs.amd.com/r/en-US/ug1144-petalinux-tools-reference-guide/Introduction> (cit. on p. 44).
- [73] *fdisk(8) – Linux manual page*, 2023, URL: <https://man7.org/linux/man-pages/man8/fdisk.8.html> (visited on 03/22/2024) (cit. on p. 44).
- [74] *Aurora 64B/66B*, 2024, URL: <https://www.xilinx.com/products/intellectual-property/aurora64b66b.html> (cit. on p. 45).
- [75] *Intelligent Platform Management Interface Specification Second Generation v2.0*, 2013 (cit. on p. 45).
- [76] J. M. Mendez et al., *CERN-IPMC solution for AdvancedTCA blades*, PoS **TWEPP-17** (2018) 053. 5 p (cit. on p. 45).
- [77] *ADT7420 Data Sheet*, 2017, URL: <https://analog.com/en/products/adt7420.html> (cit. on p. 46).
- [78] *LTC2945 Wide Range I2C Power Monitor*, 2012, URL: <https://analog.com/en/products/ltc2945.html> (cit. on p. 46).
- [79] ATLAS Collaboration, *Athena*, 2024, URL: <https://gitlab.cern.ch/atlas/athena> (cit. on p. 47).
- [80] C. ATLAS, *Technical Design Report for the Phase-II Upgrade of the ATLAS Trigger and Data Acquisition System - Event Filter Tracking Amendment*, (2022), URL: <https://cds.cern.ch/record/2802799> (cit. on pp. 55, 60).
- [81] P. V. Hough, *Method and means for recognizing complex patterns*, U.S. Patent 3,069,654. 1962 (cit. on p. 55).
- [82] R. O. Duda and P. E. Hart, *Use of the Hough transformation to detect lines and curves in pictures*, Commun. ACM **15** (1972) 11, ISSN: 0001-0782 (cit. on p. 55).
- [83] J. Gradin, M. Mårtensson, and R. Brenner, *Comparison of two hardware-based hit filtering methods for trackers in high-pileup environments*, JINST **13** (2018) P04019, arXiv: 1709.01034 [physics.ins-det] (cit. on pp. 55, 59).
- [84] M. Mårtensson, “A search for leptoquarks with the ATLAS detector and hardware tracking at the High-Luminosity LHC”, PhD thesis: Uppsala University, 2019 (cit. on pp. 55, 59).
- [85] ATLAS Collaboration, *Search for heavy WW/WZ/ZZ/WH/ZH resonances in semi-leptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (Manuscript in preparation) (cit. on p. 65).
- [86] ATLAS Collaboration, *Search for heavy diboson resonances in semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 1165, arXiv: 2004.14636 [hep-ex] (cit. on pp. 65, 68).
- [87] ATLAS Collaboration, *Search for heavy resonances decaying into a Z or W boson and a Higgs boson in final states with leptons and b-jets in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **06** (2023) 016, arXiv: 2207.00230 [hep-ex] (cit. on p. 65).
- [88] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **816** (2021) 136204, arXiv: 2008.02508 [hep-ex] (cit. on p. 65).
- [89] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, *Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory*, JHEP **04** (2021) 279, arXiv: 2012.02779 [hep-ph] (cit. on p. 66).

- [90] D. Liu and L.-T. Wang, *Prospects for precision measurement of diboson processes in the semileptonic decay channel in future LHC runs*, Phys. Rev. D **99** (2019) 055001, arXiv: 1804.08688 [hep-ph] (cit. on pp. 66, 113).
- [91] ATLAS Collaboration, *Measurement of $WW/WZ \rightarrow \ell\nu qq'$ production with the hadronically decaying boson reconstructed as one or two jets in pp collisions at $\sqrt{s} = 8$ TeV with ATLAS, and constraints on anomalous gauge couplings*, Eur. Phys. J. C **77** (2017) 563, arXiv: 1706.01702 [hep-ex] (cit. on p. 66).
- [92] ATLAS Collaboration, *Measurements of W^+W^- production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2023) (cit. on p. 67).
- [93] ATLAS Collaboration, *Observation of gauge boson joint-polarisation states in $W^\pm Z$ production from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **843** (2023) 137895, arXiv: 2211.09435 [hep-ex] (cit. on p. 67).
- [94] ATLAS Collaboration, *Measurements of differential cross-sections in four-lepton events in 13 TeV proton-proton collisions with the ATLAS detector*, JHEP **07** (2021) 005, arXiv: 2103.01918 [hep-ex] (cit. on p. 67).
- [95] ATLAS Collaboration, *Measurements of $W^+W^- + \geq 1$ jet production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **06** (2021) 003, arXiv: 2103.10319 [hep-ex] (cit. on p. 67).
- [96] ATLAS Collaboration, *Differential cross-section measurements for the electroweak production of dijets in association with a Z boson in proton-proton collisions at ATLAS*, Eur. Phys. J. C **81** (2021) 163, arXiv: 2006.15458 [hep-ex] (cit. on p. 67).
- [97] ATLAS Collaboration, *Measurement of $Z\gamma\gamma$ production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **83** (2023) 539, arXiv: 2211.14171 [hep-ex] (cit. on p. 67).
- [98] ATLAS Collaboration, *Differential cross-section measurements of the production of four charged leptons in association with two jets using the ATLAS detector*, JHEP **01** (2024) 004, arXiv: 2308.12324 [hep-ex] (cit. on p. 67).
- [99] ATLAS Collaboration, *Measurement and interpretation of same-sign W boson pair production in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2023) (cit. on p. 67).
- [100] A. M. Sirunyan et al., *W^+W^- boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. D **102** (2020) 092001, arXiv: 2009.00119 [hep-ex] (cit. on p. 67).
- [101] A. Tumasyan et al., *Measurement of the inclusive and differential WZ production cross sections, polarization angles, and triple gauge couplings in pp collisions at $\sqrt{s} = 13$ TeV*, JHEP **07** (2022) 032, arXiv: 2110.11231 [hep-ex] (cit. on p. 67).
- [102] A. Tumasyan et al., *Measurement of $W^\pm\gamma$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV and effective field theory constraints*, Phys. Rev. D **105** (2022) 052003, arXiv: 2111.13948 [hep-ex] (cit. on pp. 67, 114).
- [103] A. M. Sirunyan et al., *Measurement of the cross section for electroweak production of a Z boson, a photon and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on anomalous quartic couplings*, JHEP **06** (2020) 076, arXiv: 2002.09902 [hep-ex] (cit. on p. 67).
- [104] A. Tumasyan et al., *Measurement of the electroweak production of $W\gamma$ in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. D **108** (2023) 032017, arXiv: 2212.12592 [hep-ex] (cit. on p. 67).
- [105] A. M. Sirunyan et al., *Measurements of production cross sections of WZ and same-sign WW boson pairs in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **809** (2020) 135710, arXiv: 2005.01173 [hep-ex] (cit. on p. 67).

- [106] A. M. Sirunyan et al., *Evidence for electroweak production of four charged leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **812** (2021) 135992, arXiv: 2008.07013 [hep-ex] (cit. on p. 67).
- [107] A. M. Sirunyan et al., *Search for anomalous electroweak production of vector boson pairs in association with two jets in proton-proton collisions at 13 TeV*, Phys. Lett. B **798** (2019) 134985, arXiv: 1905.07445 [hep-ex] (cit. on p. 67).
- [108] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, Eur. Phys. J. C **83** (2023) 681, arXiv: 2211.16345 [physics.data-an] (cit. on p. 67).
- [109] ATLAS Collaboration, *ATLAS simulation of boson plus jets processes in Run 2*, ATL-PHYS-PUB-2017-006, 2017, URL: <https://cds.cern.ch/record/2261937> (cit. on p. 71).
- [110] ATLAS Collaboration, *Cross-section measurements for the production of a Z boson in association with high-transverse-momentum jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **06** (2023) 080, arXiv: 2205.02597 [hep-ex] (cit. on p. 71).
- [111] ATLAS Collaboration, *Measurement of the cross-section of high transverse momentum vector bosons reconstructed as single jets and studies of jet substructure in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, New J. Phys. **16** (2014) 113013, arXiv: 1407.0800 [hep-ex] (cit. on p. 72).
- [112] CMS Collaboration, *Search for a heavy vector resonance decaying to a Z boson and a Higgs boson in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2021) 688, arXiv: 2102.08198 [hep-ex] (cit. on pp. 72, 78).
- [113] ATLAS Collaboration, *Performance of W/Z taggers using UFO jets in ATLAS*, (2021) (cit. on pp. 73, 77).
- [114] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, JHEP **09** (2013) 076, arXiv: 1306.4945 [hep-ex] (cit. on p. 74).
- [115] ATLAS Collaboration, *Recommendations for the Modeling of Smooth Backgrounds*, (2020) (cit. on p. 78).
- [116] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006, ISBN: 0-262-18253-X (cit. on pp. 80, 83, 85, 96).
- [117] A. M. G. Deborah Hughes-Hallett and W. G. McCallum, *Calculus: Single and Multivariable, 8th Edition*, Wiley, 2021, ISBN: 978-1-119-69429-8 (cit. on p. 85).
- [118] M. Taboga, *Linear combinations of normal random variables*, 2021, URL: <https://www.statlect.com/probability-distributions/normal-distribution-linear-combinations> (visited on 07/31/2023) (cit. on p. 86).
- [119] *Project Jupyter*, 2023, URL: <https://jupyter.org> (visited on 07/20/2023) (cit. on p. 86).
- [120] C. R. Harris et al., *Array programming with NumPy*, Nature **585** (2020) 357 (cit. on p. 86).
- [121] J. E. Gentle, *Computational Statistics*, Springer, 2009, ISBN: 9780387981444 (cit. on p. 87).
- [122] S. Schmitt, *Data Unfolding Methods in High Energy Physics*, EPJ Web Conf. **137** (2017) 11008, ed. by Y. Foka, N. Brambilla, and V. Kovalenko, arXiv: 1611.01927 [physics.data-an] (cit. on p. 104).
- [123] J. Chan and B. Nachman, *Unbinned profiled unfolding*, Phys. Rev. D **108** (2023) 016002, arXiv: 2302.05390 [hep-ph] (cit. on p. 104).
- [124] ATLAS Collaboration, *Inclusive and differential cross-section measurements of $t\bar{t}Z$ production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, including EFT and spin-correlation interpretations*, (2023), arXiv: 2312.04450 [hep-ex] (cit. on pp. 104, 106).

- [125] ATLAS Collaboration, *Measurement of the total and differential cross-sections of $t\bar{t}W$ production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2024), arXiv: 2401.05299 [hep-ex] (cit. on p. 104).
- [126] ATLAS Collaboration, *Measurements of inclusive and differential cross-sections of $t\bar{t}\gamma$ production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2024), arXiv: 2403.09452 [hep-ex] (cit. on p. 104).
- [127] G. Panico, F. Riva, and A. Wulzer, *Diboson interference resurrection*, Phys. Lett. B **776** (2018) 473, arXiv: 1708.07823 [hep-ph] (cit. on p. 112).
- [128] G. F. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi, *The Strongly-Interacting Light Higgs*, JHEP **06** (2007) 045, arXiv: hep-ph/0703164 (cit. on p. 113).
- [129] J. Elias-Miro, J. R. Espinosa, E. Masso, and A. Pomarol, *Higgs windows to new physics through $d=6$ operators: constraints and one-loop anomalous dimensions*, JHEP **11** (2013) 066, arXiv: 1308.1879 [hep-ph] (cit. on p. 113).
- [130] S. Willenbrock and C. Zhang, *Effective Field Theory Beyond the Standard Model*, Ann. Rev. Nucl. Part. Sci. **64** (2014) 83, arXiv: 1401.0470 [hep-ph] (cit. on p. 113).
- [131] Q.-H. Cao, Y. Liu, and S.-R. Yuan, *Verification of Equivalence of Operator Bases in Unitarity Analysis*, (2023), arXiv: 2309.14079 [hep-ph] (cit. on p. 113).
- [132] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur.Phys.J.C 73, 2501 (2013)], arXiv: 1007.1727 [physics.data-an] (cit. on pp. 115, 117, 120).
- [133] G. Aad et al., *Measurement of single top-quark production in association with a W boson in the single-lepton channel at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 720, arXiv: 2007.01554 [hep-ex] (cit. on p. 116).
- [134] G. Aad et al., *Measurement of single top-quark production in the s -channel in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **06** (2023) 191, arXiv: 2209.08990 [hep-ex] (cit. on p. 116).
- [135] R. Barlow, *Extended maximum likelihood*, Nucl. Instrum. Methods Phys. Res. A **297** (1990) 496, ISSN: 0168-9002 (cit. on p. 119).
- [136] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219, ISSN: 0010-4655 (cit. on p. 120).