SEARCHES FOR NEW PHYSICS USING JETS WITH THE ATLAS DETECTOR

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Aviv Ruben Cukierman
June 2020

This dissertation is online at: http://purl.stanford.edu/kq805mr1848

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Ariel Schwartzman, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Patricia Burchat

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Dong Su

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Natalia Toro

Approved for the Stanford University Committee on Graduate Studies.

Stacey F. Bent, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

The Large Hadron Collider produces particle collisions at the highest energies ever observed in a scientific experiment. This apparatus is built to test the predictions of the Standard Model of Particle Physics, including the existence and properties of the Higgs boson. As a proton-proton collider, quarks and gluons are produced in abundance, which quickly fragment and hadronize into collimated showers of energy called jets. These jets are detected and measured in the ATLAS detector, which is built around the point of the proton-proton collisions to observe the products of these interactions.

Three major original research efforts are presented using data from proton-proton collisions observed in ATLAS. The first analyzes events with jets and photons to search for a beyond-the-Standard-Model decay of the Higgs boson. The second utilizes novel techniques in weak supervision to perform a generic data-driven resonance search in events with two jets. The third formalizes the calibration of the jet energies observed in the ATLAS detector, and further proposes a new method to improve this calibration with machine learning.

The work presented here addresses some of the key questions in particle physics today. By searching for new physics, it is possible to shed light on the nature of the Higgs boson and the possibility of physics beyond the Standard Model. These searches focus on processes involving multiple jets in the final state, which motivates innovations in the reconstruction of jet energies. In addition to setting new bounds on theoretically interesting models, the innovations in object reconstruction and analysis techniques developed in this work can be applied in other ATLAS efforts using currently available data or data gathered in the future.

"Your mother needs three doctors in the family to keep up her health."
— Valentin Cukierman

Acknowledgements

They say it takes a village to raise a PhD student. While many people have contributed in some way to this milestone, my PhD experience was characterized by a few intense relationships which resulted in extraordinary research output.

The first acknowledgement goes to Ariel. When I first started in the SLAC ATLAS group I had some vague notions about what I wanted to work on over the course of my degree. Ariel inundated me with new creative ideas, far more than I could possibly work on, and all of them presenting exciting new avenues of discovery. As an adviser, Ariel provided me not only with guidelines of interesting things to look into, but also with the freedom to pursue my own interests or go off on unpredicted tangential projects, for which I am extremely grateful. These tangential projects turned into some of the most valuable original research and publications presented here.

Perhaps no one single person is more responsible for helping me attain my PhD than Ben. As first a more senior student, and then an easily accessible postdoc, Ben provided both a mentorship and peer role which was immensely helpful for me to understand my own ideas and how they could be used and communicated in the community. Of the seven publications I have been majorly involved with over the course of my PhD, three of them were produced by our two-person team, including three of the four key chapters in this Thesis, and an additional two were produced in small teams involving the two of us. This represents a remarkable collaboration, and my thanks goes to Ben for being such an excellent research partner.

The fourth key chapter resulted from another two-person team involving Francesco. Francesco served as a close mentor and colleague during the key transition period between a student and a researcher that is the goal of a PhD. Through him I learned

what questions were the right ones to ask and how to do the work necessary to answer them satisfactorily. My thanks goes to Francesco for providing the scaffolding for me to grow into a fully-fledged researcher.

While my relationships to the following people were not as intense as those mentioned above, they all contributed in some way and can rest assured that I would not have been as successful without them.

My thanks goes to the other senior members of the SLAC ATLAS group - Su Dong, Rainer, Charlie, Michael, Caterina, Lauren, Catrin, and Nicoletta - who provided much-needed feedback on my work from a senior perspective. I have special gratitude to the SLAC computing team, who provided an excellent service which enabled my computationally-heavy research and allowed me to almost entirely avoid using the CERN grid over the course of my PhD. There were the many postdocs and older students who served as more on-the-ground sources of knowledge: Max, Qi, Zihao, Rafael, Valentina, Zijun, Pascal, Katie, and Ke. Finally, there were the younger students who helped me clarify my ideas and also reminded me to lighten up a little bit: Nicole, Murtaza, Jannicke, Sanha, and Rachel.

I am extremely grateful for my reading committee, which other than Ariel and Su Dong comprises Pat Burchat and Natalia Toro, and for my defense chair David Reis. Their valuable feedback was necessary for communicating my ideas to a wider audience.

Some special thanks goes to Adrienne and Maria for making sure I got paid, stayed on track with my requirements, and graduated on time.

There is the entire ATLAS collaboration, including too many individuals to mention with whom I spent an inordinate amount of time speaking to remotely and in person. These include in particular members of the JET/MET group, or anyone who attended HCW, HFSF, or BOOST, and with whom I had detailed conversations about exciting projects. There are also the Exotics and HDBS groups, whose conveners and members allowed us to pursue our non-standard analysis ideas and supported us through the byzantine ATLAS publication process.

Before joining ATLAS I already had some research projects under my belt, and my thanks goes to Mike, Adam, and Peter for giving me valuable research experience

which set me up for success in my PhD.

All of the professors and teachers who taught me how to learn in graduate school, undergraduate, and even high school and middle school laid the foundations for me being able to discover new knowledge via research.

Finally I turn to personal relationships. Celine has been a constant source of happiness in my life for the past decade, and I can only hope we continue to make each other happy forever. My thanks also go out to all my other friends from Stanford, MIT, and from TJ or before who never let me down when I was looking to destress and take some distance from school. Ari has been there my whole life and probably understands me better than anyone - I couldn't ask for a brother whose thought processes are so similar and yet different enough to my own to allow for such spirited and deep questions and discussions. The original acknowledgements of course go to my very cool and totally normal parents, who raised me to be kind, generous, and responsible, provided support and love my entire life, and taught me to always exceed expectations and to constantly look for ways to grow and better myself.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Over the past century, a remarkable picture has emerged about the nature of the universe - that fundamentally, everything we observe around us, and every way in which those things interact, is governed by elementary *particles*, which are themselves local perturbations in quantum *fields*. The theory that describes these particles and their interactions, called the *Standard Model*, makes many predictions which have been confirmed to be accurate by extremely precise measurements. One prominent example of one of these predictions is the existence of the *Higgs boson* [21–24], the discovery of which [25, 26] earned a Nobel Prize in Physics [27]. However, there are still big unanswered questions about certain phenomena we observe around us, including the nature of *Dark Matter* [28–30], the *hierarchy problem* [31–33], the strong CP problem [34, 35], and other problems with the Standard Model. The success of the Standard Model in describing most of the phenomena we observe, and in predicting the outcomes of experiments, lends confidence to the idea that these unexplained phenomena may be explained by particles and fields. Some proposed new theories include *Supersymmetry* [36–43], *axion* models [35, 44–48], and other more exotic theories [49–54], including exotic decays of the Higgs boson [55].

The Standard Model and these new theories are tested using the *ATLAS* experiment [56] to detect the results of proton-proton collisions in the *Large Hadron Collider* [57–60], or *LHC*, located at *CERN* near Geneva, Switzerland. The Standard Model describes the interactions of protons and their constituent *quarks* and

gluons via *Quantum Chromodynamics*, or *QCD* [61–64]. The products of these interactions very often involve the phenomenon of *jets* [64–67] as collimated showers of particles originating from produced quarks and gluons. As ubiquitous objects in a proton-proton collider like the LHC, jets are essential for testing Standard Model predictions [68, 69], including properties of the Higgs boson [70], and for searching for evidence of new physics [69, 71–73].

Jets are complex objects and there are many different aspects of their reconstruction. The jet showers are often *groomed* to account for the effects of other radiative particles in the event and multiple simultaneous or out-of-time collisions (*pile-up*) [74–79]. The kinematic observables associated to jets need to be calibrated to account for the detector response, the effect of pile-up, the provenance of the jet, and differences between simulation and data [80–85]. The identity of the originating particle can also be determined using a variety of algorithms based on the properties of the shower in the detector [69, 78, 86–95].

This Thesis presents original research in efforts to search for new physics beyond the Standard Model using jets observed in the ATLAS detector, including work on improving the reconstruction of jets intended to improve the sensitivity of these searches. The work presented here addresses some of the key questions in particle physics today. By searching for new physics, it is possible to shed light on the nature of the Higgs boson and the possibility of physics beyond the Standard Model. These searches target processes involving multiple jets in the final state, which motivates innovations in the reconstruction of jets. This Thesis focuses particularly on the reconstruction of jet energies, which is essential for reconstructing the final states used in these searches and in many other searches in ATLAS. In addition to setting new bounds on well-motivated theoretical models, the innovations in object reconstruction and analysis techniques developed in this Thesis can be applied in other ATLAS efforts using currently available data or data gathered in the future.

Chapter 2 gives an overview of the Standard Model and goes into detail about some relevant aspects. Chapter 3 is devoted to the understanding of what exactly are jets and how they are used. In Chapter 4, the LHC is described, and in Chapter 5, the ATLAS experiment is detailed.

The remaining chapters cover original work by the author using data gathered with the ATLAS experiment.

Two chapters are devoted to independent searches for new physics using jets in ATLAS. Chapter 6 covers a search for a beyond-the-Standard-Model decay of a Higgs boson. In Chapter 7, a generic data-driven resonance search using machine learning is described.

Two following chapters discuss work to improve the reconstruction of the energy of jets from the measurements in the ATLAS calorimeter and tracking systems. Chapter 8 describes efforts to formalize the process of calibration of jets observed in the ATLAS detector, and Chapter 9 discusses a new method to further improve this calibration with machine learning.

Finally, Chapter 10 recounts the described efforts, puts them into context, and discusses future developments.

A few Appendices add to the body of work presented in this Thesis, mostly to supplement the Chapters in the main body of the text.

1.1 Units

Natural units are used, with $c = \hbar = 1$. The unit of energy to be used most often is the GeV, which is also the unit of mass and of momentum in natural units.

The metric signature is $(+,-,-,-)$, which is the convention commonly used in particle physics.

1.2 Coordinates

As described in Chapter 5, ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the center of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The *pseudorapidity* is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. The *rapidity* is defined

for an object with observed energy E and momentum p as $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$; in the limit that the object's mass m goes to zero, the rapidity and pseudorapidity are equal. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

The four-momentum of a particle with energy E and vectorial momentum $\vec{p}$ is $p^\mu = (E, \vec{p})$, with $m = \sqrt{p^2} = \sqrt{E^2 - |\vec{p}|^2}$ the mass of the particle. The four-momentum is entirely determined by (m, p_T, η, ϕ), where $p_T = \sqrt{p_x^2 + p_y^2}$ is the *transverse momentum*:

$$E = \sqrt{m^2 + p_T^2 \cosh^2 \eta} \tag{1.1}$$

$$p_x = p_T \cos\phi \tag{1.2}$$

$$p_y = p_T \sin\phi \tag{1.3}$$

$$p_z = p_T \sinh\eta \tag{1.4}$$

Sometimes the *transverse energy* $E_T = \sqrt{m^2 + p_T^2}$ will be specified instead of p_T.

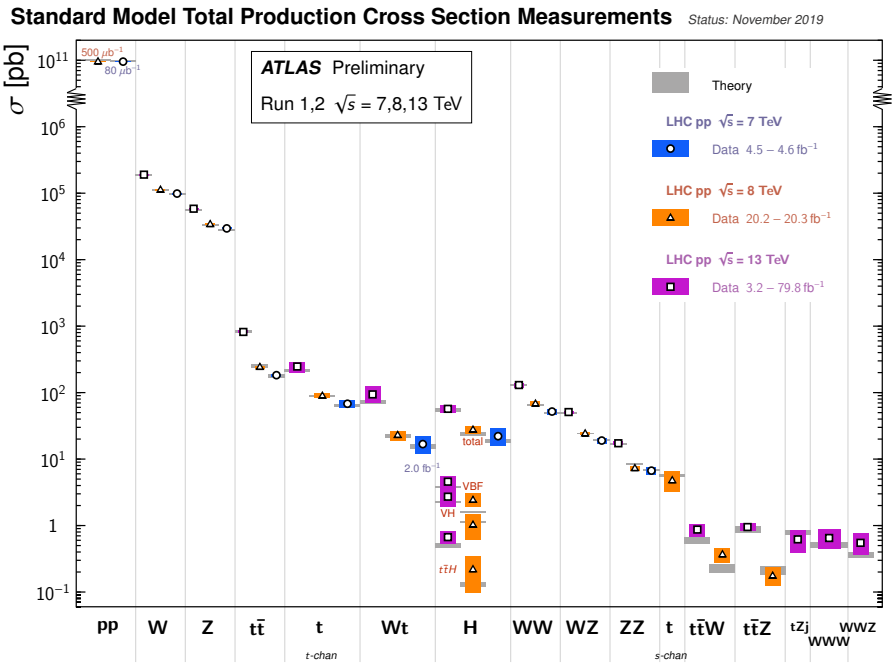
Chapter 2

The Standard Model of Particle Physics

2.1 Introduction

The *Standard Model of Particle Physics*, or *SM*, is the overarching theory that describes physics at the most fundamental level. In the SM, *particles* are localizations of *quantum fields*, and these particle-fields account for both matter and all known forces or interactions other than gravity. The SM is a remarkably well-measured and well-tested theory. Figure 2.1 shows the cross sections of a variety of physics processes measured with the ATLAS detector (Chapter 5) and their theoretical predictions using the SM. Every single measurement, spanning many orders of magnitude, agrees with the theoretical prediction within the uncertainties.

The SM describes three fundamental forces, each with a corresponding gauge symmetry group. The electromagnetic and weak forces are combined into a single *electroweak* interaction under SU(2)×U(1) (Section 2.2). The SU(2) symmetry is spontaneously broken by the Higgs boson, giving rise to massive bosons and to the mass of quarks. The strong force is described by *quantum chromodynamics*, with symmetry group SU(3) (Section 2.3). The strong force is unique in that, due to the strengthening of the coupling constant at low energy scales, quarks *confine* into multi-quark *hadrons* like protons and neutrons. However, the strong force gets weaker at



(a)

Figure 2.1: Summary of several cross section measurements at ATLAS, compared to their theoretical predictions using the SM. Figure sourced from [96].

higher energy scales, a phenomenon called *asymptotic freedom*. The SM is summarized in Section 2.4. Despite the successes of the SM, there are still unexplained phenomena like dark matter and other curious features (which, if left unexplained, are at least unsatisfying) like the quark mass hierarchy (Section 2.5). There are many theories *beyond the SM*, or *BSM*, which attempt to answer these questions, many of which predict new particles that may be produced at the LHC (Chapter 4). The two searches presented in this Thesis (Chapter 6 and Chapter 7) are both looking for evidence of any BSM particles. The analysis in Chapter 6 is searching for BSM decays of the Higgs boson, while the analysis in Chapter 7 is searching for generic new massive particles decaying hadronically.

This Chapter serves as a brief overview of the SM on a technical level; there are a variety of textbooks and other sources which provide much more detailed and complete information. Some textbooks in particular are [97–100], which are the textbooks the Author used to learn about quantum field theory and the SM as a student. In addition to serving as the primary sources for most of this Chapter, these textbooks were the catalysts which (in part) spurred the interest in Particle Physics ultimately resulting in the work presented in this Thesis.

2.2 Electroweak Sector

The Standard Model is described by *Yang-Mills* or *non-Abelian* gauge theories [101, 102], with interactions mediated by vector bosons between fermionic matter particles with spin $\frac{1}{2}$. The electromagnetic and weak forces are unified using the *Glashow-Weinberg-Salam* model [103–106] for electroweak unification, with gauge symmetry SU(2)×U(1). The gauge bosons corresponding to these symmetries are W^a_μ, with a running over the 3 generators of SU(2), and B_μ, respectively. The U(1) symmetry does not correspond to the familiar electromagnetic charge, but rather corresponds to a *hypercharge* Y; we will see that the photon corresponds to a linear combination of B_μ and one of the W^a_μ. There is in addition a complex doublet H with hypercharge

$\frac{1}{2}$ called the *Higgs multiplet* [21–24, 27][1]. Altogether, the kinetic term for these fields is

$$\begin{aligned}\mathcal{L}_{\text{kin}} &= -\frac{1}{4}(W^a_{\mu\nu})^2 - \frac{1}{4}B^2_{\mu\nu} \\ &\quad + (D_\mu H)^\dagger (D_\mu H) + m^2 H^\dagger H - \lambda (H^\dagger H)^2, \qquad (2.1)\end{aligned}$$

$$D_\mu H = \partial_\mu H - igW^a_\mu \tau^a H - \frac{1}{2} ig' B_\mu H, \qquad (2.2)$$

where $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ is the field strength; $W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g f^{abc} W^b_\mu W^c_\nu$ is the field strength of the W, and $f^{abc} = \epsilon^{abc} \neq 0$ the totally anti-symmetric tensor are the structure constants of SU(2); m and λ are free parameters of the theory; D_μ is the covariant derivative, with corresponding couplings g, g' which are free parameters of the theory; $\tau^a = \frac{1}{2}\sigma^a$ are the canonically normalized generators of SU(2), with σ^a the Pauli matrices; and the factor of $\frac{1}{2}$ in front of $B_\mu H$ is due to the hypercharge $\frac{1}{2}$ of the H.

The gauge bosons are all massless, and in fact terms corresponding to the mass of these fields (e.g., $B^\mu B_\mu$) would violate gauge invariance. These bosons (or rather, linear combinations of them) gain mass by the SU(2) symmetry being spontaneously broken by the Higgs. The Higgs potential term, $V(H) = m^2|H|^2 - \lambda|H|^4$, has a minimum away from $H = 0$, inducing a vacuum expectation value (VEV). Without loss of generality we let the VEV be real, $v = \frac{m}{\sqrt{\lambda}}$, and in the lower component, so that expanding about the minimum we have

$$H = \exp\left(2i\frac{\pi^a \tau^a}{v}\right) \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} + \frac{h}{\sqrt{2}} \end{pmatrix}, \qquad (2.3)$$

where we have introduced a new scalar field h which corresponds to motion about the minimum at v; and the π^a are the explicit SU(2) gauges.

We choose $\pi^a = 0$ to simplify the calculations. Then expanding the covariant

[1] Before data-taking at the LHC, the Higgs field was only theoretical. In 2012 the existence of this particle was confirmed [25, 26].

derivative terms, we get:

$$|D_\mu H|^2 = g^2 \frac{v^2}{8} \left[(W_\mu^1)^2 + (W_\mu^2)^2 + \left(\frac{g'}{g} B_\mu - W_\mu^3 \right)^2 \right] + \text{terms involving } h \tag{2.4}$$

The terms not involving h correspond to mass terms for the gauge bosons, while the terms that do involve h correspond to interactions between h and the gauge bosons and itself.

We identify a massless and a massive boson with

$$Z_\mu = \cos\theta_w W_\mu^3 - \sin\theta_w B_\mu, \tag{2.5}$$

$$A_\mu = \sin\theta_w W_\mu^3 + \cos\theta_w B_\mu, \tag{2.6}$$

with $\tan\theta_w = \frac{g'}{g}$; θ_w is called the Weinberg angle. A_μ corresponds to the familiar photon, with mass 0, and Z_μ is the Z boson, which has mass $m_Z = \frac{1}{2\cos\theta_w} g v$.

The coupling to A and therefore the normal electromagnetic charge of the gauge bosons is determined by $g[A_\mu, W_\mu^a \tau^a] = g \sin\theta_w W^3 W^a [\tau^3, \tau^a]$, with the electromagnetic coupling strength

$$e = g \sin\theta_w = g' \cos\theta_w \tag{2.7}$$

Clearly, with $a = 3$, the charge is 0 - the Z boson is neutral. W_1 and W_2 do not have definite charges, so we define instead $\tau^\pm = \frac{1}{\sqrt{2}} \left(\tau^1 \pm i\tau^2 \right)$, so that $[\tau^3, \tau^\pm] = \pm\tau^\pm$, and the (linear combination of the) W bosons coupling to $\tau^\pm$ have charge $\pm$, which turn out to be $W^\pm = \frac{1}{\sqrt{2}} \left(W^1 \mp i W^2 \right)$. In addition to $W\pm$ having electric charge $\pm$, their mass is $m_W = \frac{gv}{2}$. In particular, one immediate prediction is that $m_W < m_Z$.

The Higgs boson itself is a scalar, and uncharged under either hypercharge or SU(2). However, it does interact via 3- and 4-point interactions with the $W^\pm$ and Z bosons; the search in Chapter 6 targets Higgs bosons produced via this interaction in the vector-boson-fusion mode. There are also 3- and 4-point self-interactions of the Higgs. The mass of the Higgs is $m_H = \sqrt{2} m$, which again is a free parameter of the

theory.

The four original free parameters m, λ, g, g' are therefore related to experimentally observable quantities [107]: $\alpha_e = \frac{e^2}{4\pi} \approx \frac{1}{137}$ is the fine structure constant; $m_Z \approx 91.2$ GeV; $m_W \approx 80.4$ GeV; and $m_N \approx 125$ GeV. In particular, $\sin^2\theta_w \approx 0.22$ and $v \approx 246$ GeV.

2.2.1 Fermions

The coupling of the weak interaction to fermions is *chiral*, meaning the interactions are different between left- and right-hand components of the fermion fields. In fact, the weak interaction couples only to left-handed fermions, and the SU(2) symmetry is often written as $SU(2)_L$ because of this. There are 3 generations each of the *leptons* and the *quarks*, whose left-handed components are doublets under SU(2):

$$L^i = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \tag{2.8}$$

$$Q^i = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}, \tag{2.9}$$

where i goes over the 3 generations; and their corresponding right-handed components, which are singlets:

$$e_R^i = e_R, \mu_R, \tau_R \tag{2.10}$$

$$\nu_R^i = \nu_{eR}, \nu_{\mu R}, \nu_{\tau R} \tag{2.11}$$

$$u_R^i = u_R, c_R, t_R \tag{2.12}$$

$$d_R^i = d_R, s_R, b_R \tag{2.13}$$

The fields are appropriately named to correspond to the familiar leptons and quarks - the electron e, the muon μ, and the tau particle τ; the neutrinos; the up type quarks up, charm, and top; and the down type quarks down, strange, and bottom. It should be noted that the right-handed neutrinos, which are not charged

Table 2.1: Charges of Standard Model fields under U(1) hypercharge, SU(2), and SU(3). - means the field does not transform under the symmetry group, i.e. a singlet. **2, 3** mean the field transforms in the fundamental representation of the group, i.e. a doublet for SU(2) and a triplet for SU(3), respectively.

Field	$L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	e_R	ν_R	$Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	u_R	d_R	H
SU(3)	-	-	-	**3**	**3**	**3**	-
SU(2)	**2**	-	-	**2**	-	-	**2**
U(1)	$-\frac{1}{2}$	-1	0	$\frac{1}{6}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{2}$

under U(1) and are singlets under SU(2), do not interact via the electroweak force; they are also color neutral (Section 2.3), so they are called *sterile neutrinos*.

Furthermore, the quark fields listed above are in the interaction eigenstate basis, which is different from the mass eigenstate basis, as will be described below.

The charges of these fields under hypercharge and SU(2) are given in Table 2.1 (as well as the charge under SU(3) (Section 2.3).

The interaction terms between the fermions and the gauge bosons are:

$$\begin{aligned}
\mathcal{L}_{\text{int}} = {} & i\bar{L}_i \left(\not{\partial} - ig\not{W}^a\tau^a - ig'Y_L\not{B}\right) L_i \\
& + i\bar{Q}_i \left(\not{\partial} - ig\not{W}^a\tau^a - ig'Y_Q\not{B}\right) Q_i \\
& + i\bar{e}^i_R \left(\not{\partial} - ig'Y_e\not{B}\right) e^i_R \\
& + i\bar{\nu}^i_R \left(\not{\partial} - ig'Y_\nu\not{B}\right) \nu^i_R \\
& + i\bar{u}^i_R \left(\not{\partial} - ig'Y_u\not{B}\right) u^i_R \\
& + i\bar{d}^i_R \left(\not{\partial} - ig'Y_d\not{B}\right) d^i_R,
\end{aligned} \tag{2.14}$$

where, e.g., $\not{A} = \gamma^\mu A_\mu$, with γ^μ the Dirac γ matrices; Y are the hypercharges corresponding to the given set of fields (Table 2.1); and there are implicit projection operators $P_L = \frac{1}{2}(1 - \gamma_5)$, $P_R = \frac{1}{2}(1 + \gamma_5)$.

The familiar elementary charge corresponds to the coefficient of coupling of the fields with A_μ. Expanding out in terms of W and B, we have (for the electron and

neutrino fields, and the same, *mutatis mutandis*, for the up and down quark fields):

$$\begin{aligned}\mathcal{L}_{\text{int}} = &\bar{e}^i_L \left(-\frac{1}{2} g \not{W}^3 + g' Y_L \not{B}\right) e^i_L \\ &+ \bar{\nu}^i_L \left(\frac{1}{2} g \not{W}^3 + g' Y_L \not{B}\right) \nu^i_L \\ &+ g' Y_e \bar{e}^i_R \not{B} e^i_R + g' Y_\nu \bar{\nu}^i_R \not{B} \nu^i_R \\ &+ \text{off-diagonal terms}\end{aligned} \tag{2.15}$$

The off-diagonal terms contain terms like, e.g., $\bar{e}^i_L g(\not{W}^1 - i\not{W}^2)\nu_L = \bar{e}^i_L g \not{W}^+ \nu_L$. These are the flavor-changing charged current interactions with the W boson - e.g., e to ν_e or c to s.

Changing basis to the A and Z bosons, we then have:

$$\begin{aligned}\mathcal{L}_{\text{int}} = &\, e\left(\left(-\frac{1}{2} + Y_L\right)\bar{e}^i_L \not{A} e^i_L + \left(\frac{1}{2} + Y_L\right)\bar{\nu}^i_L \not{A} \nu^i_L + Y_e \bar{e}^i_R \not{A} e^i_R + Y_\nu \bar{\nu}^i_R \not{A} \nu^i_R\right) \\ &+ Z \text{ terms}\end{aligned} \tag{2.16}$$

Just to be clear, $e = g \sin\theta_w$ is a constant. Reading off Table 2.1, we therefore have that the electromagnetic charges of the left- and right-handed charged leptons e are -1 and that the electromagnetic charges of the left- and right-handed neutrinos ν are 0. Examining the same terms involving the left- and right-handed u and d fields yields electromagnetic charges of $+\frac{2}{3}$ for the up-type quarks u and $-\frac{1}{3}$ for the down-type quarks d.

As with the gauge bosons, mass terms like $\bar{e}_L e_R$ break the gauge invariance, and mass terms arise due to couplings with the Higgs boson. These terms are called *Yukawa couplings*, e.g. for the electron

$$\mathcal{L}_{\text{mass}} = -y_e \bar{L} H e_R + \text{h.c.} \tag{2.17}$$

will generate a mass term $-m_e(\bar{e}_L e_R + \bar{e}_R e_L)$ with $m_e = \frac{y_e}{\sqrt{2}} v$ after the symmetry is spontaneously broken and the Higgs boson gets a VEV. These terms only give mass to the charged leptons and the down-type quarks. For the up quarks, we use a slightly

different term for gauge invariance:

$$\mathcal{L}_{\text{mass}} = -Y^d_{ij}\bar{Q}^i H d^j_R - iY^u_{ij}\bar{Q}^i\sigma_2 H^* u^j_R + \text{h.c.} \tag{2.18}$$

where i, j run over the quark generations, and Y^d and Y^u are general mixing matrices between the generations. After symmetry breaking we then have

$$\mathcal{L}_{\text{mass}} = -\frac{v}{\sqrt{2}}\left(Y^d_{ij}\bar{d}^i_L d^j_R + Y^u_{ij}\bar{u}^i_L u^j_R\right) + \text{h.c.} \tag{2.19}$$

We diagonalize Y^d and Y^u separately,

$$Y^d = U_d M_d K_d^\dagger \tag{2.20}$$

$$Y^u = U_u M_u K_u^\dagger \tag{2.21}$$

with U and K unitary matrices and M diagonal matrices with real positive eigenvalues. With the transformations $u_L \to U_u u_L$, $d_L \to U_d d_L$, $u_R \to K_u u_R$, $d_R \to K_d d_R$ the U and K terms disappear, and we have

$$\mathcal{L}_{\text{mass}} = -m^d_j \bar{d}^j_L d^j_R - -m^u_j \bar{u}^j_L u^j_R + \text{h.c.} \tag{2.22}$$

where m^d_j and m^u_j are the quark masses - the diagonal elements of $\frac{v}{\sqrt{s}}M^d$ and $\frac{v}{\sqrt{s}}M^u$, respectively.

In diagonalizing, the U matrices come up only in the interactions with $W\pm$, via terms like

$$W^+_\mu \bar{u}^i_L \gamma^\mu \left(U^\dagger_u U_d\right)^{ij} d^j_L + W^-_\mu \bar{d}^i_L \gamma^\mu \left(U^\dagger_d U_u\right)^{ij} u^j_L \tag{2.23}$$

So mixing occurs between the quark generations via interactions with the $W^\pm$ bosons subject to the matrix $V = \left(U^\dagger_u U_d\right)$, known as the *Cabibbo-Kobayashi-Maskawa*, or *CKM* matrix. There are 3 free angles corresponding to mixing in i, j space, and 1

Table 2.2: Lepton and quark masses [107].

Quark	d	s	b	u	c	t
Mass [MeV]	4.7	96	4.18×10^3	2.2	1.28×10^3	173.1×10^3
Lepton	e	μ	τ	ν_e	ν_e	ν_τ
Mass [MeV]	0.511	105.7	1777	$< \mathcal{O}(10^{-6})$	$< \mathcal{O}(10^{-6})$	$< \mathcal{O}(10^{-6})$

free complex phase corresponding to CP violation:

$$V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{pmatrix} \times \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13}e^{i\delta} \\ 0 & 1 & 0 \\ -\sin\theta_{13}e^{i\delta} & 0 & \cos\theta_{13} \end{pmatrix} \times \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.24}$$

The empirical values are [107] $\theta_{12} = 13.04 \pm 0.05°$, $\theta_{13} = 0.201 \pm 0.011°$, $\theta_{23} = 2.38 \pm 0.06°$, and $\delta = 68.8 \pm 4.6°$. In particular, the CKM matrix is mostly diagonal, since all the θ angles are small; the largest is θ_{12} (mixing between the first and second generation).

The leptons also have Yukawa couplings, generating their masses. However, the neutrinos are massless[2], allowing a simultaneous diagonalization of Y^ν and Y^e. This implies any mixing between the lepton generations is not allowed, and lepton number for each lepton generation is conserved under the weak interaction.

As mentioned above, while Yukawa interactions explain how lepton and quark masses are possible, the actual values are free parameters of the theory. The experimental values are summarized in Table 2.2. Both the quark masses and the charged lepton masses span many orders of magnitude, raising a potential hierarchy problem (more on this in Section 2.5).

[2] Actually, neutrinos are known to have very small but positive masses. More on this in Section 2.5.

2.3 Quantum Chromodynamics

The strong force is described by a symmetry group under SU(3). The 3 dimensions are referred to as *color charges* (red, green, and blue when names are required[3]), and so the strong force is referred to as *Quantum Chromodynamics*, or *QCD* [61–64]. The gauge boson associated with the theory is the gluon, and of the fermionic fields only the quarks are charged under this interaction. The quarks (here denoted by ψ_f, labeled by the quark flavor f) transform under the fundamental representation of SU(3), meaning each of the 6 flavors are length 3 column vectors, or triplets. The gluons G^a_μ transform in the adjoint representation as 3×3 matrices, with 8 color-anticolor states (the *octet*) corresponding to the 8 Gell-Mann matrices.

The Lagrangian of QCD is:

$$\mathcal{L}_{QCD} = i\bar{\psi}_f \not{D}_\mu \psi_f - \frac{1}{4}(G^a_{\mu\nu})^2 \tag{2.25}$$

where $\not{D} = \gamma^\mu D_\mu$; $D_\mu = \partial_\mu - ig_s G^a_\mu T^a$ is the covariant derivative; $T^a = \frac{\lambda^a}{2}$ are the generators of SU(3) (λ^a are the Gell-Mann matrices); $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu$, with f_{abc} the structure constants of SU(3), which are totally anti-symmetric.

We see immediately that the set of fundamental interactions are: 2 quarks + 1 gluon, either with a color exchange or no color exchange; 3 gluons; and 4 gluons.

In contrast to the electroweak theory (Section 2.2), there is no Higgs boson to break the symmetry, and so in some sense the model is much simpler. However, phenomenologically and empirically QCD is much more complicated, due to the manner in which the coupling g_s runs as a function of energy scale.

In every theory scattering cross sections and other observables can be calculated perturbatively in the interaction constant. Formally some of these terms, which correspond to loops in Feynman diagrams, are infinite; in order to account for these infinities calculations can instead be carried out by comparing calculations between two finite scales, so that the infinities cancel, a technique called *renormalization*. One

[3] Totally unrelated to colors in the visual spectrum that we are used to. However, the analogy does work a little, as a state with all three color charges is has 0 net color charge, i.e. "white" or colorless.

consequence of renormalization (which is necessary to make finite predictions) is that the effective strength of the coupling g in the theory changes, or *runs* depending on the scale being probed. This running of the coupling is encoded in the *beta function*:

$$\frac{\partial}{\partial \ln \mu} g = \mu \frac{\partial}{\partial \mu} g = \beta(g), \tag{2.26}$$

where μ is the energy scale being probed.

In QCD[4] $\beta(g) = -\frac{g^3}{(4\pi)^2}\beta_0$, with $\beta_0 = (\frac{11}{3}n_c - \frac{2}{3}n_f)$; $n_c = 3$ is the number of colors, and $n_f = 6$ is the number of flavors, so $\beta_0 = 7 > 0$. This implies that in QCD the coupling strength gets weaker at higher energies, and stronger at lower energies, a phenomenon known as *asymptotic freedom*[5] [66, 109, 110]. Said another way, at low energy scales g_s grows so large that perturbative calculations are impossible and the theory breaks down; at high energy scales g_s is small and perturbative calculations can be performed. The running of the coupling (expressed in terms of $\alpha_s = \frac{g_s^2}{4\pi}$) can be seen in Figure 2.2. It can be seen that the coupling gets too strong and the theory becomes non-perturbative around $\mathcal{O}(\text{GeV})$.

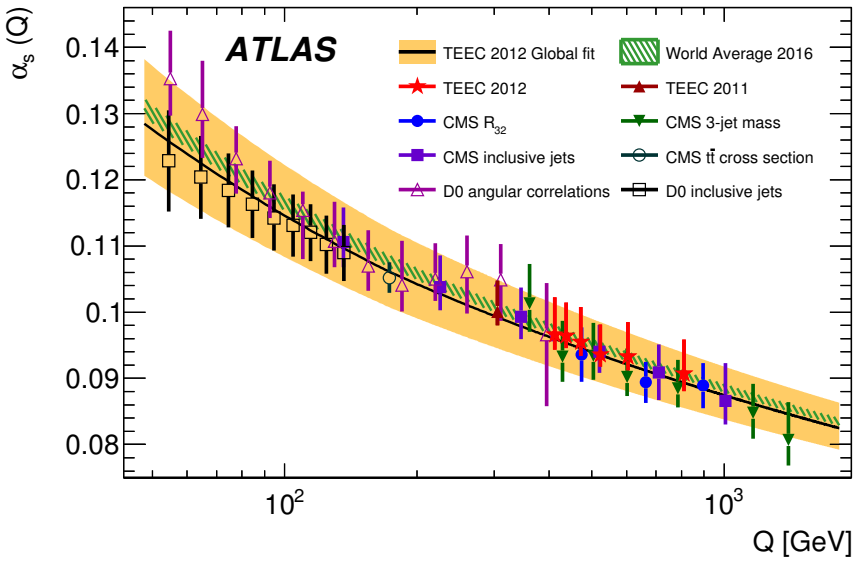
Because of the increasing strength of the strong force at low energies, quarks and gluons are *confined* [112] into color-neutral composite particles, or *hadrons*. In order to pull apart two quarks that compose a hadron, the larger the distance between them the greater the strength of the interaction, and it becomes energy favorable to pull quark-antiquark pairs out of the vacuum to create new hadrons, a process called *hadronization*.

A lone quark or gluon produced in a hard-scatter process will first *fragment*, radiating gluons which in turn convert into quark-antiquark pairs, etc., and then hadronize once low enough energy scales are reached. This collimated hadronic shower leads to *jets* [64–67] observed in the detector (Chapter 3).

Hadrons composed of 2 quarks are called *mesons*, and those composed of 3 quarks

[4] Like everything else, β is calculated perturbatively in g. Given here is the one-loop contribution.

[5] The couplings for U(1) and SU(2) also run. The U(1) coupling has a negative β function, so it gets stronger at higher energy scales, but remains perturbative even at the Planck scale, where it is expected that a higher-energy theory will become relevant. The SU(2) coupling has a positive β function, so it gets stronger at lower energy scales (larger distances), but since the W/Z bosons are massive the weak force is short-range and these energy scales are not relevant for the theory [108].



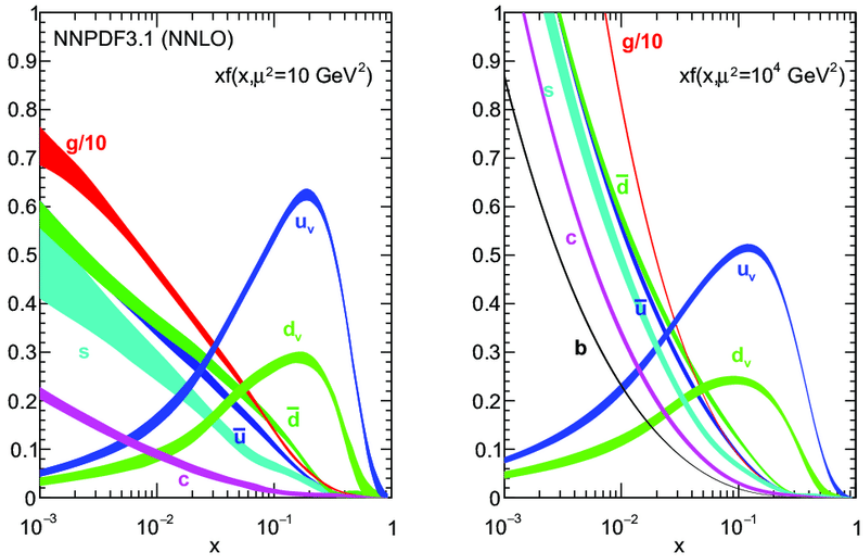
(a)

Figure 2.2: Strong coupling α_s as a function of energy scale Q, theoretical predictions and experimental measurements. Figure sourced from [111].

are called *baryons*; hadrons with four or more constituent quarks (*tetraquarks*, *pentaquarks*) have been observed [113, 114] but are very rare. The familiar protons are baryons composed of two up quarks and one down quark.

When protons collide in the LHC (Chapter 4), the energy scales are high enough such that the individual constituent quarks and gluons do interact rather than the entire hadron [115]. The QCD factorization theorem allows the cross section calculation to be factorized between the hard-scatter and perturbative $2 \to N$ process and the non-perturbative hadronic structure. However, even with this simplification, the interaction between the constituent partons has to be averaged over the distribution of momenta each could be carrying, at the energy scale of the interaction. This information is contained in the *parton distribution function* [116, 117], or *PDF*, $f_q(x, \mu^2)$, which is the probability distribution function for a parton q to be carrying momentum fraction x at energy scale μ. The hard-scatter interactions can either occur via the *valence partons*, which are the constituent up and down quarks, or via *sea partons*, which are virtual quarks of any flavor or gluons in the interior of the proton. A

calculation of the PDFs at two energy scales can be seen in Figure 2.3. At high x, the valence quarks dominate, but at lower x the sea partons, especially the gluons, become more prominent.



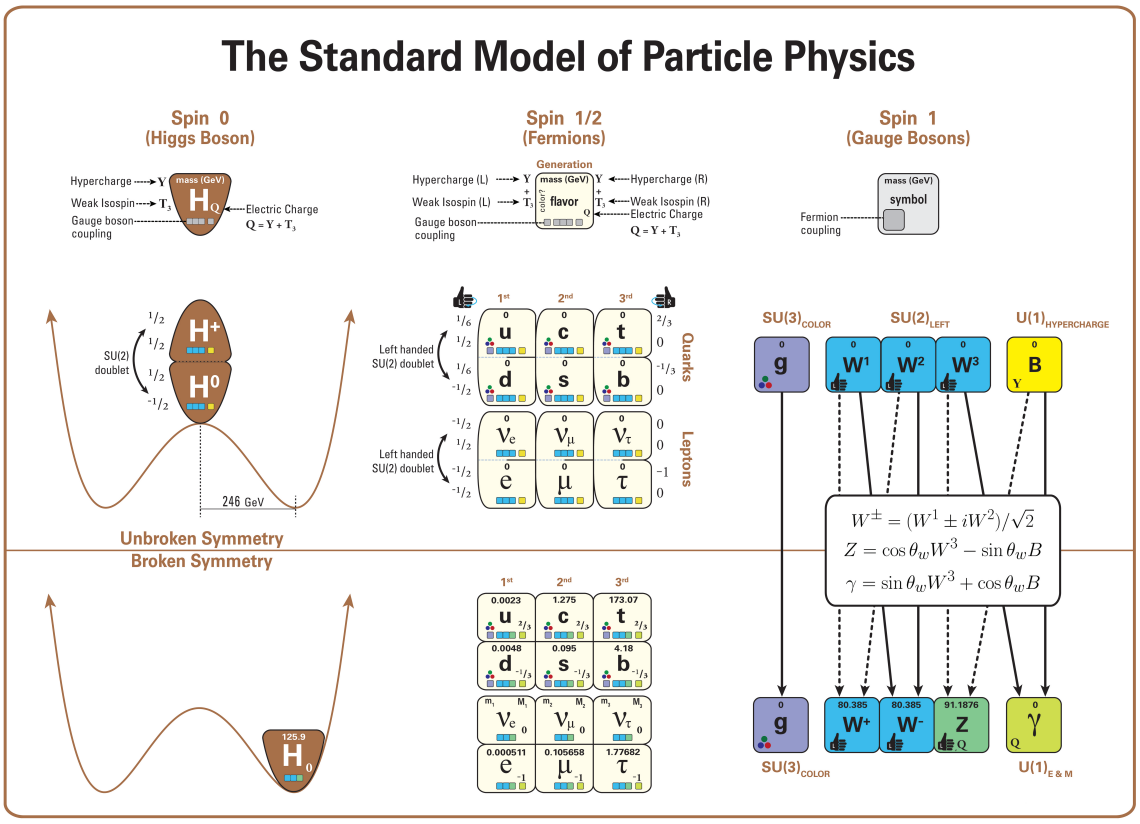
(a)

Figure 2.3: Proton PDFs at (left) $\mu^2 = 10$ GeV^2 and (right) $\mu^2 = 10^4$ GeV^2 calculated with NNPDF3.1 at next-to-next-to leading order (NNLO). Figure sourced from [118].

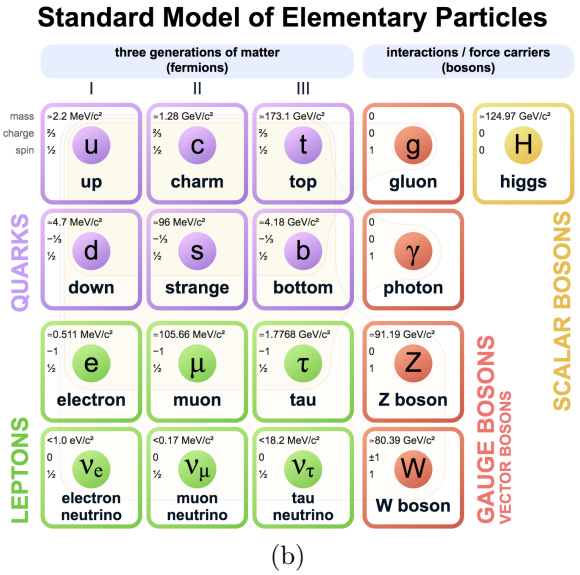
2.4 Summary

A summary of the SM can be seen in Figure 2.4, before and after spontaneous symmetry breaking of SU(2).

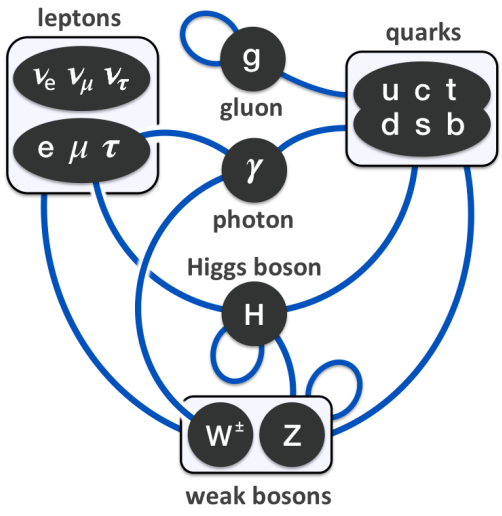
Among the spin $\frac{1}{2}$ fermions, there are 3 generations each of quarks and leptons, split into up-type and down-type. The 6 quarks are the (up-type) up, charm, top, and (down-type) down, strange, bottom; the 6 leptons are the (charged) electron, muon, and tau, and their respective neutral neutrino partners. There are also 4 vector gauge bosons - the photon, with mediates the electromagnetic force; the $W^{\pm}/Z$ bosons, which mediate the weak interaction; and the gluon, which mediates the strong force.



(a)



(b)



(c)

Figure 2.4: Summary of SM. (a) Fields before and after electroweak symmetry breaking. (b) After electroweak symmetry breaking - particle content. (b) After electroweak symmetry breaking - interactions. Figures sourced from [119].

Finally there is the Higgs boson, which spontaneously breaks the SU(2) electroweak symmetry and acquires a vacuum expectation value, which has spin 0.

The up-type quarks each have electric charge $+\frac{2}{3}$; the down-type quarks each have charge $-\frac{1}{3}$; the charged leptons each have charge -1; enabling interactions with the photon; the neutrinos each have charge 0 and do not interact via the electromagnetic force. Each of these fermions can interact with the W and Z bosons; the quarks can have flavor-changing interactions according to the CKM matrix; the leptons do not mix generations via the weak force. Only the quarks are colored and interact with the gluon via the strong force.

The phenomenology of the strong force implies color confinement, which leads to hadrons like the proton and neutron, which are the constituents of the nuclei of atoms, and the rest of the particle zoo like pions, kaons, etc.

The Higgs boson gives mass via Yukawa couplings to the charged leptons and the quarks. It also gives mass to the $W^{\pm}$ and Z bosons. The $W^{\pm}$ bosons are charged under electromagnetism and therefore interact with the photon. Finally, each of the gauge bosons other than the photon has self-interactions[6].

2.5 Beyond the Standard Model

Despite the successes of the SM, there are still underexplained or unexplained natural phenomena. "Underexplained" phenomena refer to features within the SM that are either aesthetically displeasing (also called "unnatural"), or where there is a discrepancy between the SM prediction and the observed value. "Unexplained" phenomena refer to observations of effects that have no corresponding elements in the SM outlined above.

As the SM attempts to describe physics on the most fundamental level, any unsatisfying or missing pieces indicate that the theory is not complete, implying there must exist a theory of new physics, or *beyond the SM* (*BSM*).

Some of these problems with the SM are outlined below, in (subjective) order of increasing weakness of the SM. There are many BSM models [49–54], and some of

[6] At tree level. E.g., the photon does have self-interactions involving fermion loops.

these models account for some of the problems with the SM listed below. Some of the more popular models are included with the problem they account for.

As mentioned in Section 2.3, below a certain energy scale QCD becomes non-perturbative, and the theory therefore becomes intractable analytically. One alternative formulation is lattice QCD [120], which analyzes QCD on a lattice in space-time. This formulation has had success, for example accurately predicting the hadron masses from first principles [121]. However, numerical lattice QCD calculations using Monte Carlo methods can be extremely computationally intensive, and often the only available descriptions rely on phenomenological models, e.g. to describe the hadron PDFs. While not necessarily a failure of the theory itself, there may exist a different formulation of the SM which would be able to calculate quantities and make predictions about hadrons analytically from a fundamental level.

As mentioned in Section 2.1, the SM makes predictions that can be calculated very precisely, and these predictions have been remarkably well-tested. There are only a couple of precision measurements that are in tension with the SM predictions, of the tens or hundreds of precision SM tests. The muon anomalous magnetic moment [122], or muon $g-2$, has been calculated [123] up to 10th order, including weak and hadronic effects in addition to the primary quantum electromagnetic effects, yielding a precision of 10 decimal places. This quantity has also been measured [124] to an uncertainty of 0.54 parts per million. There is a tension of almost 3σ (with σ equal to the experimental uncertainty[7]) between the experimental and theoretical values, which may be an indication of new physics or may be a statistical fluctuation. A new experiment [125] at Fermilab will reduce the experimental uncertainty by a factor of around 4, which can elucidate whether this is a real tension or not.

There is also evidence from LHCb [126] of tension with the SM in B-hadron decays [127], at the level of $3-4\sigma$. This also remains to be seen whether it is a true tension indicative of new physics or a statistical fluctuation (or unaccounted-for experimental bias).

As mentioned in Section 2.2, the quark and charged lepton masses are allowed

[7]The theoretical uncertainty is much smaller than the experimental uncertainty, and so the uncertainty on the difference is dominated by the experimental measurement.

in the SM via Yukawa couplings with the Higgs boson. However, there is a large *hierarchy*, meaning many orders of magnitude difference in free parameters of the model. This is especially apparent for the quarks, where the top quark mass and the up quark mass differ by a factor of 10^5. This indicates that there may be some theory at higher energies that explains this observed difference [128].

The hierarchy problem gets drastically worse when considering the neutrino masses. As discussed in Section 2.2, in the vanilla SM neutrinos are massless, and cosmological constraints [129] place limits on the sum of the neutrino masses at < 1 eV. However, observations of neutrinos oscillating between flavors [130] indicate that neutrinos must have finite masses, which is consistent with cosmological models [131]. The neutrino oscillation observations measure only differences between neutrino masses of different generations, with measurements of [107, 132, 133] $\Delta m^2_{12} = 7.5 \times 10^{-5}$ eV^2 and $\Delta m^2_{23} = 2.5 \times 10^{-3}$ eV^2, where the subscripts label the neutrinos in the mass basis (which is different from the interaction basis, similar to quarks). These are only mass differences, but the implication is that there must necessarily exist some neutrino with mass > 0.05 eV.

The neutrino masses can be generated via Yukawa couplings in the same manner as the charged leptons and the quarks, which is called *Dirac masses*. However, there is an even larger hierarchy between the neutrinos and the lepton masses (a factor of 10^9 between the tau and the tau neutrino), indicating these masses may come from a different mechanism [134]. For example, the neutrinoless double beta decay process (two neutrons decaying to two protons with no associated missing energy corresponding to neutrinos), which has not yet been observed and would violate the SM [135], would imply that neutrinos are their own anti-particles, allowing a *Majorana mass* term in the electroweak Lagrangian.

All particles have quantum corrections to their mass due to loop terms involving other particles. While the quantum corrections to the fermions and gauge bosons are constrained by symmetry in the SM, the Higgs mass specifically is sensitive to physics at higher energy scales [31–33]. We know there must exist some new physics at the Planck scale ($m_P \sim 10^{19}$ GeV), incorporating Quantum Gravity (see below), so there is an enormous hierarchy between that scale and the Higgs mass, $m_H \sim 10^2$ GeV, of

a factor of 10^{17}.

It is therefore proposed that there could exist new physics around the TeV scale to protect and constrain the Higgs mass. A very popular theory is called *Supersymmetry*, or *SUSY* [36–43, 136–140]. In SUSY every particle has a supersymmetric partner, swapping fermions and bosons. The lightest of these supersymmetric particles would be on the TeV scale.

Another possibility is the existence of extra dimensions [141–148]. As of course we only observe the four dimensions of spacetime, any extra dimension would have to be rolled up or compactified into a small length. 5D particles viewed in 4D would be observed as towers of particles with equally spaced masses, some of which could be at the TeV scale.

Another problem with the SM comes in the form of the observed matter-antimatter asymmetry in the universe [149]. In the SM matter and antimatter are treated symmetrically, although differences between matter and antimatter can come in the form of CP violation [150], which is only present in the weak interaction. Up to the size of this violation, matter and antimatter particles should have been created in equal quantities in the Big Bang and the early universe. It's clear that our universe is almost entirely made up of matter rather than antimatter. However, the size of the CP violation in the weak interaction is not enough to explain this difference.

Puzzlingly, the QCD Lagrangian (Section 2.3) actually does allow for a CP-violating term [34] of the form

$$\mathcal{L}_{\text{CP-violating}} = \theta \frac{1}{32\pi^2} G^a_{\mu\nu} \epsilon_{\mu\nu\rho\sigma} G^{\rho\sigma a} \tag{2.27}$$

where $\epsilon_{\mu\nu\rho\sigma}$ is the four-dimensional totally antisymmetric tensor. However, measurements of the electric dipole moment of the neutron [151] set limits of at least $\theta < 10^{-9}$. In general in the SM if terms are allowed then they do exist, so the question, called the *strong CP problem*, is why this term is so small or 0.

A proposed solution is the presence of light pseudoscalar particles called *axions* [35, 44–48]. These particles would drive the θ parameter to 0, explaining the lack of strong CP violation.

A significant phenomenon underexplained by the SM is the existence of *dark matter* [28–30]. It has been known for a long time [152] that the amount of visible matter in galaxy clusters is not enough to keep the galaxies from escaping the cluster's pull. However, the idea of missing matter in galaxies was not generally accepted until observations of galaxy rotation curves [153–155] indicated the existence of additional matter other than what could be seen, i.e. dark matter. Since then, observations using weak lensing [156] and galaxy cluster merging [157] have confirmed the existence of dark matter. Cosmological models indicate [158] that dark matter comprises about 85% of all matter, the rest being the particles included in the SM.

There are of course electromagnetically neutral[8] stable particles in the SM - neutrinos. However, neutrinos are found to be too light to explain the abundance of observed dark matter [159][9]. Instead, the predominant theory is that of cold dark matter, composed of *weakly-interacting massive particles*, or *WIMPS* - massive particles that are electrically neutral but interact via the weak force. Either axions [160] or SUSY particles [161] could serve as WIMP candidates, which would solve two problems at once for either theory to be correct.

The biggest missing piece of the SM is gravity. There is of course the enormously successful theory of general relativity, and there is a desire to come up with a theory of *quantum gravity* [162–164] - to explain gravity as a quantum field theory in the same way that the other interactions are. Such a theory would involve a spin-2 *graviton* as the force-carrying gauge boson [165]. Gravity at low energies is much weaker than the other three forces, and so its effects on the quantum level are mostly negligible. However, at around the *Planck scale*, $m_P \sim 10^{19}$ GeV, it is expected that quantum gravitational effects will become important. String theory [166], often incorporating supersymmetry [36, 167] attempts to devise a theory of quantum gravity, but it has struggled to make unique testable predictions.

Several of the models mentioned above predict the existence of particles outside

[8] The particles are dark, i.e. they do not produce light. This means they do not interact with photons, so they are electrically neutral.

[9] Which is why dark matter can be considered to be *under*explained rather than *un*explained - there is a dark matter candidate in the SM, it just doesn't adequately explain the observed phenomena.

the SM which could be directly produced at the LHC. Supersymmetry in particular was expected to be found easily at the LHC [168], as it was probing energy scales that had never been accessed before. However, after data-taking at center-of-mass energies of 7/8 and then 13 TeV in Runs 1 and 2 (Section 4.4), no evidence for supersymmetry, or for that matter any BSM particles, has been found [169]. This has led to the use of LHC data to search for even more exotic particles, in hopes of serendipitously finding evidence for BSM physics, and has also motivated the need for generic searches which would be sensitive to a wide class of new particles, any of which could be e.g. dark matter candidates.

The search presented in Chapter 6 is a search for exotic decays of the Higgs boson. There is a broad class of models in which there exists a new scalar or pseudoscalar a which the Higgs boson can decay into, $H \to aa$. These models can be motivated [55] by explaining the matter-antimatter asymmetry problem, or included in supersymmetric or axion models.

The search presented in Chapter 7 is a generic search for a particle with mass $\mathcal{O}$(TeV) (which must be BSM) decaying to two other particles with mass $\mathcal{O}(100$ GeV) which decay hadronically into jets (which could be BSM or the SM W, Z, or Higgs bosons, or the top quark), $A \to BC \to JJ$. Furthermore, the techniques presented in that search can be extended to search for generic massive particles decaying to BSM or SM particles. These new techniques enable the best possible uses of LHC data to constrain or discover BSM physics.

Chapter 3

Jets

3.1 Introduction

At a hadron collider like the LHC, quarks and gluons are produced copiously. These quarks and gluons are not observed directly - because of color confinement, these *partons* fragment and hadronize almost immediately, leading to a collimated spray of hadronic particles which are ultimately observed in the detector. Other massive particles, like W/Z bosons, the Higgs boson, top quarks, and τ leptons, are also produced and play vital roles in searches and measurements at the LHC. The hadronic decays of these massive particles are the primary way or one extremely important way of identifying their production in an event. *Jets* [64,69,86,170] are algorithmic objects intended to group together the hadronic decay products of these partons or massive particles and derive the four-momentum of the originating particle.

There are a variety of jet algorithms; the most common contemporary one at the LHC is the anti-k_t algorithm [171], with distance parameter R=0.4 (*small*-R *jets*) or R=1.0 (*large*-R *jets*). These algorithms group together constituent four-vector objects called *seeds*. Small-R jets are intended to approximate light particles, in particular quarks and gluons, though in ATLAS τ leptons, photons, and electrons are in addition reconstructed as small-R jets in the detector; large-R jets are intended to approximate boosted massive objects decaying hadronically. Both small- and large-R jets have a variety of *grooming* techniques applied to clean up the substructure in order to remove

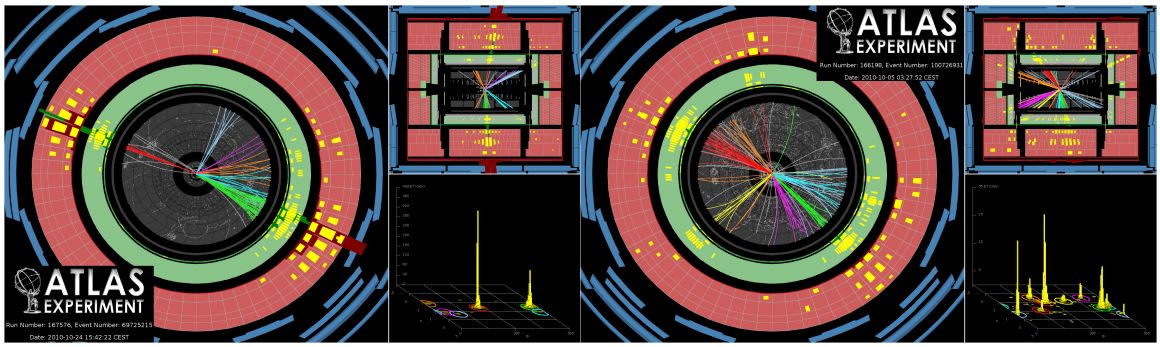
the effects of the underlying event and of multiple in and out of time simultaneous collisions (*pile-up*). There is also often a *tagging* step where the substructure of the jet is used to identify the provenance of the jet.

As almost all particles produced at the LHC are reconstructed as jets, other than muons and neutrinos, jets play an essential role in almost every search and measurement in ATLAS. This Thesis in particular discusses two searches (Chapter 6 and Chapter 7) that use jets extensively, and further delves deeply into methods for improving the calibration of the energy of jets (Chapter 8 and Chapter 9).

This Chapter is organized as follows. Section 3.2 discusses the definition of jets used in ATLAS. Section 3.3 goes over how these jets are groomed, and Section 3.4 discusses how jets are tagged. Finally, Section 3.5 summarizes the Chapter and puts it into context of the rest of the Thesis.

3.2 Jet Definition

Jet algorithms are functions for mapping a set of *seeds* observed in a detector to groupings of those seeds, called jets. For example, Figure 3.1 shows two events recorded in the ATLAS detector - one with two hard jets (*dijet*) and one with multiple (*multijet*).



(a) (b)

Figure 3.1: Two events recorded by the ATLAS detector in data-taking at $\sqrt{s} = 7$ TeV pp collisions. (a) A dijet event. (b) A multijet event. Figures sourced from [172].

In ATLAS seeds are defined either as topologically connected, noise-suppressed

cell-clusters in the electromagnetic and hadronic calorimeters [173], which may themselves be calibrated or uncalibrated; as particle flow objects [174] which combine information from the calorimeter and tracker; or as the tracks themselves without using the calorimeter. From the point of view of the jet algorithm these distinctions are irrelevant, as the seeds are abstracted into individual points with some transverse momentum (p_T) and some position (η, ϕ) in the detector[1]. Figure 3.2a shows an event display with the seeds in the (η, ϕ) plane.

Jets have a long history in high energy physics and various definition with different properties have been used [170]. The Snowmass Accords [175] laid out a series of properties jet definitions should have in order to be both useful experimentally and also tractable theoretically and computationally. In particular, jets should be *IRC* safe, which is a portmanteau of *IR* and *collinear* safety. IR safety requires that soft (i.e., infrared) emissions should not affect the set of jets the algorithm produces, as soft emissions can grow without bound in theoretical computations. Collinear safety requires that if a single seed splits into multiple collinear seeds sharing the original seed's momentum, this should also not affect the output of the algorithm; again, hard partons can undergo many collinear splittings, which can lead to divergences in perturbative QCD calculations. In practice detectors tend to be both IR and collinear safe, as noise-suppression removes soft emissions and the finite size of calorimeter cells and tracking pixels sum up nearly collinear particles; but it is still desired for the algorithm itself to obey these properties for theoretical calculations.

The most common contemporary jet algorithms are called *sequential recombination* algorithms, though there is a long history of other types of algorithms which will not covered here [170]. In these algorithms, every seed is first denoted to be a *proto-jet*, and a distance parameter d_{ij} is defined between every pair of proto-jets i and j. The algorithm then proceeds as follows:

1. Calculate the distance d_{ij} between every pair of existing proto-jets.

2. Find the minimum d_{ij} over all pairs of proto-jets i, j, d_{min}.

[1] Seeds may also be given a small mass, e.g. the pion mass, for the purposes of four-momentum summation.

3. If d_{min} is less than some threshold, then combine proto-jets i and j into a single proto-jet and go back to step 1.

4. Else, terminate and let all existing proto-jets be the output jets of the algorithm.

The k_t jet algorithms use the distance metric

$$d_{ij} = \min\left(p_{Ti}^{2p}, p_{Tj}^{2p}\right)\frac{\Delta R_{ij}^2}{R^2}, \qquad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{3.1}$$

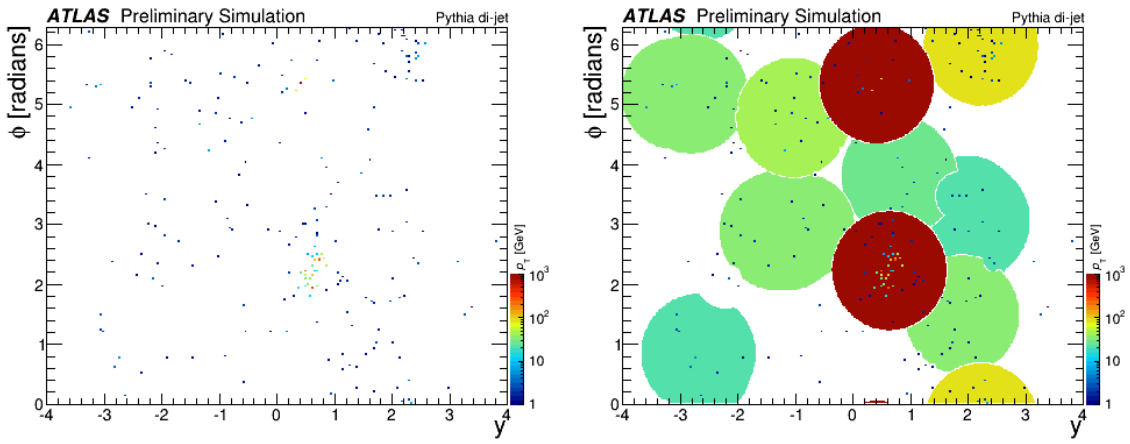
$$d_{iB} = p_{Ti}^{2p} \tag{3.2}$$

with p and R free parameters of the algorithm[2]. d_{iB} is called the beam distance, and if d_{min} is a d_{iB} the proto-jet is removed from the event and added to the list of final jets; there is no minimum threshold but rather the algorithm continues until there are no more proto-jets in the event.

The algorithm with $p = 1$ is called the k_t algorithm [176], with $p = 0$ is called the Cambridge-Aachen algorithm [177], and with $p = -1$ is called the anti-$k_t$algorithm [171]. The anti-k_t algorithm is the standard jet algorithm used in ATLAS and CMS. First, it is manifestly IRC safe - low p_T seeds get added to high p_T proto-jets if $\Delta R < R$ without affecting the proto-jet itself in the limit $p_T \to 0$, and collinear splittings are combined together since $\Delta R \to 0$. Second, the algorithm tends to make circular jets with radius R for isolated clusters, and if there are two clusters within $R < \Delta R < 2R$ then the highest p_T cluster is circular while the second cluster is a crescent. The results of an anti-k_t jet algorithm can be seen in Figure 3.2b.

If there are N seeds in the event, then the k_t algorithms require calculating the pair-wise distance d_{ij} among all proto-jets, which takes $O(N^2)$ time; and then repeating that process until all jets have been found, which in general is $O(N)$. Thus the whole algorithm naively takes $O(N^3)$ time, which is prohibitive for events in

[2]Here the rapidity y is used instead of the pseudorapidity η. y is a kinematic quantity: $y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}$; while η is a geometric quantity: $\eta = -\ln\tan\theta/2$. For massless particles $y = \eta$, and the distinction is meaningless; however, proto-jets may and in general do have non-zero mass. Differences in rapidity are invariant under longitudinal boosts, while differences in pseudorapidity are invariant only for massless particles.



(a) (b)

Figure 3.2: A simulation of a dijet event. (a) The seeds in the (η, ϕ) plane. (b) The results of the anti-k_t jet algorithm with distance parameter $R = 1.0$ (jets with $p_T < 20$ GeV removed). Figures sourced from [178].

which N can be in the hundreds. However, faster algorithms using nearest neighbor algorithms [179] have reduced this computation time to $O(N \log N)$, a significant improvement. In practice this algorithm is executed with the FastJet package [180].

As mentioned above, the most common distance parameters used are $R = 0.4$ (*small*-R *jets*) and R=1.0 (*large*-R *jets*). The distance parameters correspond exactly to the size of the circular cone formed by the jet, and are intended to target the decays of particles with typical sizes on the order of the respective R.

The decay products of light quarks and gluons have a spread due to the hadronic fragmentation. The calculation of the size of this spread is in general very complicated, but jets with $R = 0.4$ are used to capture the decay products of these partons. Gluons are slightly wider than quarks due to the difference in their color charge, and for $R = 0.4$ roughly 89% of the originating particle p_T for gluons and 95% for quarks ends up within the catchment area of the jet [170].

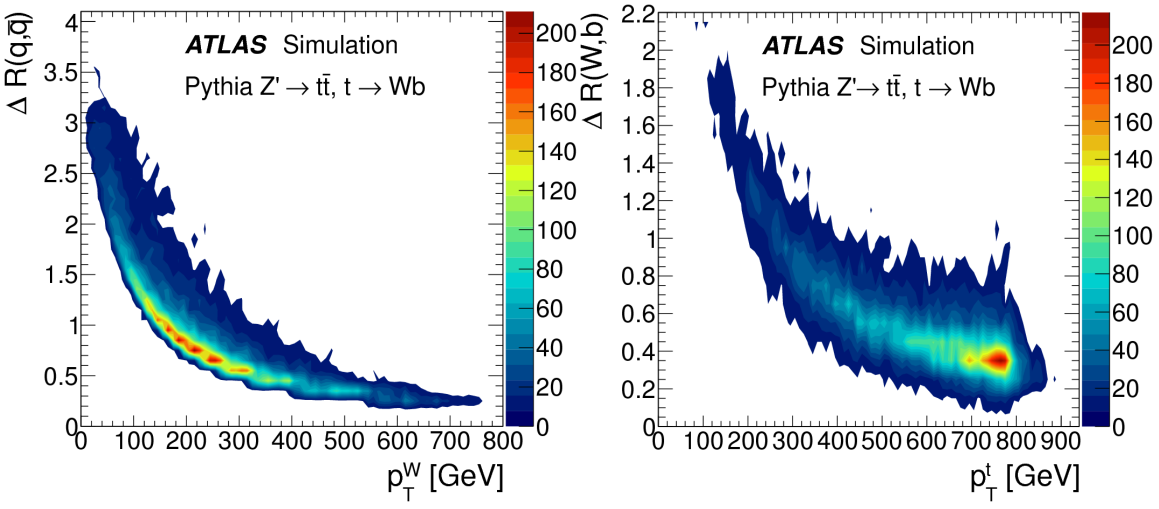
Photons and electrons shower primarily in the electromagnetic calorimeter, but as the seeds are formed from clusters in the electromagnetic and hadronic calorimeters, and the typical size of photon and electron decays is ~ 0.2 [181], these are also reconstructed as small-R jets. However jets originating from photons and electrons

can be identified with high discrimination power based on the three-dimensional energy shower shape [181] (Section 5.5.3), and typically jets that overlap with these are identified and removed from consideration of the hadronic activity in the event ("overlap-removed"). In addition, jets originating from bottom quarks take advantage of the relatively long lifetime of hadrons containing bottom quarks and the resulting presence of tracks with large impact parameters and/or a secondary vertex [182–184] (Section 5.5.4.3).

For massive particles, the opening angle of the decay products is generically roughly $\Delta R \sim \frac{2m}{p_T}$ [185]. With $p_T \gtrsim 2m$, the opening angle is therefore $\Delta R \lesssim 1.0$, and the particle can be reconstructed as a single large-R jet. When the decay products of a massive particle can be reconstructed as a single jet, it is called *boosted*, while for lower p_T it is called *resolved*. τ leptons decay hadronically roughly 65% of the time [107], although due to their light mass they are reconstructed as small-R jets. Again, there are dedicated tagging algorithms with high discrimination power that can identify hadronically-decaying τ leptons [186] (Section 5.5.4.4). Large-R jets are used for particles on the electroweak scale, like W/Z bosons, top quarks, and even boosted Higgs bosons, for $p_T \gtrsim 200$ GeV, which are typical scales for these particles at the LHC. The opening angle ΔR for the decay products of W bosons and top quarks can be seen for example in Figure 3.3. These large-R jets have substructure that can be used to tag the provenance of the jet; these techniques are discussed further in Section 3.4.

3.3 Grooming

A typical pp collision at the LHC produces many radiative particles other than the hard-scatter interaction just from the interactions of all the constituent quarks and internal gluons in the protons; this is referred to as *underlying event*. In addition there are secondary hard-scatter interactions, or *multiple parton interactions*. Furthermore, as will be mentioned in Section 4.3, the collisions occur in whole bunches of 10^{11} protons, so that there are multiple simultaneous interactions in a single bunch crossing; this is referred to as *in-time pile-up*. There is also *out-of-time pile-up*, which



(a) (b)

Figure 3.3: Simulated distributions of ΔR between decay products of massive particles as a function of the particle p_T. (a) For hadronically decaying $W \to q\bar{q}$. (b) For hadronically decaying $t \to Wb$. Figures sourced from [75].

are energy deposits from previous and future bunch crosses which are integrated into the signal shape when reconstructing a seed. Taken together these effects form a layer of soft particles spread diffusely over the entire event, which can end up in the catchment area of jets and add noise to the jet measurements.

Because of this, there is a desire to *groom* jets by removing whatever seeds ended up in their catchment areas and originate from underlying event or pile-up.

Small-R jets are not typically groomed directly, but rather the first step of the calibration process is a jet-wide pile-up subtraction formed by estimating the pile-up density in the event ρ and area of the jet A. However, there are a series of exciting new techniques proposed which do groom small-R jets on the constituent level. The Author has contributed towards implementing some of these techniques [187] by estimating the catchment areas of seeds using Voronoi areas and subtracting pile-up from the seeds directly; these subtracted seeds can further be removed as a grooming procedure.

Large-R jets have a much larger catchment area and are therefore more susceptible to underlying event and pile-up. Though the soft radiation from underlying event and pile-up does not affect the overall energy that much, especially for jets with high p_T,

the jet mass can be significantly adversely affected by soft wide-angle radiation. Since large-R jets are typically used to identify massively decaying particles, it is therefore important to remove this source of noise.

There have been a variety of grooming techniques proposed for large-R jets [75]. For example, one proposed alternative method to form large-R jets is to first find small-R jets, correct for pile-up and calibrate those jets, and then *recluster* them into large-R jets by using the small-R jets as seeds [76].

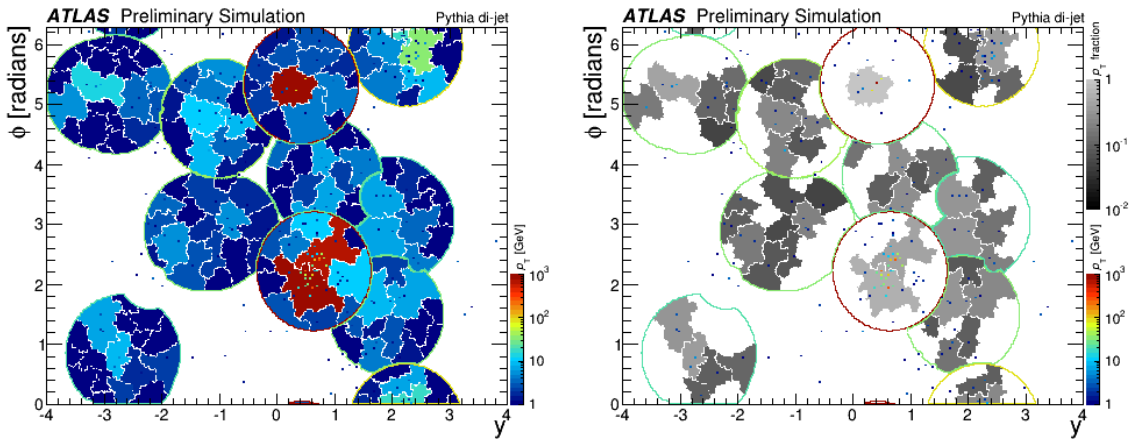
The standard contemporary grooming technique in ATLAS is jet trimming [77]. In trimming, the constituents of a jet are reclustered using a different jet algorithm with smaller distance parameter in order to find distinct subjets. These subjets are then removed ('trimmed') if their p_T fraction of the whole jet is less than some f_{sub}; then the only remaining subjets are hard clusters of energy, removing a significant amount of underlying event and pile-up. In ATLAS the reclustering uses the k_t algorithm with R parameter 0.2, and $f_{sub} = 0.05$; the effect of this trimming can be seen in Figure 3.4. The choice of trimming and these parameters were optimized at the beginning of Run 2 for boosted W tagging [78], though there is still an active effort to understand the best way to groom large-R jets [79].

3.4 Tagging

After grooming, the substructure of jets can be used to tag the provenance of the jet. The term "substructure" here is taken to mean functions of the distribution of the seeds associated to a jet in (η, ϕ, p_T) space. For many of the following taggers the tracks resulting from charged particles that lie within the catchment area of the jet are used to wholly or partially supplant the seeds themselves as tracks are significantly more robust to pile-up than the calorimeter seeds.

Jet substructure has a long history in high energy physics [69, 86]; its use can be traced back to LEP [188–191], Tevatron [192, 193], and HERA [194–196].

Jet substructure for small-R jets focuses on distinguishing quark-initiated from



(a) (b)

Figure 3.4: A simulation of a dijet event, showing the effect of trimming. (a) The reclustered large-R jets, showing the subjets (this event uses $R = 0.3$ instead of the current standard $R = 0.2$). (b) After trimming with $f_{sub} = 0.05$. Figures sourced from [178].

gluon-initiated jets at ATLAS [87–90] and CMS [197–200]. The discrimination typically focuses on the fact that gluon-initiated jets are slightly wider than quark-initiated jets. Jet substructure has also been used to discriminate between small-R jets that arise from a single particle and those that arise stochastically from multiple pile-up interactions [91].

For large-R jets a variety of techniques are employed to tag those jets originating from W/Z bosons [78,92,94], top quarks [93,94] and boosted Higgs [95] decaying to two bottom quarks. The fully hadronic decays that these taggers target are $W/Z \to q\bar{q}$, $t \to Wb \to q\bar{q}b$, and $H \to b\bar{b}$, respectively (there are also decays involving leptons or photons which as mentioned have their own dedicated taggers). Of course, in addition to the substructure itself the decays involving b quarks can use the dedicated b-tagging algorithms mentioned above to further discriminate. The intention of these taggers is primarily to discriminate these jets from those originating from quark and gluons; though as mentioned above small-R jets are sufficient to capture these particle showers, expanding the R parameter in the jet algorithm still finds these, albeit with a small hard core. The most obvious substructure variable is the mass of the jet

itself [201]; but in addition, each of the above taggers rely on the fact that the decays of these objects will have multiple hard cores. While all of these taggers now make use of a complicated combination of multiple substructure variables, a prominent one is n-subjettiness [202], which gives a sense of how compatible a given jet is with having n hard subjets.

3.5 Conclusion

Jets are ubiquitous at a hadronic collider like the LHC. As generic groupings of detector seeds, almost every Standard Model particle other than muons and neutrinos are reconstructed as jets. The precise definition of the jet algorithm is a choice, governed by the physics requirements and implications of that choice. Both small-R and large-R jets are groomed to remove soft additions due to underlying event or pile-up. Finally, a suite of taggers has been developed to tag jets originating from each individual particle.

Jets are interesting theoretical, algorithmic, and physical objects in and of themselves. In this Thesis in particular, Chapter 6 and Chapter 7 discuss two searches that use jets extensively, and Chapter 8 and Chapter 9 further delve deeply into methods for improving the calibration of the energy of jets. In light of this material, the use of jets can be considered a central theme of this Thesis.

Chapter 4

The Large Hadron Collider

4.1 Introduction

The *Large Hadron Collider* [57–60], or *LHC*, is a high energy particle accelerator and collider located at *CERN* near Geneva, Switzerland. The LHC is the largest machine ever built, with a circumference of 27 km, running 100 m underground and straddling the border between Switzerland and France. The purpose of the LHC is to produce the highest energy particle collisions ever, with a design center-of-mass energy of $\sqrt{s} = 14$ TeV[1]. A view of the LHC superimposed on an aerial view of Geneva and the surrounding areas can be seen in Figure 4.1, including the four main LHC experiments: ATLAS [56], CMS [203], LHCb [126], and ALICE [204]. It was built at a cost of about 4.5 billion USD over the course of almost 20 years [57], in addition making it one of the most expensive scientific experiments ever built.

The main LHC experiments are built around particle interaction points and are themselves extremely impressive scientific apparatuses. This Thesis uses data gathered at the ATLAS experiment (Chapter 5), which is the largest particle detector ever built by volume, with a diameter of 25 m and a length of 44 m [206]. The

[1] The *center-of-mass energy* of two colliding beams is $\sqrt{s}$, where s is the Mandelstam variable $s = (p_1 + p_2)^2$, and p_1 and p_2 are the four-momenta of the two beams. In the case that the beams collide head-on, with $E_1 = E_2 \equiv E \gg m_1 = m_2$, $\sqrt{s} = 2\sqrt{E_1 E_2} = 2E$ (which is the case in the LHC). I.e., the protons in each beam have energy 7 TeV, and they collide basically head-on, so $\sqrt{s} = 14$ TeV.

(a)

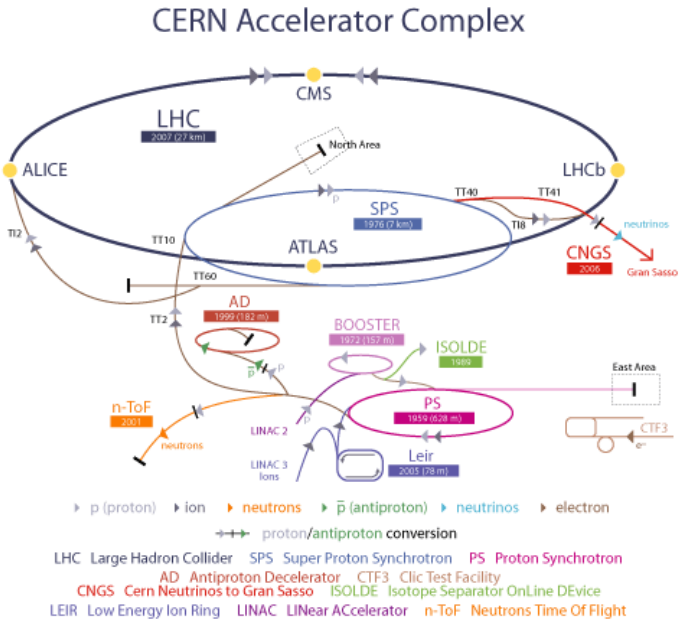
Figure 4.1: An image of the LHC (yellow circle) superimposed on an aerial view of Geneva and the surrounding areas. The white dashed lined indicates the French/Swiss border. The two main CERN sites at Meyrin and Prévessin, and the four main LHC experiments (ATLAS, CMS, LHCb, and ALICE) are indicated. Geneva is at the tip of Lac Léman, and Mont Blanc can be seen in the background. Figure sourced from [205].

"rival" experiment to ATLAS is CMS, which is slightly smaller than ATLAS but is the heaviest particle detector ever, at 14000 tons [207]. The LHC also represents a remarkable international scientific collaboration, with over 10000 scientists from at least 40 countries [57] working on one of the experiments or on LHC operations.

This Chapter goes over the design of the LHC (Section 4.2), the conditions of particle collisions (Section 4.3), and finally the data taking history and future plans for the LHC (Section 4.4).

4.2 Design

The LHC is the final stage of the CERN accelerator complex, which accelerates protons from rest to TeV-scale energies through a series of accelerators, with a final speed of $> 99.99999\%$ of the speed of light. The entire CERN accelerator complex can be seen in Figure 4.2.



(a)

Figure 4.2: The CERN accelerator complex. Figure sourced from [208].

Protons start as ordinary ionized hydrogen, which are accelerated [57] in the *Linear*

Accelerator 2 (Linac 2) to an energy of 50 MeV. They are then passed through the *Proton Synchrotron Booster* (PS Booster) and the *Proton Synchrotron* (PS) to final energies of 1.4 GeV and 25 GeV, respectively. The final intermediate stage before the LHC is the *Super Proton Synchrotron* (SPS), which accelerates the protons an injection energy of 450 GeV. The LHC itself then accelerates the protons to an energy of up to 7 TeV, though operating histories have been less than this - 3.5 TeV in 2010, 4 TeV in 2011-2012, and 6.5 TeV in 2015-2018 data taking (Section 4.4).

The LHC is also used to accelerate and collide lead ions, which reach energies of up to 2.56 TeV.

The particles are accelerated using radiofrequency cavities [209, 210] which in the LHC operate at 400 MHz - every tenth wave is filled with a *bunch* of about 10^{11} protons, for a bunch rate of 40 MHz, and 2808 bunches circulating the LHC per proton beam. There are two beams circulating in opposite directions in adjacent beam pipes, which come together at a few collision points around the ring, with a *bunch crossing* every 25 ns; the detectors are built around these collision points.

Accelerating charged particles radiate [211], so that the accelerators have to constantly input energy in order to maintain constant beam energy. However, the radiation from charged particles moving in circular motion, called *synchrotron radiation*, goes as $\frac{E^2}{m^4}$, so this radiation is strongly suppressed for protons relative to electrons and positrons at the same energy. As a historical note, the LHC reuses the tunnel built for the *Large Electron-Positron Collider* [212–214], or *LEP*, which collided electrons and protons at a maximum center-of-mass energy of 209 GeV. However, when increasing the energy of the beams for the LHC, it was decided to switch to proton-proton collisions which suppresses the synchrotron radiation.

The particles are kept along a circular trajectory with a perpendicular magnetic field produced by dipole magnets[2]. The radius of curvature R, magnetic field B, and

[2]There are also quadrupole, sextupole, etc. magnets to provide second-order effects like stabilization and beam focusing.

momentum p[3] are related [210] via

$$BR = \frac{p}{|q|}, \tag{4.1}$$

where q is the charge (for protons, $q = +1$ elementary charges). As the protons are accelerated around the ring the magnetic field increases proportionally in order to keep the radius fixed. A given magnet can only support a magnetic field range of $\mathcal{O}(100)$, which is why the energy increases by a factor of $20-25$ in each stage of the accelerator complex before moving to the next stage with a larger R.

In the final stage, in the LHC, 1232 dipole magnets provide a magnetic field of up to 8.3 T (there are almost 10000 total magnets, including those with higher orders) [57]. The magnets are made of superconducting NbTi cables cooled down to 1.9 K with superfluid helium.

The LHC consumes about 750 GWh per year [215], most of which is for the cryogenics system, and CERN in general consumes about 1300 GWh per year; in comparison, the entire Geneva canton consumes about 3000 GWh per year[4].

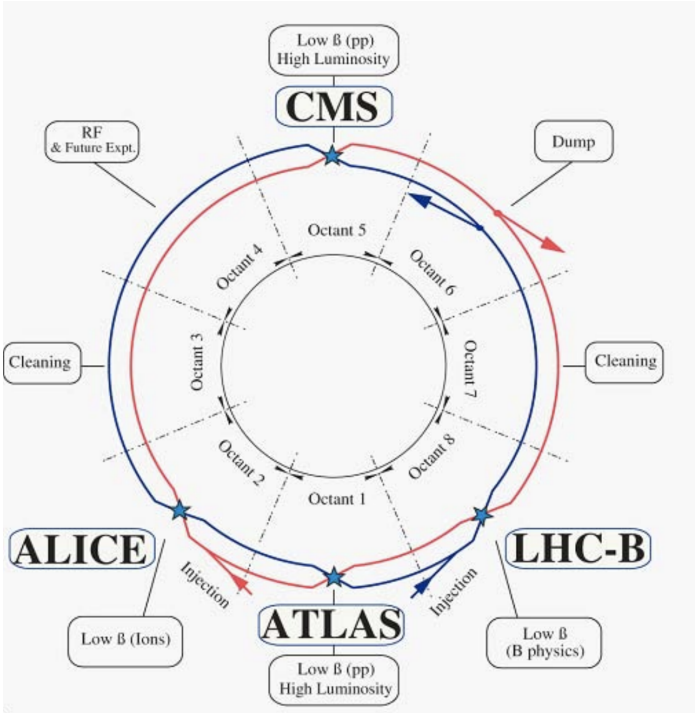
The LHC is split into eight *octants*, each of which supports its own cryostat. As can be seen in Figure 4.3, each octant has some special feature, like the two injection sites, beam dump, beam cleaning, and of course the four collision points. ATLAS, in the center of Octant 1, has the distinction of being located at "Point 1", which is also near where (on surface level) the main CERN Meyrin campus is, and the easiest to access from downtown Geneva; CMS on the other hand is at "Point 5", at the CERN Prévessin campus in the middle of the French countryside.

4.3 Luminosity and Pile-up

The rate of collisions is called the *luminosity* $\mathcal{L}$ [100] (or *instantaneous luminosity*), and has units of inverse area per unit time. The expected number of events N per

[3]Note that, as the particles are highly relativistic, $p = \gamma m v$. Also, since $m_p \approx 1$ GeV $\ll \mathcal{O}(\text{TeV}) =$ beam energy, $p \approx E$.

[4]CERN is actually connected to the French electrical grid, not the Swiss [216].



(a)

Figure 4.3: Schematic of the LHC showing the two circulating beams, the four main experiments, and the eight octants with some of their special features. Figure sourced from [60].

unit time of a particular process with cross-section σ is

$$\frac{dN}{dt} = \mathcal{L}\sigma. \tag{4.2}$$

In particle physics a commonly used unit of area is the *barn*, which is equal to 10^{-24} cm^2 [217]. The LHC has a peak luminosity of about 10^{34} cm^{-2}s^{-1} = 10 nb^{-1}s^{-1} [60][5]. As a point of comparison, the total proton-proton cross section at 13 TeV is about 10^8 nb, as can be seen in Figure 4.4. While at the LHC the total proton-proton cross section is measured (e.g. in ATLAS, broken down into total elastic cross section [218] and total inelastic cross section [219]), one primary physics goal of the LHC is to provide inelastic scattering of the proton constituents (partons) at the electroweak scale ($\mathcal{O}(100)$ GeV), e.g. in particular the production of a Higgs boson. The cross section for these "interesting" processes are at least a factor of 10^{-5} less than the total cross section, and many (like production of the Higgs boson) are much rarer; therefore in order to produce enough events to claim discovery of a new particle with statistical significance, increasing luminosity is critical.

The luminosity at the LHC is given by the formula

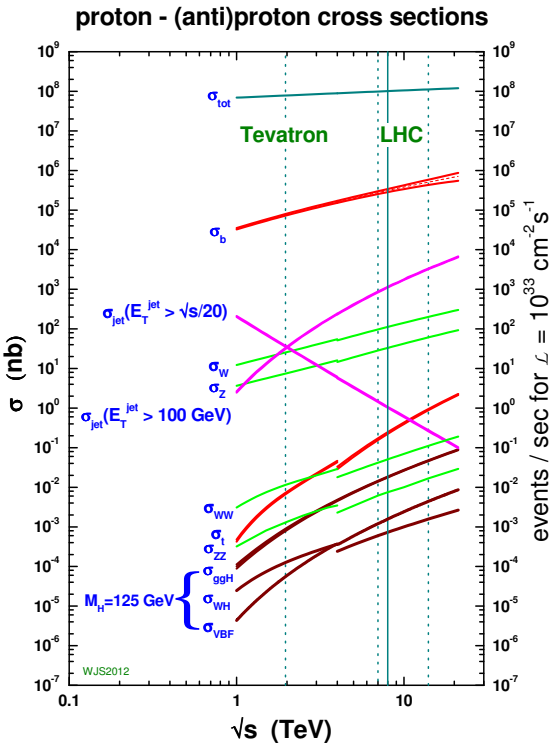
$$\mathcal{L} = \frac{N^2 n f \gamma}{4\pi\epsilon^*\beta^*}F, \tag{4.3}$$

where N is the number of particles per bunch, n the number of bunches per beam, f the revolution frequency, γ the relativistic factor, ϵ^* the *normalized emittance*, β^* the *beta function* at the collision point, and F a geometric factor based on the angle of the beam crossing.

The *emittance* ϵ [221, 222] is a quantity that measures the area or spread in phase space (position-momentum space $x - p$) taken up by the beam - the locus of points in phase space usually forms an ellipse, and the emittance is the area of this ellipse[6]. An example locus of points in phase space can be seen in Figure 4.5.

[5] Just to be clear, 1 nb^{-1} = 10^{+9} b^{-1}, etc.

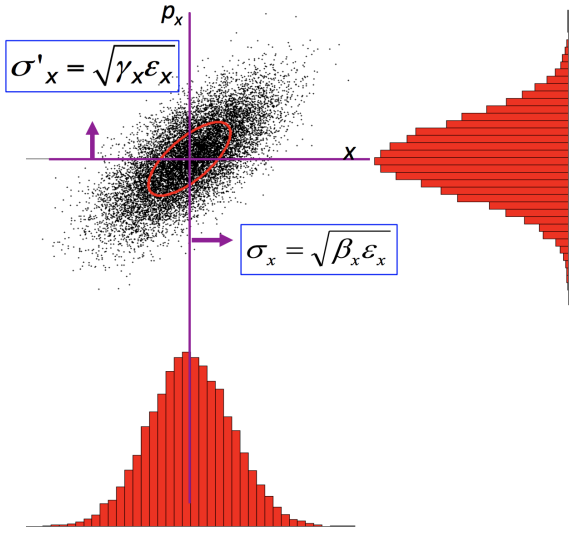
[6] More precisely, the set of points in the beam can be considered to be drawn from a two-dimensional Gaussian distribution in phase space, for which the contours of constant probability density are ellipses. The area is appropriately defined according to some integral over this probability density.



(a)

Figure 4.4: Cross sections for pp interactions as a function of $\sqrt{s}$ center-of-mass energy. Figure sourced from [220].

According to Liouville's theorem [223,224], the emittance is an invariant for a beam at constant energy being acted on by external magnetic fields (disregarding synchrotron radiation). As the beams are accelerated to increase the energy and longitudinal momentum, the emittance shrinks as $\frac{1}{p}$, a phenomenon known as *adiabatic damping*. The *normalized emittance* $\epsilon^* = \frac{p}{m}\epsilon$ takes into this into account and is an invariant for the beam the whole time it is being accelerated.



(a)

Figure 4.5: Example locus of points in a beam in phase space, showing the distributions in x and p_x. The red ellipse corresponds to a contour of constant probability density, and the area of the ellipse is (up to a constant factor) the emittance ϵ. Given ϵ, the spread in x is given by β, and the spread in p_x is given by γ. Figure sourced from [225].

Though the (normalized) emittance is an invariant, there is still flexibility to control the shape of the beam locus in phase space. The parameter that controls the spread along the position axis is called β and the spread along the momentum axis γ[7]. This is illustrated in Figure 4.5. In order to maximize the luminosity, β is desired to be as small as possible in order to reduce the size of the beam in the transverse

[7]These, along with a cross-term parameter α, are called the *Twiss parameters* in Hamiltonian mechanics. Not to be confused with the relativistic $\beta = v$ (with c=1) and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ - there are unfortunately too few Greek letters for all the uses physics has for them.

plane[8], at the expense of increasing the spread in momentum space. The minimum β reached at the interaction point is called β^*.

As mentioned above, the rate of proton-proton interactions is roughly $\mathcal{L}\sigma_{pp} \approx 10^9$ s^{-1}, and the rate of bunch crossings, or *events*[9], is 40 MHz (= 25 ns between bunch crossings). This means that there is more than one proton-proton interaction per bunch crossing, a phenomenon known as *pile-up*. The average number of interactions per bunch crossing μ is a direct measurement of the luminosity and amount of pile-up. Also as mentioned above, almost all proton-proton interactions are "uninteresting", so that even with $\mu > 1$ most events need to be thrown away due to disk writing and space requirements; in ATLAS there is a trigger (Section 5.4) that decides in real-time whether or not to store an event based on whether at least one of the interactions was "interesting". In addition, even if an event is triggered, it is almost certain that there is at most one "interesting" interaction in the event; however, the remaining interactions cover the detector with a mostly uniform soft layer of particles, introducing a major source of noise in the energy measurements of physics objects in the detector. In particular, jets (Chapter 3) can include energy deposits from pile-up interactions, requiring this contribution to be subtracted and calibrated, which is the focus of Chapters 8 and 9.

In ATLAS, the luminosity is measured and monitored [226] using detectors downstream from the interaction point, almost along the beam line, and using the well-measured total differential inelastic proton-proton cross section.

The luminosity gives the rate of proton-proton collisions per unit time. The *integrated luminosity* $\int \mathcal{L} dt$ is the integral of the luminosity over the data-taking run, and has units of inverse area. The total number of events produced over a data-taking run for a process with cross section σ is $N = \sigma \int \mathcal{L} dt$, and so the integrated luminosity is a key quantity for understanding the statistical power of a dataset.

[8] There is an emittance corresponding to each of the longitudinal direction and the two transverse directions. For the colliding beams the relevant quantity is the size in the transverse directions - the two transverse directions are treated symmetrically, assumed to have an equal ϵ^* and β^*.

[9] Since the bunch crossing occurs in a short span of time (much less than the time between bunch crossings), the products of all interactions in a single bunch crossing are recorded in the detector at once and so the bunch crossing is the natural discretization unit.

4.4 Data Taking History and Future

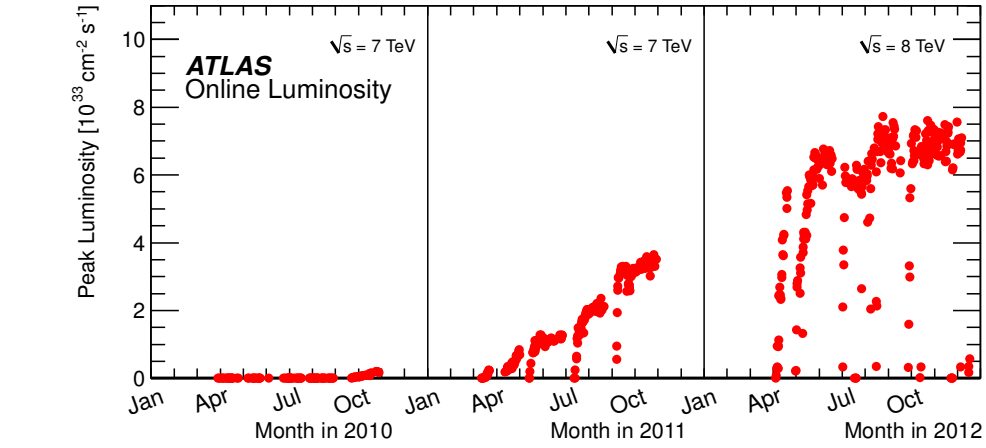
The LHC turned on for the first time on 10 September 2008 [227], though due to "the incident" on 19 September 2008 [228], it did not turn back on again until 20 November 2009, and did not start recording collisions at its target (less than design) center-of-mass energy of 7 TeV until 30 March 2010. In 2010 and 2011 the LHC ran successfully at $\sqrt{s} = 7$ TeV and in 2012 this was bumped up to $\sqrt{s} = 8$ TeV. The data-taking run consisting of 2010-2012 is called *Run 1*, and was particularly important as ATLAS and CMS jointly announced the discovery of the Higgs boson [25, 26] based on this dataset. Figure 4.6 shows the luminosity, μ distributions, and integrated luminosity over time in Run 1[10]. The luminosity was gradually ramped up over the course of the Run to a maximum of almost 8×10^{33} cm^{-2}s^{-1}. This was achieved in part by increasing μ from $5 - 15$ in 2010 and 2011 to $10 - 30$ in 2012. The total integrated luminosity delivered was 48.1 pb^{-1} in 2010 and 5.46 fb^{-1} in 2011 at $\sqrt{s} = 7$ TeV and 22.8 fb^{-1} in 2012 at $\sqrt{s} = 8$ TeV [229].

After Run 1 the LHC was shut down for *Long Shutdown 1* until resuming collisions at $\sqrt{s} = 13$ TeV in June 2015. The center-of-mass energy remained at 13 TeV for data-taking in 2015, 2016, 2017, and 2018, a period known as *Run 2*. The searches in this Thesis use data taken during Run 2 - the search in Chapter 6 uses data from 2015-2016, while the search in Chapter 7 uses the entire Run 2 dataset 2015-2018. Figure 4.7 shows the luminosity, μ distributions, and integrated luminosity over time in Run 2[11]. The luminosity was gradually ramped up over time and reached a maximum of 21×10^{33} cm^{-2}s^{-1} in 2018. The μ distribution was higher than in Run 1, broadly ranging from $10 - 60$ with a mode at around $\langle\mu\rangle \approx 30$. The total integrated luminosity delivered was 4.2 fb^{-1} in 2015, 38.5 fb^{-1} in 2016, 50.2 fb^{-1} in 2017, and 63.3 fb^{-1} in 2018, for a total of 156 fb^{-1} [230].

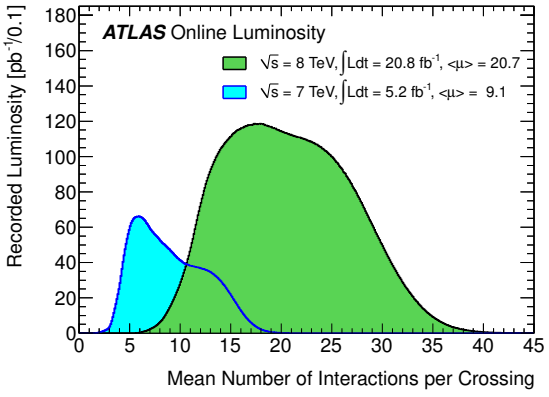
At time of writing, the LHC is currently in *Long Shutdown 2*, which is projected to last until 2021. Data taking in *Run 3* will last from 2021 to 2024 [231] at the design

[10] What is being shown is the ATLAS recorded luminosity, which is slightly less than the LHC delivered luminosity.

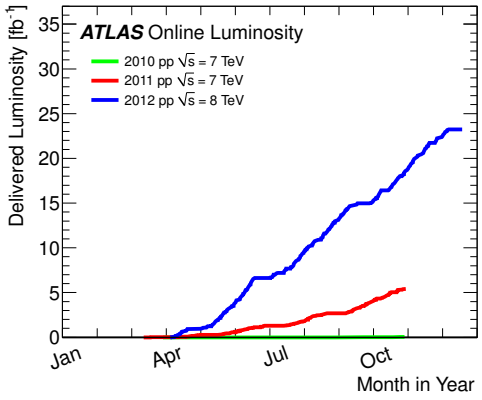
[11] Again, what is being shown is the ATLAS recorded luminosity, which is slightly less than the LHC delivered luminosity.



(a)



(b)



(c)

Figure 4.6: (a) Luminosity vs time in Run 1. (b) μ distributions in 2010-2011 at $\sqrt{s} = 7$ TeV and in 2012 at $\sqrt{s} = 8$ TeV. (c) Integrated luminosity vs time in Run 1. Figure sourced from [229].

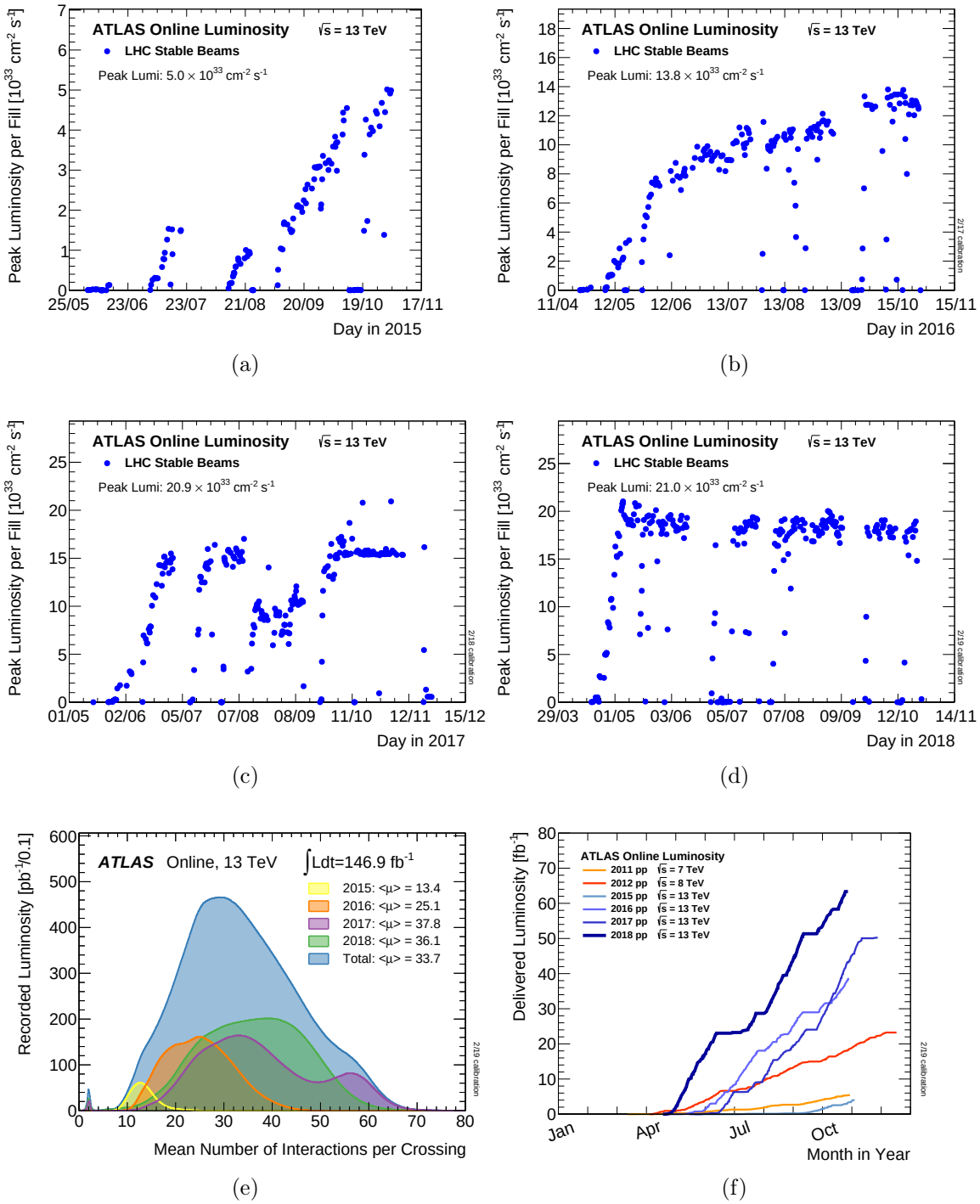


Figure 4.7: (a,b,c,d) Luminosity vs time in 2015, 2016, 2017, 2018, respectively. (b) μ distributions in Run 2. (c) Integrated luminosity vs time in Runs 1 and 2. Figure sourced from [230].

center-of-mass energy of 14 TeV and a peak luminosity of 2.0×10^{34} cm^{-2}s^{-1} [232], with $\langle \mu \rangle \approx 55$, similar to data-taking in 2018. Run 3 is targeted to provide about 100 fb^{-1} per year, and a total of $300 - 400$ fb^{-1} total [233].

The LHC will then be shut down again for *Long Shutdown 3* until 2027. From 2027 to 2030 the LHC will be running in *Run 4*, then shutting down for *Long Shutdown 4* until *Run 5* from 2032 to 2034 [231]. Runs 4 and 5 are also known as *high-luminosity LHC* or *HL-LHC* [234], as the luminosity will be significantly increased. The target peak luminosity is 5.0×10^{34} cm^{-2}s^{-1} [232], achieved mostly by decreasing β^* by a factor of almost 2, and increasing the number of protons per bunch also by a factor of almost 2 relative to Run 2. The pile-up in HL-LHC will be at $\langle \mu \rangle$ between 150 and 200, presenting a significant challenge to suppress this amount of noise, especially in low energy jets. Due to this increased luminosity, HL-LHC is expected to deliver almost 3000 fb^{-1} over the course of the two Runs.

Chapter 5

The ATLAS Detector

5.1 Introduction

The ATLAS detector [56] is a general-purpose particle physics detector with nearly 4π coverage in solid angle around the collision point.[1] The physics results that ATLAS produces are enabled not only by the hardware that measures the properties of outgoing particles, but also by software which simulates, stores, and processes this enormous amount of data.

The detector itself is designed to have many concentric layers serving different purposes, in particular to detect and identify all kinds of particles that may be encountered from the collisions at the LHC (Section 5.2).

The physics phenomena that occur in the pp collisions provided by the LHC and their subsequent interactions with the ATLAS detector are simulated using a variety of generators and a detailed detector simulation (Section 5.3).

Because of the extremely high rate of events and large amount of data read out per event, there is a trigger system which lowers the rate that is written to disk (Section 5.4).

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the center of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

Finally, the readouts of all the various detector subsystems are combined to construct objects intended to match individual Standard Model particles (Section 5.5). Tracks and calorimeter clusters are combined to form detector-level photons, electrons, taus, jets (Chapter 3), and muons. All of these objects are taken together to calculate the missing energy in the event which could be due to neutrinos or beyond-the-Standard-Model physics.

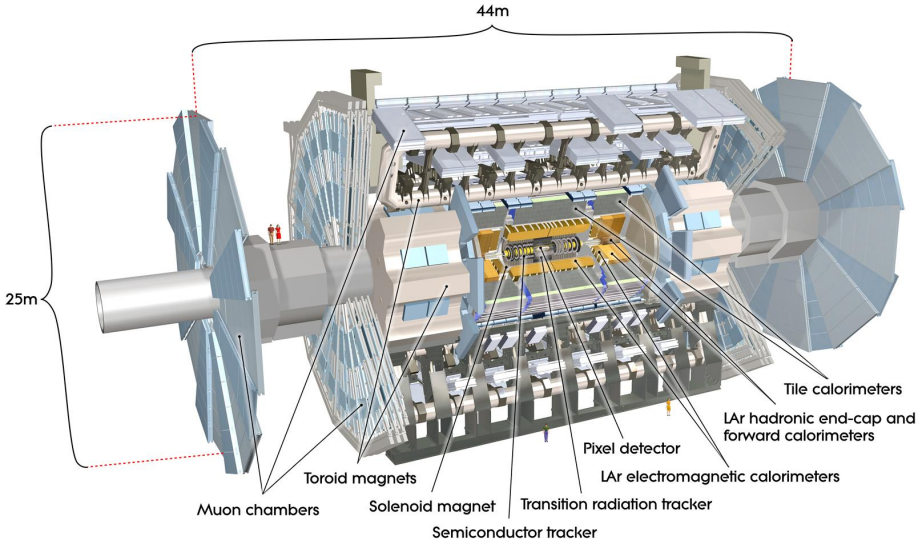
5.2 Hardware

As mentioned above, the ATLAS detector (A Toroidal LHC ApparatuS) is a general purpose particle physics detector built around the nominal interaction point for pp collisions provided by the LHC[2]. It is an enormous instrument, roughly cylindrical with a diameter of about 25 m and a length of about 44 m. A cutout of the ATLAS detector with some parts labeled can be seen in Figure 5.1.

ATLAS consists of an inner tracking detector surrounded by a superconducting solenoid providing a 2 T axial magnetic field (Section 5.2.1), a system of calorimeters (Section 5.2.2), and a muon spectrometer incorporating three large superconducting toroid magnets (Section 5.2.3). The various layers target different kinds of particles based on their interactions with matter, as can be seen in Figure 5.2.

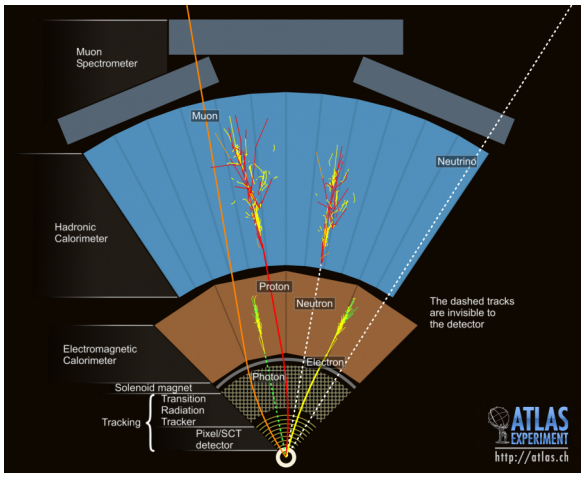
This schematic only shows the ideal or targeted case; in practice the object identification is a non-trivial problem (Section 5.5). Charged particles leave tracks in the tracker, which is surrounded by a solenoid magnet to bend their tracks and measure charge and momentum. Photons and electrons are stopped and their energies measured in the electromagnetic calorimeter, while hadrons interact to a lesser extent and are ultimately stopped and measured in the hadronic calorimeter. Muons pass through the entire calorimeter system and are measured in the muon system, which is surrounded by superconducting toroids. Finally, neutrinos interact only very weakly with matter and pass right through the detector; these can only be reconstructed as

[2]This Section is sourced mainly from the description of ATLAS in [56]; however, the Author is also grateful for the additional explanations that can be found in previous SLAC ATLAS students' PHD theses, in particular [235] and [236]. The later Sections of this Chapter also benefit from these sources.



(a)

Figure 5.1: A cutout view of the ATLAS detector with major subsystems labeled. People included for scale. Figure sourced from [56].



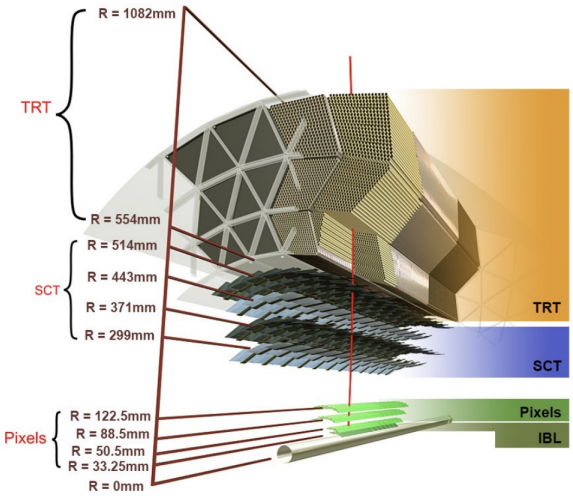
(a)

Figure 5.2: A schematic of various particles passing through the ATLAS detector. Figure sourced from [237].

missing energy and momentum in the event.

5.2.1 Tracker

The tracker, or inner detector (as it is the closest part of ATLAS to the beamline), consists of 3 layers with different technologies: the silicon pixel detectors, the semiconductor strip tracker (SCT), and the transition radiation tracker (TRT). The entire tracking system is immersed in a 2 T magnetic field provided by a surrounding solenoid. These components, when taken together, provide charged-particle tracking in the range $|\eta| < 2.5$. A cutout view of the tracking system can be seen in Figure 5.3.



(a)

Figure 5.3: A cutout view of the tracking system showing the various layers. Figure sourced from [238].

The pixel detectors are the innermost part of the ATLAS detector to the beamline. The original design included 3 layers, although a 4th layer, the insertable B-layer (IBL), was installed between Run 1 and Run 2 in 2014. The pixel detectors are composed of silicon and operate as ionizing radiation detectors. As charged particles pass through the material, electrons are knocked loose and these are measured in each individual pixel, without substantially affecting the momentum of the charged

particle. The pixel system has very good spatial resolution, with each pixel having a size of 50×400 μm^2 in the outer 3 layers and 50×250 μm^2 in the IBL. There is both a cylindrical set of pixel detectors in the barrel and an endcap set on the ends. The precision of the pixels is 10×115 μm^2 in $R - \phi \times z$ in the barrel and also 10×115 μm^2 in $R - \phi \times R$ in the endcaps. There are roughly 80.4 million independent pixel channels.

The next outermost layer is the SCT, which operates under very similar principles to the pixel detectors. However, instead of small pixels, long and thin strips are used, which provide spatial information in only one direction. Because of this, each of the 4 layers of the SCT are actually composed of 2 layered and slightly offset strips at an angle of 40 mrad to each other, in order to get a (rough) measurement in the second direction as well. In the barrel region these strips are parallel to the beam direction; in the endcaps they are radial. The strips have a precision of 17×580 μm in $R - \phi \times z$ in the barrel and also 17×580 μm in $R - \phi \times R$ in the endcaps. There are approximately 6.3 million readout channels in the SCT.

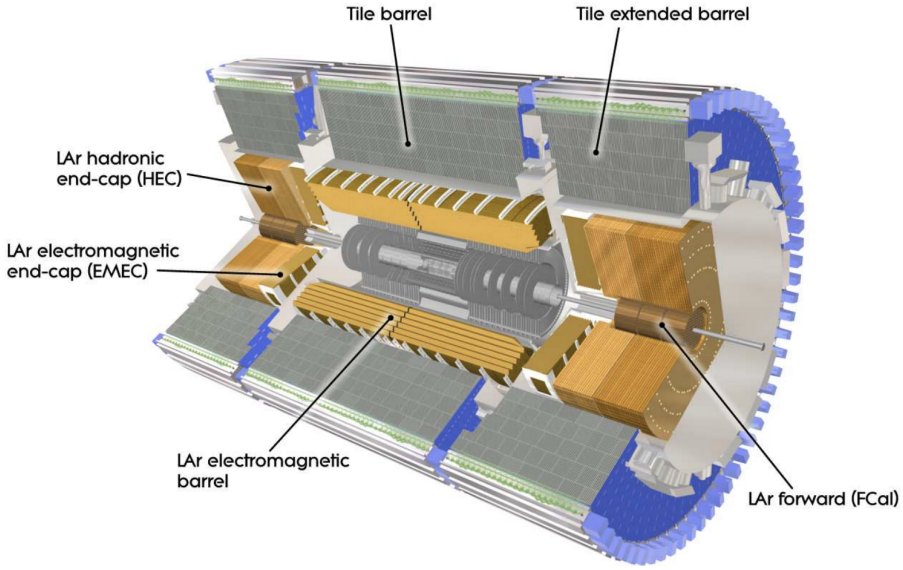
The final layer consists of the TRT, which is composed of 4 mm (in diameter) drift tubes. These drift tubes are not made out of silicon, but rather are filled with gas which, as charged particles pass through, gets ionized; this signal is amplified by a large voltage difference (1530 V) between the center and exterior of the tube. The tubes do not provide any z resolution, but only information in $R - \phi$ with an precision of 130 μm, both in the barrel and endcap regions. However, this is mitigated somewhat by every particle passing through about 30 tubes before exiting the tracker. There are approximately 351000 readout channels in the TRT.

The entire tracking system is surrounded by a superconducting solenoid which generates an axial field of 2 T. As charged particles pass through this magnetic field, their total momentum is not changed as magnetic fields do no work, but the direction of the momentum curves in the ϕ direction with the radius of curvature determined by the charge-to-(transverse) momentum ratio.

The individual hits in the various layers are combined together in software to identify paths of charged particles through the tracker and magnetic field, measuring both the charge and momentum of the charged particle (Section 5.5.1).

5.2.2 Calorimeters

The ATLAS calorimeter system is designed to stop and measure the energy of all charged and neutral particles that exit the tracker other than muons and neutrinos. These calorimeters cover the region $|\eta| < 4.9$, in comparison to the tracker which only covers $|\eta| < 2.5$. The calorimeter system consists of two main subsystems: the inner electromagnetic calorimeter, which is intended to interact electromagnetically and measure particles like photons and electrons, and the outer hadronic calorimeter, which is intended to interact both electromagnetically and via nuclear interactions in order to measure hadronic particles. A cutout view of the calorimeter system can be seen in Figure 5.4.



(a)

Figure 5.4: A cutout view of the calorimeter system showing the various subsystems. Figure sourced from [56].

Both subsystems are sampling calorimeters which operate under the principle of alternating passive and active layers. The passive layers are made of some dense material (lead, steel, copper, or tungsten) that have high probability of interacting with the energetic particles passing through them, causing a cascade of lower energy radiation that is easier to measure (*sample*) in the active materials. The active materials

use ionization or scintillation to measure the energies of the particles passing through them. In ioniziation, the material is ionized by electrons being knocked free; the free electrons are then drifted to the side of the cell and measured. In scintillation, excited molecules emit photon radiation which can then be read out by photomultiplier tubes. All systems are non-compensating, meaning they do not account for energy loss in the passive layers; this is one cause for the need to calibrate the energy of physics objects formed in the calorimeter, in particular jets (Section 5.5.4.1). However, the innermost layer of electromagnetic calorimeter is a presampler which does compensate for the energy loss in the tracker.

High energy charged electrons and positrons interacting with nuclei in the materials are dominated by bremsstrahlung, which is photon radiation due to deceleration in the material [239]. High energy photons, in turn, primarily convert to electron-positron pairs in the field of nuclei, which further lose energy to bremsstrahlung; this back-and-forth is referred to as an *electromagnetic shower*. Below a certain energy (depending on the material, but usually some MeV), other processes take over. Muons can be considered to be heavy electrons; however, because of their higher mass, bremmstrahlung does not dominate until the muon energy is above ~ 1000 GeV [240]; thus muons tend to pass through the entire calorimeter without losing much energy and are reconstructed using hits in the muon spectrometer (combined with tracks from the ID - Section 5.2.3).

The loss of energy in an electromagnetic shower at high energies can be characterized as

$$\frac{dE}{dx} = -\frac{E}{X_0}, \tag{5.1}$$

where X_0 is called the *radiation length*, and is characteristic of the interacting material. The solution to 5.1 implies an exponential loss of energy as a function of distance in the material, meaning the shower length is logarithmic in initial energy and also that a fixed size detector can cover many orders of magnitude of energy.

For nuclear interactions the underlying processes are more complicated, but the *hadronic interaction length* λ gives a similar length scale for hadrons passing through a

material and forming *hadronic showers* [240] of pions, photons, and positrons/electrons. For dense materials $\lambda > X_0$ by a factor of 5-10 [239], implying that hadronic showers are much longer and occur later than electromagnetic showers; this is why the hadronic calorimeter is outside the electromagnetic calorimeter.

The electromagnetic calorimeter is broken down into the barrel, which covers $|\eta| < 1.475$, and two end-caps which cover $1.375 < |\eta| < 3.2$. Each of these have three layers in addition to the innermost presampler layer. Both of these use lead as the passive material and liquid argon (LAr) as the active material, which measures the energy of particles via ionization. The total thickness of the electromagnetic calorimeter is $> 22X_0$ in the barrel and $> 24X_0$ in the endcaps. The size of the calorimeter cells vary with η and depth, but the smallest cells, which occur in the second layer, are 0.025×0.025 in $\Delta\eta \times \Delta\phi$.

The liquid argon cells take a long time to collect the ionization energy, on the order of 450 ns [241], which is significantly longer than the time between bunch crossings (25 ns). Luckily, the ionization pulse shape can be fit in order to determine the bunch crossing the energy deposit corresponds to. However, there is some uncertainty on this assignment, leading to energy deposits from neighboring bunch crossings adding a source of noise to the set of energy deposits corresponding to a single event; this noise is called *out-of-time* pile-up, to contrast it with the other interactions in the same bunch crossing (Section 4.3), which is called *in-time* pile-up[3].

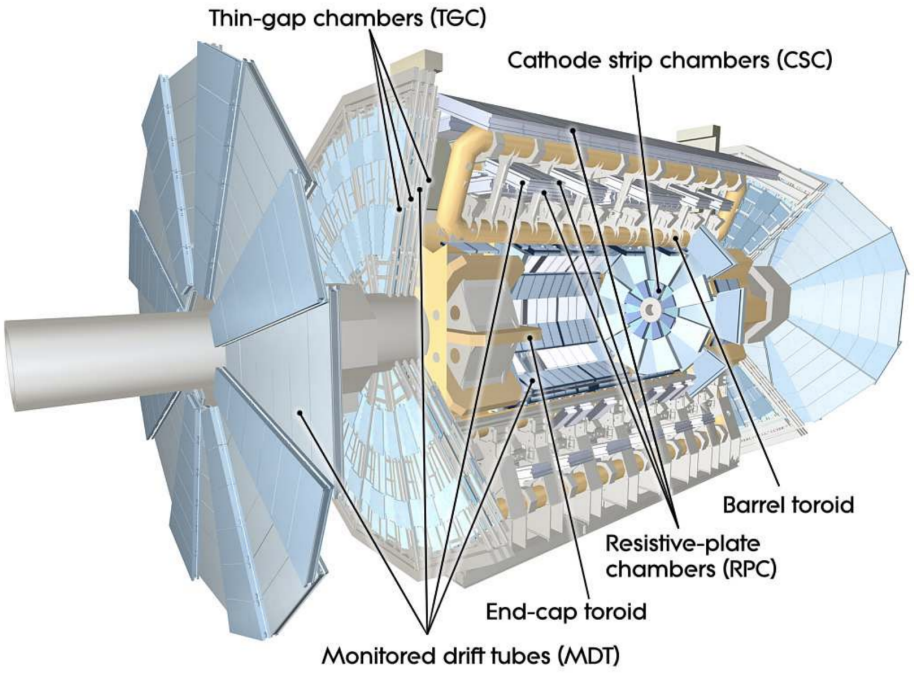
The hadronic calorimeter consists of the barrel, which covers $|\eta| < 1.7$, two end-caps which cover $1.5 < |\eta| < 3.2$, and two forward calorimeters which cover $3.1 < |\eta| < 4.9$. These three components consist of 3, 4, and 3 independent layers respectively. The barrel uses steel as the passive material and scintillating tiles as the active material; however, the end-caps and forward calorimeters use ionizing LAr as the active material, with copper (in the end-caps and the first layer of the forward calorimeter) and tungsten (in the second and third layers of the forward calorimeter) as the passive materials. The total thickness of the hadronic calorimeter is $\gtrsim 10\lambda$ over the entire detector. The smallest calorimeter cells in the barrel and end-caps are 0.1×0.1 in $\Delta\eta \times \Delta\phi$. In the forward calorimeters, η increases rapidly with θ, so the

[3]When not clear, *pile-up* refers to the contribution from both of these sources.

sizes are simply measured on an absolute scale, with the smallest cells 3.0×2.6 cm^2 in $\Delta x \times \Delta y$. In contrast to the electromagnetic calorimeter, the readout time for the tile cells is quite fast, and there is basically no contribution of out-of-time pile-up in the hadronic calorimeter [242].

5.2.3 Muon System

The outermost radial component of the detector is the muon system, or muon spectrometer (MS). In principle due to the interactions with the calorimeters the only particles that can make it out so far are muons; the MS provides measurements of muon tracks out to $|\eta| < 2.7$, with an additional triggering system that goes out to $|\eta| < 2.4$. However, very energetic hadrons can "punch through" to the MS; the energy of jets therefore not measured in the calorimeter can be corrected (Section 5.5.4.1). A cutout view of the MS can be seen in Figure 5.5.



(a)

Figure 5.5: A cutout view of the muon system showing the various subsystems. Figure sourced from [56].

The MS is surrounded by a toroid magnet system which bends muons in the $\pm z/\eta$ direction in the barrel. In the barrel, there are eight toroids arranged symmetrically about the beam axis; there are in addition two end-cap toroids.

The main component of the MS are the monitored drift tubes (MDT), which operate similarly to the TRTs in Section 5.2.1. These cover the region $|\eta| < 2.7$, with three layers out to $|\eta| < 2.0$ and two beyond that. As drift tubes, they provide a resolution of 35 μm in the z direction and no measurement in the ϕ direction; this choice of orientation is intended to measure the trajectory of the muons in the magnetic field and therefore their momentum. In addition, the drift time in the MDTs is large (approximately 700 ns) relative to the frequency of bunch crossings (25 ns), so that other systems must be used to trigger (Section 5.4) on muons.

Cathode-strip chambers (CSC) provide tracking measurements in the end-cap region $2.0 < |\eta| < 2.7$ with alternating layers of perpendicular strips. The CSCs are multi-wire proportional chambers which drift electrons from the inside to the outside of the chamber. Because of the perpendicular strips, the CSCs provide measurements in both directions, with a resolution of 40 μm $\times$ 5 mm in $R \times \phi$.

The muon triggering system is provided by the resistive plate chambers (RPC) in the barrel region $|\eta| < 1.05$, and by thin-gap chambers (TGC) in the end-cap region $1.05 < |\eta| < 2.4$. The RPCs are parallel plate capacitors filled with gas which are segmented in order to provide measurements in both directions, 10×10 mm^2 in $z \times \phi$. The drift time in the parallel plates is significantly less than in the MDTs, allowing for use in triggering. The TGCs are multi-wire proportional chambers similar to the CSCs, and provide resolution of about 5×5 mm^2 in $R \times \phi$.

5.3 Simulation

Every analysis in ATLAS relies on simulated events in some way. Both of the analyses presented in this Thesis (Chapter 6 and Chapter 7) are searches for new BSM physics; since BSM particles have never been observed, simulations are required to understand the sensitivity of the analysis to these new signals. Many analyses in ATLAS use simulations to model their background - both of these searches avoid this by

estimating the background in a data-driven way, but still require simulations of the background in order to set up and validate the analysis chain. Furthermore, a major step of object calibrations is understanding the effect of the detector in simulated events, which is the subject of Chapters 8 and 9. These are just some of the ways that simulations are used in ATLAS.

Simulations of physics events are *factorized*, often using different software programs entirely for matrix element calculations and parton showering, fragmentation and hadronization, and detector simulation [243]. Fundamentally each step of this process is random due to quantum effects, so events are simulated via Monte Carlo, or *MC*, methods to fully populate the relevant probability distributions. Since even the most common "interesting" physics processes have cross sections on the order of 10^{-5} of the total pp cross section, and many other important processes are much rarer still, events are typically simulated not inclusively but rather by first specifying the underlying *hard-scatter* (i.e., the tree-level pp interaction) process and proceeding from there.

The matrix element calculations cover the initial $2 \to N$ interactions of the constituent quarks in the colliding protons at the very high energies of the beams. The parton showering further simulates the produced partons as they radiate down to the hadronization scale around $\mathcal{O}(\text{GeV})$. Because the energies in these steps are above the hadronization scale, they can be calculated perturbatively at fixed order in α_s, and the generators are specified as to what order they go to (leading order "LO", next-to-leading order "NLO", etc.). The three most common generators in use are Pythia [244, 245], Sherpa [246], and Herwig [247, 248]. These generators can use Powheg-Box [249, 250] to interface between the matrix element and parton showering. Often a single generator will be used primarily for the simulated events in an analysis, and one or multiple other generators will be used to estimate the uncertainties related to the theoretical calculations.

The fragmentation and hadronization covers the conversion of partons to hadrons. This process is not described well by any fundamental theory, and so the generators use phenomenological models that are tuned to empirical data (e.g., [251]); similarly, generators are specified by this hadronization tune. The underlying event that comes

from the soft interactions of the proton constituents is also simulated at this stage, with the parton distribution functions again coming from phenomenological models [252]. These are also typically covered by the above-mentioned generators. The decays of heavy flavor hadrons (those containing bottom and charm quarks) are also simulated at this stage using dedicated programs like EvtGen [253].

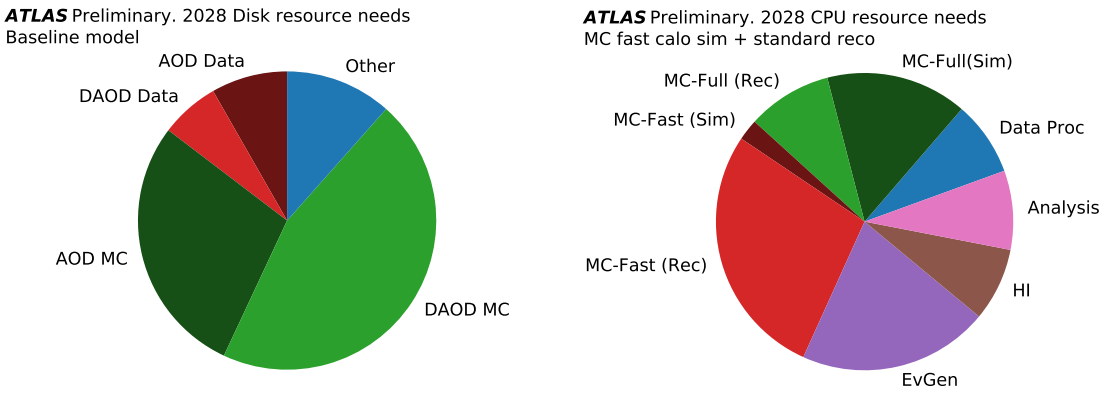
The other simultaneous pp collisions in the event (pile-up) are also simulated [254], but there are proposals to overlay minimum bias (i.e., minimal triggering requirements) data events on top of the simulated ones, according to the expected number of interactions in the event[4].

The above-mentioned steps constitute what is known as the "truth" event. The observables related to the particles at this stage (in particular, the four-momentum) are not accessible in real life, because they have to be observed in the detector (Section 5.2). The simulated events are passed through a full simulation of the ATLAS detector modeled in GEANT [255]. This simulates both the particle interactions with the material and the resulting digitization of the signals. This is considered to be the "reconstructed" event, which is in the exact same format that real data would be observed. The reconstructed event is passed through various post-processing stages (Section 5.5), the overall goal of which is to attempt to construct the "truth" event that would have given rise to the "reconstructed" (nonsimulated) event as well as possible.

The majority of ATLAS computing resources are dedicated to running simulations. Figure 5.6 shows the fraction of disk space and CPU time projected to be needed in 2028 for the various ATLAS computing activities. For the disk space, about 75% of the roughly 1500 petabytes needed by ATLAS by 2028 [256] will be dedicated to simulations ("MC" in the Figure). Similarly, for the CPU time, about 75% of the roughly 20 MHS06[5] needed by ATLAS by 2028 [256] will be dedicated to simulations ("MC" and "EvGen" in the Figure).

[4] In fact, this is what is done for heavy-ion collisions in ATLAS.

[5] A HS06 is a CPU benchmark tailored for high energy physics typical use cases [257]. Typical high-performance CPUs are equivalent to 500-1000 HS06. A MHS06 is a mega-HS06, or 1 million HS06.



(a) (b)

Figure 5.6: Projected computing resources needed by ATLAS in 2028. (a) Disk space. (b) CPU resources.Figures sourced from [256].

5.4 Trigger

The trigger system lies somewhere between the hardware and software systems[6]. Because of the enormous rate of collisions provided by the LHC (40 MHz), it is infeasible to store every event. Every event takes up $\mathcal{O}$(MB) [261] on disk, so storing every event would be equivalent to $\mathcal{O}$(100 TB) per second rate of writing, or about 10^6 PB for a full year's worth of data-taking[7]. Moreover, as discussed in Section 4.3 it is not desirable to store every event, because the vast majority of events are "uninteresting" and one of the main physics goals of the LHC and ATLAS is to observe or discover extremely rare processes.

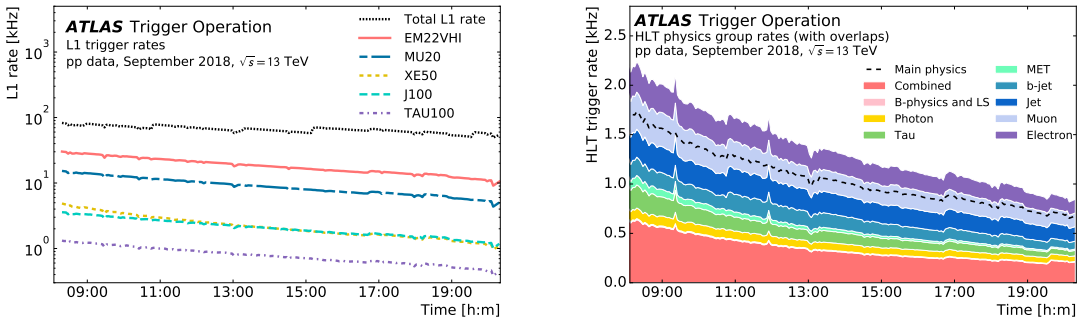
Because of the reasons outlined above, ATLAS institutes a *trigger* system [258] to decide whether or not to store each event, based on whether the reconstructed objects and topology in the event are indicative of some particular physics process. The reconstruction algorithms in use in ATLAS are distinguished by whether they are *online*, meaning they are in use in the trigger and are subject to the relevant time and space constraints, or if they are *offline*, meaning they are used on events passing

[6]This Section is sourced mainly from [258], with additional more recent information from [259] and [260] (which is not yet public at the time of writing).

[7]As a piece of trivia, this amount of data is called a *zetta*byte, or ZB [217].

the trigger and can be optimized for performance using the full detector information and much looser time constraints[8].

The trigger consists of two major subsystems. First, the hardware-based *Level 1* trigger, or *L1*, reduces the rate of events from 40 MHz to about 100 kHz based on fast but approximate algorithms. Then, the software-based *high-level trigger*, or *HLT*, uses offline-like algorithms (Section 5.5) to further reduce the rate to about 1 kHz. The rates of each of these two trigger levels, along with some of the main individual triggers, can be seen in Figure 5.7.



(a) (b)

Figure 5.7: Trigger rates for (a) L1 and (b) HLT or a typical run in September 2018. The overall rate (black dashed lines) is broken down into some main individual triggers. Figures sourced from [262].

The trigger broadly works by identifying detector objects that are likely to correspond to individual particles and thus indicate a hard-scatter pp interaction. The rate of the trigger is controlled by specifying some energy threshold the detector object has to meet in order to be passed to the next trigger in the case of L1 or written to disk in the case of HLT. The collection of trigger object definitions and corresponding energy thresholds is called the *trigger menu*, and is set every year of data taking (with some minor changes between periods in the same year) [263–266] according to the instantaneous luminosity and pile-up conditions and physics objectives.

Trigger considerations are vital for any analysis, because if the physics being targeted by an analysis does not pass any trigger then the data are simply not stored

[8]Subject to the CPU resources available to the offline analyzer.

and the analysis cannot be done. For example, the low-mass extension of the search in Chapter 6 discussed in Appendix B spends considerable effort on finding a suitable trigger for the new signature in that search; luckily, there is an existing trigger (intended for other uses) which does have efficiency on those new signals. However, there are a few different kinds of triggers on the trigger menu which can enable analyses in light of the trigger limitations other than the *primary* triggers, which store the entire event every time the trigger is fired. Triggers can be *prescaled*, meaning there is some probability < 1 that the trigger fires, which allows for lower energy thresholds - these can be used for background or efficiency studies or simply for monitoring the beam, detector, or trigger systems. Also, there are situations where the energy threshold is lowered but only small fraction of the total information in the event is stored - these can be used for detector calibration or to directly conduct *trigger-level* analyses (e.g. [267]) to be sensitive to kinematic regions that would otherwise be inaccessible with the primary triggers.

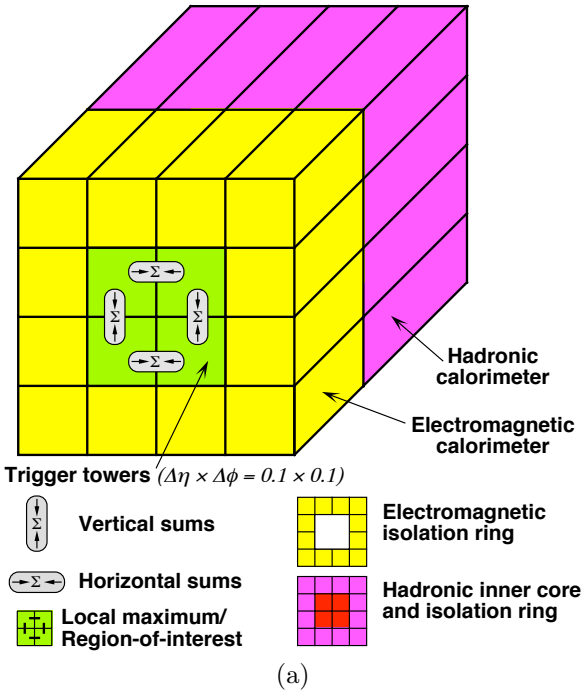
There are also special runs with entirely different trigger menus than the regular physics menu. For example, a *minimum-bias* trigger is run as the only item on the trigger menu for a short period of time (i.e., trigger on every event with some minimum quality controls with some high prescale). These minimum-bias events are used, among other things, for measuring online trigger efficiencies offline and for overlaying pile-up events on top of simulations (Section 5.3).

5.4.1 Level 1 Trigger

The L1 trigger has two major subsystems, one based on observations in the calorimeter (L1Calo) and one based on observations in the muon system (L1Muon). There is also a topological trigger (L1Topo) which combines information from the two subsystems to make trigger decisions on a whole-event basis. In particular, there is not enough time at L1 to reconstruct tracks from hits in the tracker.

L1Calo sets up 0.1×0.1 in $\Delta\eta \times \Delta\phi$ trigger towers in the calorimeter and forms *regions of interest*, or *RoIs*, to identify physics objects. The electron/photon and tau triggers use 2×2 groupings of trigger towers in the EM calorimeter that are

local maxima in energy, with additional isolation requirements in the surrounding towers and in the hadronic calorimeter. The isolation requirements are intended to distinguish between isolated photons and collimated $\pi^0 \rightarrow \gamma\gamma$ decays. The jet triggers use 4×4 and 8×8 groupings of trigger towers in the EM and hadronic calorimeters with the central 2×2 grouping a local maximum in energy. A schematic of the L1Calo trigger towers and RoIs can be seen in Figure 5.8.



(a)

Figure 5.8: Schematic view of the trigger towers used as input to the L1Calo trigger algorithms. Figure sourced from [258].

The L1Muon system uses the information from the RPCs and TGCs (Section 5.2.3) to identify muon candidates in time to pass on to the HLT.

There is finally an L1Topo system which combines information from L1Calo and L1Muon to evaluate event-wide information, in particular the *missing transverse energy* or E_T^{miss}, which corresponds to imbalances of momentum in the transverse plane (Section 5.5.6).

5.4.2 High Level Trigger

As mentioned above, the HLT uses offline-like reconstruction algorithms to target a wide range of more specific physics goals than the L1. In particular, tracking information from hits in the tracker, finer-granularity information in the calorimeter, and more precise measurements in the MS allow more precise reconstruction than is available at L1.

The tracking algorithms first run a fast tracking algorithm with rough pattern recognition[9] within RoIs identified at L1. These fast tracks are then used as seeds for an offline-like (Section 5.5.1) track reconstruction algorithm.

Calorimeter reconstruction algorithms are used to identify electrons, photons, taus, and jet candidates and finally E_T^{miss} global reconstruction. The first step is to construct clusters of energy in the calorimeter based on RoI seeds from L1, for which there are two different algorithms.

The electron and photon reconstruction algorithms use a sliding-window approach by finding a window of size 0.075×0.175 in $\Delta\eta \times \Delta\phi$ that is a local maximum of energy in projective towers. The shower shape is then found by determining layer-by-layer in the calorimeter the center of energy of the cells behind the sliding window and summing up a fixed size window around that. This shower shape is then used to positively identify electrons and photons, e.g. using the ratio of energy deposited in the hadronic calorimeter to the electromagnetic calorimeter.

The tau, jet, and E_T^{miss} reconstruction algorithms use a global topo-clustering algorithm [173] very similar to the offline reconstruction algorithm (Section 5.5.2), which are built up iteratively from high-energy cell seeds. Jets are then reconstructed using the anti-k_t algorithm with $R = 0.4$ or $R = 1.0$ (Section 3.2) using these clusters as seeds. The jets are calibrated in a similar manner to the offline procedure (Section 5.5.4.1). Small-R jets are calibrated in particular incorporating a pile-up subtraction step, an overall MC-based p_T correction (the exploration of which is the subject of Chapter 8), and the global sequential calibration (GSC) (the improvement

[9] Not to be confused with the similarly-named but ultimately shuttered *Fast TracKer* [268], or FTK, which was a project to implement tracking directly into the hardware in order to speed up performance.

of which is the subject of Chapter 9). Large-R jets in 2017 and 2018 are trimmed and have a mass cut, but do not use any other jet substructure techniques (Section 3.4) for identifying boosted massive objects.

Tau identification uses similar principles to the offline selection (Section 5.5.3), looking for small-R jets with narrow calorimeter energy deposits and low numbers of associated tracks.

Jets originating from b-quarks can also be tagged (Section 5.5.4.3), by identifying secondary vertices off the beamline that are characteristic of such jets. In order to judge if a secondary vertex is significantly far enough from the beamline, it is essential to have a precise measurement of the *beamspot*, which is the 3-dimensional ellipsoidal region in which pp collisions can occur[10] [269]. The Author has contributed to a project intended to more precisely measure the online beamspot using Bayesian inference [270], which could improve the b-tagging at HLT.

As jets with no such tagging are the most generic detector objects formed in ATLAS, the rate of jet production is very high - e.g., the cross section of producing a jet with $E_T > 100$ GeV is at least a factor of 10 higher than any other "interesting" electroweak process (see Figure 4.4). Therefore, the energy thresholds for jets in the trigger are the highest of any other object, and analyses using only jet triggers are therefore limited to very high energies. For example, the search in Chapter 7 uses a single jet trigger, corresponding to offline $p_T > 500$ GeV; this limits the minimum dijet invariant mass $m_{JJ} > 1.1$ TeV just to meet the trigger requirements. Because of these high requirements, there have been concerted efforts for trigger-level analyses using lower p_T jets [267].

There are a variety of different trigger algorithms for E_T^{miss}, but the most basic algorithm sums up the vectorial $\vec{p_T}$ of all cells in the calorimeter to find any imbalance. Other algorithms build on this by adding pile-up corrections or jet-based or topo-cluster-based associations.

The muon reconstruction algorithm first runs a fast algorithm which then seeds RoIs to be used in a precision step similar to the offline algorithm (Section 5.5.5). The

[10] With roughly a multivariate Gaussian probability distribution.

fast algorithm matches the RoIs identified at L1Muon to data from the MDT chambers (Section 5.2.3) to form tracks in the MS. These tracks are then back-extrapolated back to the interaction point and combined with tracks in the inner detector. The precision step basically repeats this process with a refined track-finding and p_T measurement, with an additional step of extrapolating tracks from the inner detector to hits in the MS in case the back-extrapolation fails. Dimuon triggers (with dimuon mass requirements) are also used for tagging resonant hadrons, e.g. $J/\psi \to \mu\mu$.

Finally, there are dedicated triggers with low rates for exotic signatures, e.g. long-lived particles [271].

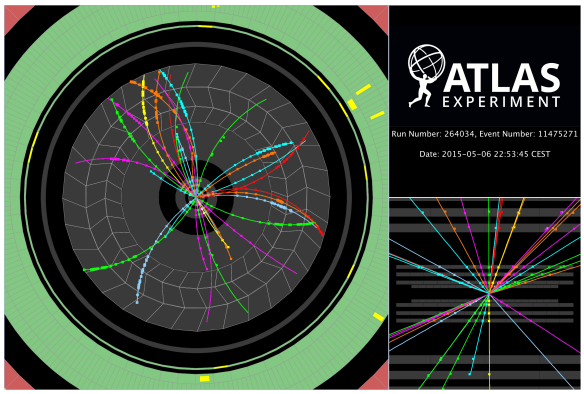
5.5 Object Reconstruction

Events passing the triggers (Section 5.4) are written to disk and the detector observables are *reconstructed* with algorithms into objects intended to correspond to particular physics particles. Tracks derived from hits in the inner detector (Section 5.5.1) and topological clusters of cells in the calorimeter (Section 5.5.2) are used to build up objects corresponding to photons and electrons (Section 5.5.3) and jets (Section 5.5.4). Muons are also reconstructed using information from the MS (Section 5.5.5). Finally, these objects are combined together to estimate the missing transverse energy (Section 5.5.6).

5.5.1 Tracks

Tracks, corresponding to the paths of charged particles in the inner detector, are reconstructed from hits in the various layers of the tracker [238, 272]. An example event display showing hits in the inner detector and the tracks formed from those hits can be seen in Figure 5.9.

Hits in the pixel (including the IBL) and SCT layers are clustered in order to account for a single particle passing through multiple adjacent pixels or multiple particles passing through the same or adjacent pixels; these clusters are then abstracted as three-dimensional *space-points*. Track seeds are then formed from sets of



(a)

Figure 5.9: Event display showing hits in the inner detector and the tracks formed from those hits. Figure source [273].

three space-points. A combinatorial Kalman filter [274] is then used to incorporate additional space-points from other layers of the pixel and SCT detectors which are compatible with the particle trajectory into a track candidate. An ambiguity solver is then used to identify clusters with track candidates, using a neural network to identify and predict merged clusters resulting from multiple tracks [275]. The clusters are identified with the track candidates in order of quality of the track, considering criteria like the number of holes (missing clusters) across the layers and χ^2 of the track fit. Track candidates are also rejected if they do not meet basic requirements like $p_T > 400$ MeV (in which case the particle's trajectory would curve so much in the magnetic field it would never leave the inner detector)[11]. Finally, a track fit is performed using all available information (in particular including information from the TRTs) to determine the track parameters. The radius of curvature of the track, when combined with the known magnetic field, gives the charge to momentum ratio q/p (Equation 4.1), and the direction of curvature gives the sign of the charge. The other relevant track parameters are the direction in η, ϕ after exiting the tracker (after which the expected trajectory of the charged particle is straight), transverse

[11] With the $B \approx 2$T magnetic field, the radius of curvature has to be $R \gtrsim 0.5$m in order to escape the ID at a distance of ~ 1m from the beamline. Therefore the momentum has to be $p_T \gtrsim 300$ MeV (Equation 4.1), with $q = \pm 1$.

impact parameter off the beamline d_0 (i.e., the closest approach of the track trajectory to the beamline), and longitudinal displacement along the beamline z_0 (at the point of closest approach). The tracking reconstruction is particularly difficult within the dense environments of jets, where there can be large numbers of tracks in a small region.

The resolution of track q/p, d_0, and z_0 as a function of p_T can be seen in Figure 5.10. The momentum resolution is quite good, around $1\% \oplus 0.05\% * p_T/\text{GeV}$ -

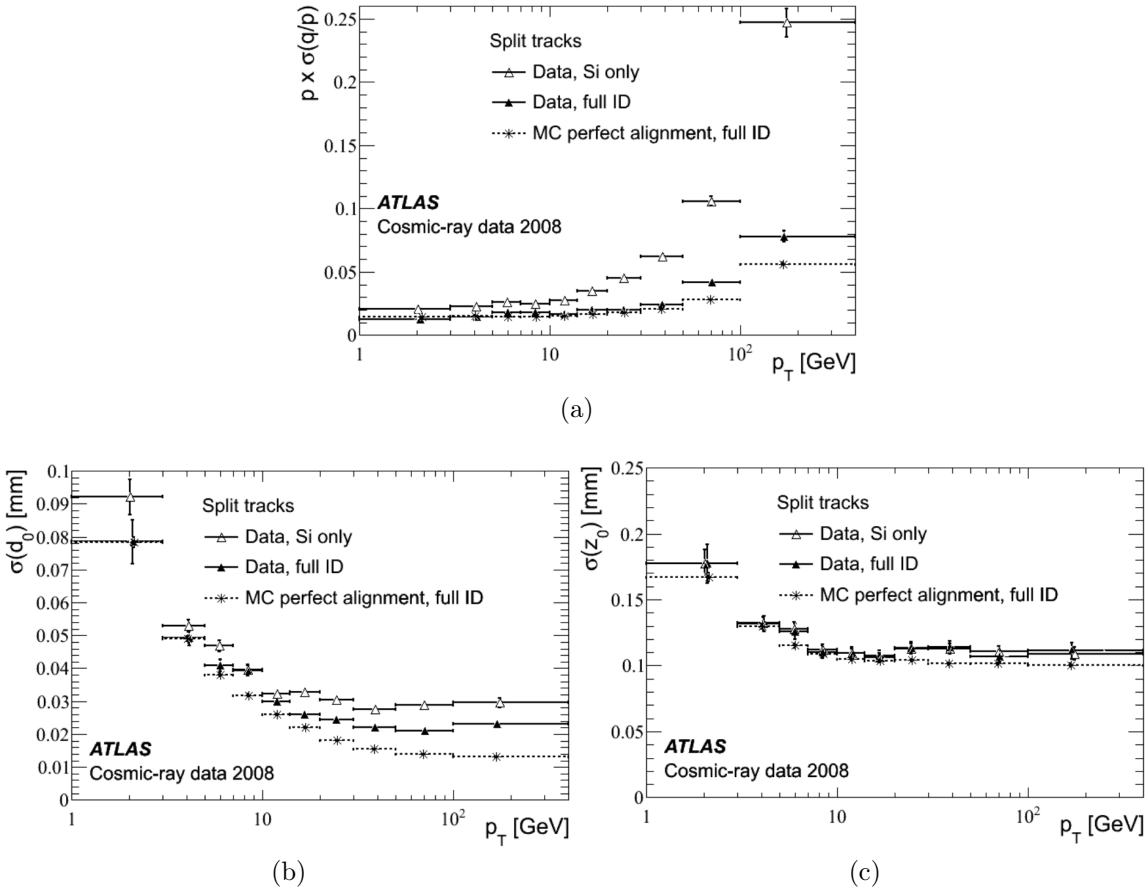


Figure 5.10: Track resolution of (a) q/p, (b) d_0, and (c) z_0 as a function of p_T in cosmic-ray data. The measurements are compared among tracks measured with silicon-only components, full inner detector, and full inner detector in simulation. Figures sourced from [276].

the momentum resolution gets worse at high momentum as the track has less curvature in the magnetic field. The d_0 resolution is important for constructing decays off the beamline, which is essential for b-tagging (Section 5.5.4.3). The z_0 resolution is important for constructing primary vertices which correspond to interactions along the beamline (Section 5.5.1.1).

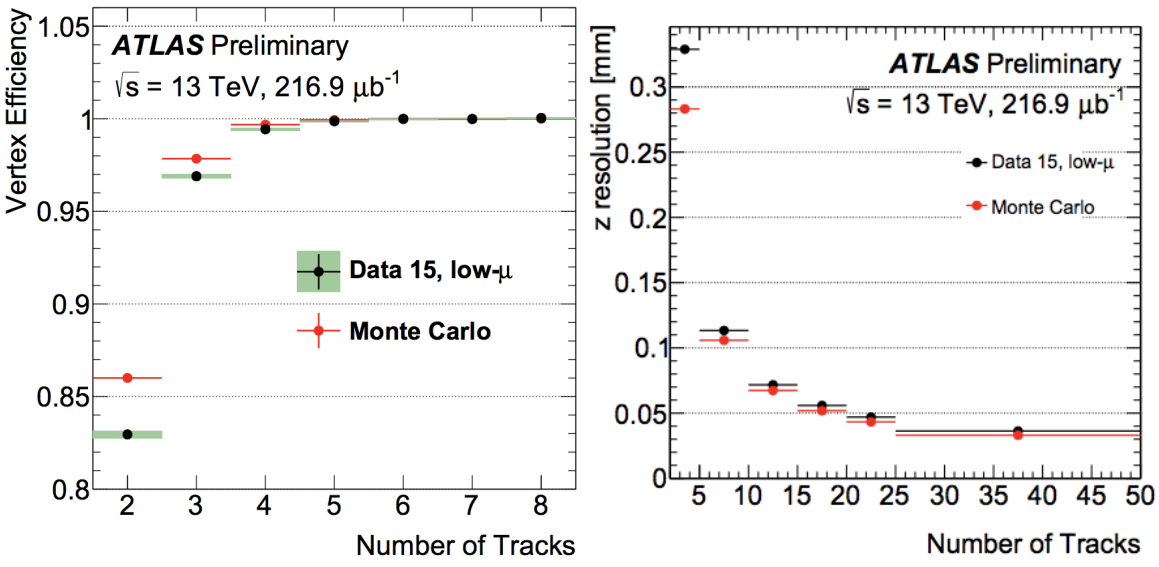
5.5.1.1 Primary Vertices

Primary vertices [277–279] are collections of tracks emanating from a single point indicating a pp interaction at that point. The average number of interactions per bunch crossing μ is > 1, which is directly related to the instantaneous luminosity; $\mu > 1$ is the cause of pile-up interactions in the event other than the *hard-scatter* interaction or (roughly) the one that caused the event to pass the trigger. In the ideal case the expected number of reconstructed primary vertices would scale linearly with μ; however, the dominant cause of nonlinearity is vertex merging due to the high density of interactions in the luminous region. There can also be vertex splitting, vertex fakes, and inefficiencies in vertex reconstruction.

The vertex reconstruction algorithm proceeds iteratively - after choosing a vertex seed, the algorithm alternates between finding tracks compatible with coming that vertex (subject to their z_0 and d_0 resolutions) and re-fitting the vertex position with the associated tracks until some stopping condition is met (vertices must have at least 2 associated tracks). The found vertex and its associated tracks are then removed from consideration and the process is repeated until all tracks are associated to a vertex or no more vertices can be found. The output of this algorithm is a set of vertex positions and the covariance matrix of their resolutions.

The efficiency of vertex reconstruction as a function of number of associated tracks can be seen in Figure 5.11a. The vertex reconstruction efficiency is basically 1 when the vertex has 5 or more associated tracks. Figure 5.11b shows the z resolution of the primary vertices. The z resolution is important for distinguishing between different interactions along the beamline.

The primary vertex with the highest $\sum p_T^2$ over the associated tracks is designated



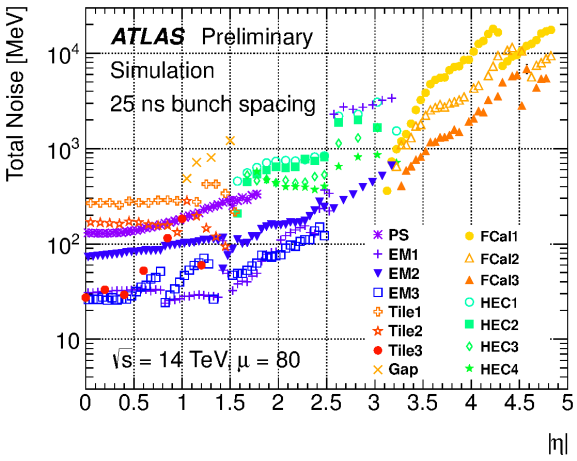
(a) (b)

Figure 5.11: (a) Efficiency of vertex reconstruction as a function of number of associated tracks. Also, resolution of vertices in (b) x, (c) y, and (d) z. Measured in low-μ ($\mu \sim 0.01$) data and in MC. Figures sourced from [278].

as the *hard-scatter* vertex (and the associated tracks are considered to be the hard-scatter tracks), as it corresponds to the hardest interaction in the event and therefore the most likely to be "interesting". The remaining vertices and their associated tracks are labeled *pile-up*. For most reconstruction algorithms considered below, if they use tracks, only the hard-scatter tracks are considered. However, the pile-up tracks are used for the removal of calorimeter jets originating from pile-up interactions (Section 5.5.4.2), as the calorimeter by itself does not have nearly good enough angular information to distinguish between energy clusters due to hard-scatter and pile-up interactions. Also, tracks with large impact parameters, but within some small distance of the hard-scatter primary vertex, are used to reconstruct secondary vertices for b-tagging (Section 5.5.4.3).

5.5.2 Clusters

Energy deposits in calorimeter cells are combined together using a topological clustering algorithm [173], which are then used as seeds in jet-finding. The expected amount of noise in a given cell $\sigma_{\text{noise,cell}}$ is estimated in order to define an energy significance $\zeta_{\text{cell}} = E_{\text{cell}}/\sigma_{\text{noise,cell}}$. The noise in a given cell is due both to electronic noise and to soft energy deposits from the pile-up interactions in the event; thus, the cell noise is estimated as a function of the position in the detector (layer of the calorimeter and η) and the amount of pile-up μ. This estimate can be seen in Figure 5.12.



(a)

Figure 5.12: Expected total noise per cell (in simulation) as a function of position in the detector (layer and η) at $\mu = 80$, corresponding to conditions similar to those in 2018 data-taking. Figure sourced from [280].

The algorithm for topological clustering proceeds as follows. First, seeds are formed as cells with energy deposits $\zeta_{\text{cell}} > 4$. The cluster then grows from the seed to all neighboring cells (either adjacent if in the same layer or close in η, ϕ if in different layers) with energy deposits $\zeta_{\text{cell}} > 2$; this growth continues iteratively until there are no more cells passing the threshold to spread to. Finally, the cluster spread to all immediately neighboring cells with $\zeta_{\text{cell}} > 0$[12]. Clusters can then be split if there are local maxima of energy within the cluster. The result of the clustering algorithm

[12]Note that the significance can be < 0 due to pulse-shape calibrations in the cell.

is a set of 3-dimensional irregularly-shaped but topologically-connected clusters. The progression of this algorithm can be seen in Figure 5.13.

Each cluster is assigned an (η, ϕ) position based on the barycenter of the constituent cells, and a total energy E. Clusters are assumed to be massless when forming their four-momentum.

The electromagnetic showers (due to electrons and photons) have a different response in the calorimeter than hadronic showers due to the different kinds of interactions with the detector at play. Because of this, there is a desire to calibrate individual clusters to a consistent scale. This calibration does not have to be applied - in the case that it is not applied, the clusters and the jets formed from them are considered to be at the *EM scale*. For example, the small-R jets used in the search presented in Chapter 6 use jets at the EM scale. However, large-R jets usually do use calibrated clusters, as they often take advantage of substructure information in their constituent clusters, which benefits from the clusters being at a consistent energy scale. The large-R jets used in the search presented in Chapter 7 use jets formed from these calibrated clusters.

The calibration is a *local cluster weighting*, or *LCW*. First, the entire cluster is classified according to the likelihood it comes from an electromagnetic or hadronic shower based on its position in the detector and various cluster moments. Then, each cell in the cluster is corrected as a weighted average between a hadronic correction and an electromagnetic correction according to this likelihood. The hadronic corrections are derived from simulations of single-particle charged pions interacting with the detector, and the electromagnetic corrections are derived from simulations of single-particle neutral pions ($\rightarrow \gamma\gamma$). The per-cell hadronic correction consists of three components: a correction for the difference between hadronic and electromagnetic sources; a correction for out-of-cluster cells; and a correction accounting for dead material in front of the calorimeters. The per-cell electromagnetic correction consists of the second two components in the hadronic correction (out-of-cluster and dead material effects). The entire cluster four-momentum is then recalculated using the weighted cells. This calibration process is outlined in Figure 5.14.

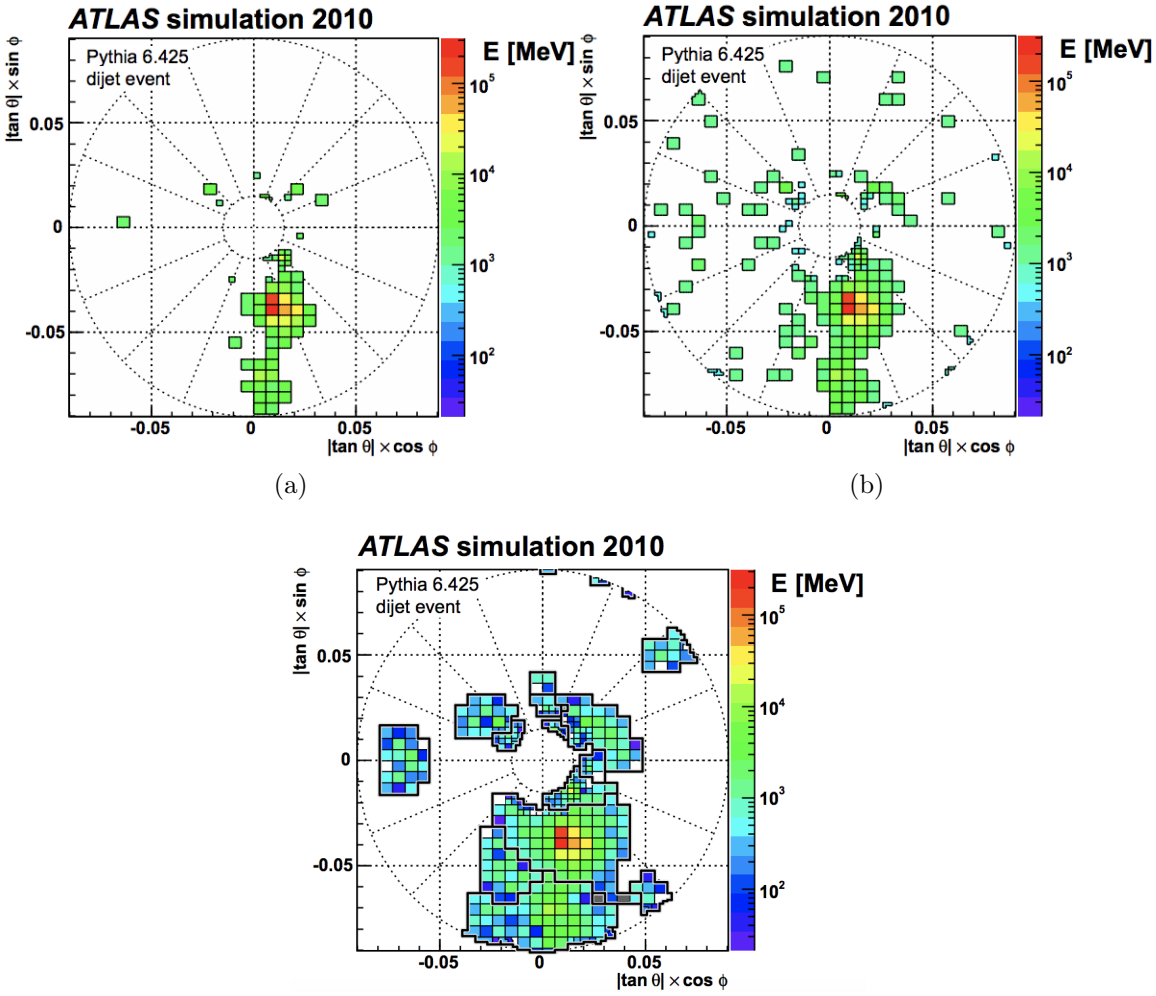
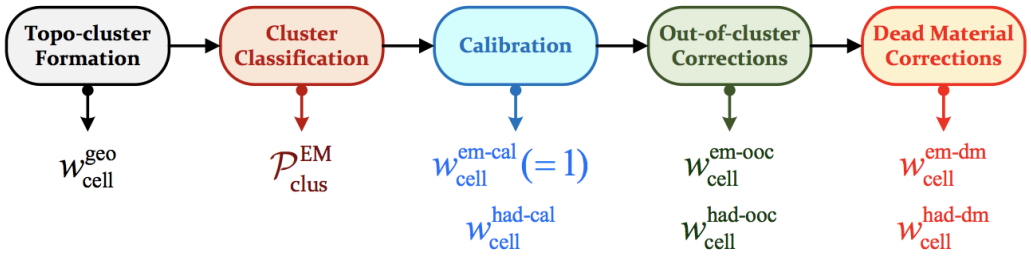


Figure 5.13: Progression of the topological cell clustering algorithm in a single layer of the calorimeter. (a) Cell seeds passing $\zeta_{\text{cell}} > 4$. (b) Growth of clusters to cells passing $\zeta_{\text{cell}} > 2$. (c) Final clusters including cells with $\zeta_{\text{cell}} > 0$ and cluster splitting. Figures sourced from [173].



(a)

Figure 5.14: Steps in LCW cluster calibration scheme. After the topo-clustering formation and splitting, clusters are classified based on the likelihood of coming from a hadronic or electromagnetic shower. This likelihood is used as the weighting between a hadronic and electromagnetic correction. The correction includes a correction for differences between hadronic and electromagnetic showers (for which the electromagnetic correction is 1), out-of-cluster effects, and dead material. Figure sourced from [173].

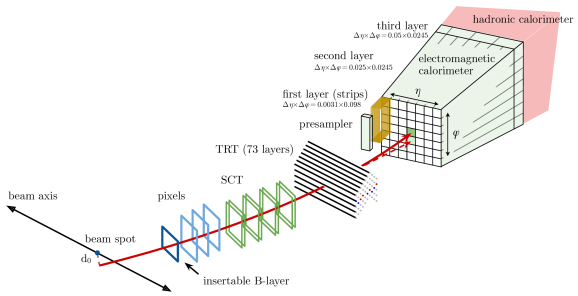
5.5.3 Photons and Electrons

The reconstruction of photons and of electrons [181] are quite similar to each other. This is due to the fact that, as mentioned in Section 5.2.2, electrons and positrons radiate bremsstrahlung photons when interacting with the detector, and photons in turn often *convert* into electron/positron pairs - so the electromagnetic showers produced by electrons and photons in the calorimeter are very similar. A schematic of an electron passing through the detector can be seen in Figure 5.15, showing in addition a radiated bremmstrahlung photon[13].

Electrons and photons are distinguished mostly by tracking information - electrons have a matched track emanating from the primary vertex, while photons either have no associated tracks (if unconverted) or a set of paired tracks emanating from a secondary conversion vertex (if converted).

The reconstruction begins with topoclusters (Section 5.5.2), with a preselection that $> 50\%$ of the topocluster energy is in the electromagnetic calorimeter. Tracks

[13]Without the dashed line, this is also just a generic path of a charged particle through the detector.

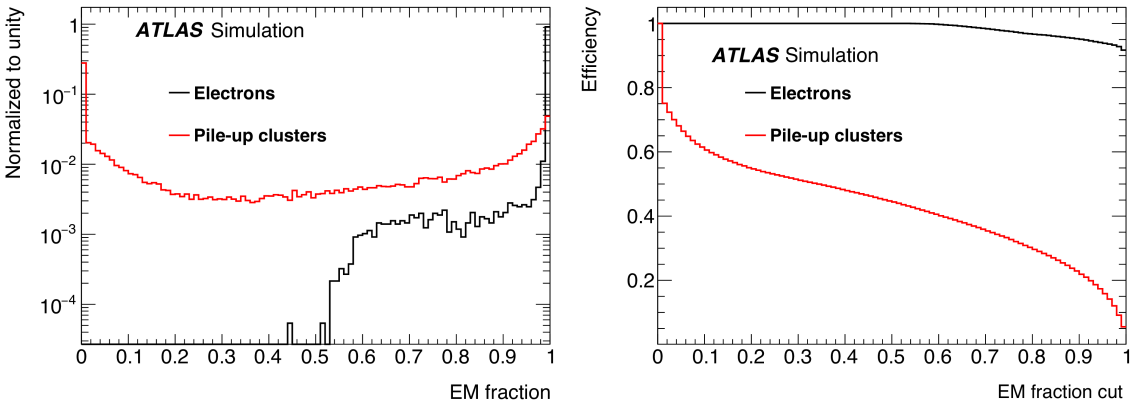


(a)

Figure 5.15: Schematic of the path of an electron through the ATLAS detector. The red solid line can be reconstructed as a track in the inner detector, and the red dashed line shows the path of a radiated bremsstrahlung photon due to material interactions in the inner detector. Figure sourced from [281].

are then extrapolated from the tracker into the calorimeter and matched to topoclusters in η, ϕ. As mentioned above, matched tracks are then determined to come from the primary vertex or combined with other nearby tracks to find secondary vertices indicating a converted photon. The topoclusters are then used as seeds to expand into *superclusters* intended to capture the full electromagnetic shower. Topoclusters with $E_T > 1$ GeV and a matched track, or topoclusters with $E_T > 1.5$ GeV and no requirement on matched tracks are used as superclusters seeds. In order of seed E_T, the seeds are expanded to all neighboring topoclusters within $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$. If there is a matched primary vertex track to the seed then any other neighboring clusters matched to the same track are also added; and if there is a matched conversion vertex to the seed then any other neighboring clusters with tracks matching the conversion vertex are added. Finally, tracks are refit to the superclusters, and they are labeled unambiguous electrons if there is a matched primary vertex track, unambiguous photons if there are no matched tracks or a set of matched tracks forming a conversion vertex, or ambiguous if both are true.

A variety of features related to the shower shape are used for electron and photon identification and discrimination from hadronic showers. For example, the fraction of energy of the supercluster in the electromagnetic calorimeter is used as a discriminatory variable; the distribution of this feature and discrimination power in electrons and in pile-up clusters (mostly hadronic) can be seen in Figure 5.16. These features



(a) (b)

Figure 5.16: Fraction of supercluster energy in electromagnetic calorimeter, compared between true electrons and pile-up, in simulation. (a) Distribution; (b) Efficiency of cut. Figure sourced from [181].

(formed separately for electrons and photons) are combined together into a multivariate likelihood in bins of E_T and η to be used in the final identification. More information about the electron identification can be found in [281]. The shower shape variables are particularly useful for distinguishing between prompt photons and neutral pions decaying to collimated $\gamma\gamma$. Both photons and electrons have working points at `Loose`, `Medium`, or `Tight`, which can be chosen at the analysis level based on the signal efficiency - background rejection trade-off.

In addition, a set of isolation variables are defined based on calorimeter and tracking activity outside the supercluster but inside a cone of $\Delta R = 0.2$. The isolation is intended to discriminate between prompt electrons and electrons from heavy flavor decays or light hadrons misidentified as electrons, and further between prompt photons and neutral pions decaying to collimated $\gamma\gamma$. Isolation working points are defined at `Loose`, `Tight`, and `CaloOnly` which uses only calorimeter information.

The efficiency of the identification and isolation working points are evaluated in data using high-fidelity $Z \to ee$ and $J/\psi \to ee$ events for electrons [282] and $Z \to ee\gamma$ and $Z \to \mu\mu\gamma$ events for photons [283, 284]. The differences between the simulation and data are corrected, and the uncertainties on these corrections are applied as

systematic uncertainties on simulated events.

The energy of electrons and photons are calibrated [285, 286] to account for energy lost in the material upstream of the calorimeter, for energy deposited in the cells neighbouring the cluster, and for energy lost beyond the electromagnetic calorimeter. The corrections are first derived in simulation using a boosted decision tree (BDT) incorporating multiple shower shape variables in bins of η and E_T. The photon and energy calibrations do not use numerical inversion (Chapter 8), meaning the calibrations can be biased by the prior used to calibrate. However, as the energy resolution is small ($\mathcal{O}(5\%)$ for electrons and photons with $5 < E_T < 10$ GeV, and better for higher energies), this effect is not as important as it would be for jets, which have much worse resolutions (Section 5.5.4.1).

The calibrated energies are evaluated in data using the above-mentioned high-fidelity samples. A data-simulation, or *in situ*, correction is applied based on any observed differences. The uncertainties on these corrections are applied as systematic uncertainties on simulated events.

5.5.4 Jets

There are a variety of jet collections in use in ATLAS. The most common are those formed using topoclusters (Section 5.5.2) as seeds, with $R = 0.4$ (small-R jets) using topoclusters at the EM scale and $R = 1.0$ (large-R jets) using topoclusters at the LCW scale. Large-R jets are typically trimmed as described in Section 3.3.

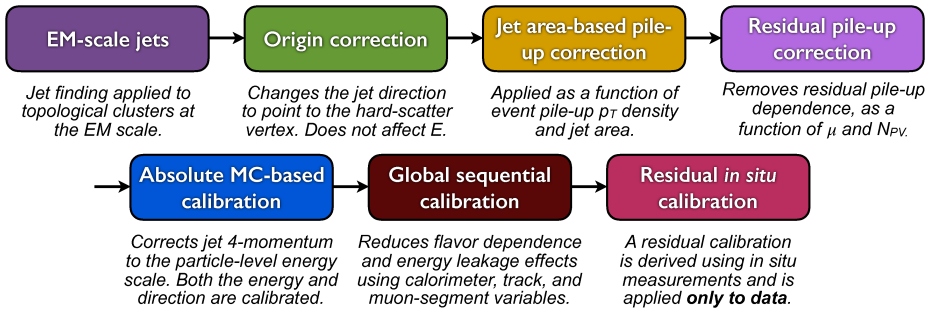
Jets can also be formed using particle flow objects [174] which combine information from the calorimeter and tracker; Track-CaloClusters [287] which similarly combine information from the calorimeter and tracker; or reclustering from smaller-radius jets [76]. Topoclusters themselves can also be corrected other than the LCW calibration - the Author has contributed to a project [187] with the idea of subtracting pile-up from topoclusters before running the jet algorithm.

However, even if the seeds are in principle themselves corrected or calibrated, there is still a need to calibrate the jets formed from those seeds to further account for pile-up, out-of-cone, and nonlinearity effects.

5.5.4.1 Jet Calibration

The jet energy and momentum calibration procedure [80–82] is achieved through a series of steps.

The steps for the calibration of small-R jets are shown in Figure 5.17. Following



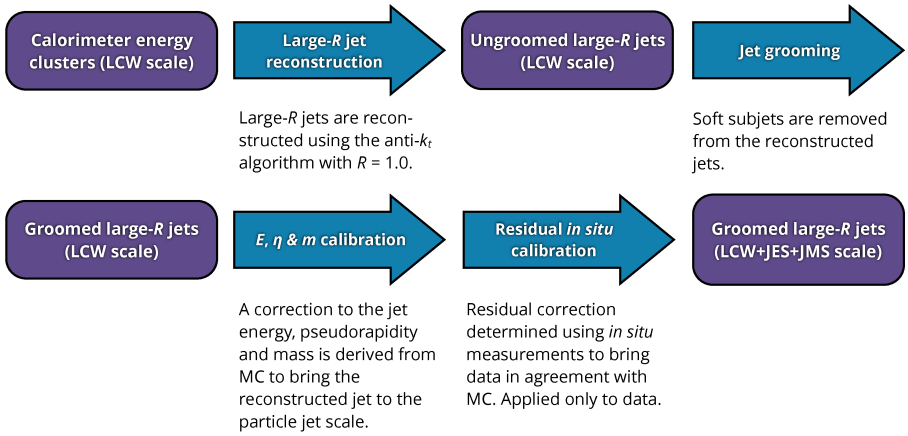
(a)

Figure 5.17: Steps in the small-R jet calibration. Figure sourced from [80].

jet reconstruction from the calorimeter cell-clusters, there is a minor correction to the jet direction to point to the hard-scatter primary vertex rather than the nominal detector center. The impact of pile-up is corrected for using a jet area-based [83, 84] approach followed by a residual correction sensitive to both in-time and out-of-time pile-up [85]. Following the pile-up correction, an absolute p_T scale correction is derived in simulation, which also corrects the jet direction. There is then a sequence of corrections, called the *global sequential calibration* or *GSC*, to further correct the residual dependence of the jet p_T on various jet quantities using information from the tracker, calorimeter, and MS. The final step of the jet calibration procedure applied only to data is an in-situ correction that accounts for the residual differences between data and simulation.

The steps for the calibration of large-R jets are shown in Figure 5.18. After jet reconstruction with LCW topoclusters and grooming with trimming, the jet energy, mass, and η are corrected using an absolute MC-based correction. There is finally an in-situ correction applied only to data to correct for differences in the energy scale between data and simulation.

Each of the above steps, other than the pile-up correction for small-R jets, operate



(a)

Figure 5.18: Steps in the large-R jet calibration. Figure sourced from [82].

under the principle of *numerical inversion* (Chapter 8) in order to be independent of the underlying prior distribution of the reference energy or momentum.

In the simulation-based corrections, the reference object is "truth" jets formed from detector-stable simulated particles ($c\tau > 10$ mm) other than muons and neutrinos, absent the detector simulation (Section 5.3). Reconstructed jets are geometrically matched to truth jets with $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Truth jets are matched to truth partons using ghost association [84]; the type of the highest energy parton matched to a truth jet is used as the label.

In the in situ corrections, the goal is to correct any differences in the energy scale between data and simulation. The relative p_T response of the jet to the reference object is compared between data and simulation and this relative ratio is corrected for. For lower p_T jets, the in situ corrections use $Z \to ee/\mu\mu$ or γ as the reference object. For higher p_T jets, systems of multiple lower p_T small-R jets are used as the reference objects.

The first step of the small-R jet correction is simply a geometric correction to account for the fact that η is measured as an angle from the center of the detector, but jet angle should instead be measured from the interaction point or the hard-scatter primary vertex.

The pile-up correction for small-R jets [83] first estimates the average extra p_T

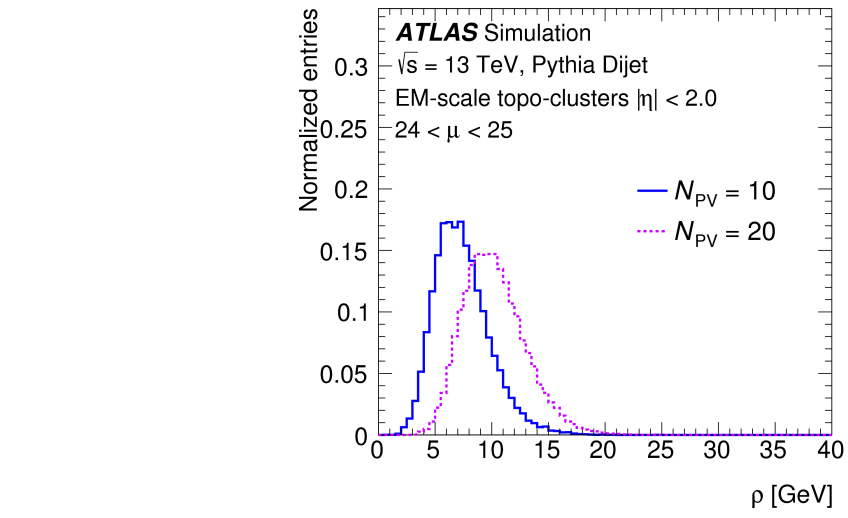
density due to pile-up in the event ρ, and the catchment area of the jet A; then ρA is subtracted from the jet p_T as the expectation of pile-up contribution in the jet. The areas of jets [84] are estimated using *ghost association*. The whole event is blanketed uniformly in η, ϕ with zero-momentum *ghosts*, and the jet finding process is repeated. The set of output jets is not affected by the presence of ghosts, since the algorithm is IRC safe (Section 3.2), but the set of ghosts associated with each jet gives a definition of the jet area. ρ is calculated by rerunning the jet finding using the k_t algorithm with $R = 0.4$ (most of which will be due to pile-up interactions), calculating the jet momentum density p_T/A, and taking the median over all jets. The distribution of ρ at different values of the number of primary vertices (NPV), an estimate of the pile-up contribution in the event, can be seen in Figure 5.19. It can be seen that the peak of the ρ distribution increases roughly linearly with NPV.

The Author has been involved in a project [187] to subtract pile-up from topoclusters in the first place rather than from the entire jet. In this project, ρ is calculated in the same way, but the area of each topocluster is estimated using a Voronoi tessellation [288–290] of the calorimeter in η, ϕ with the topoclusters as seeds.

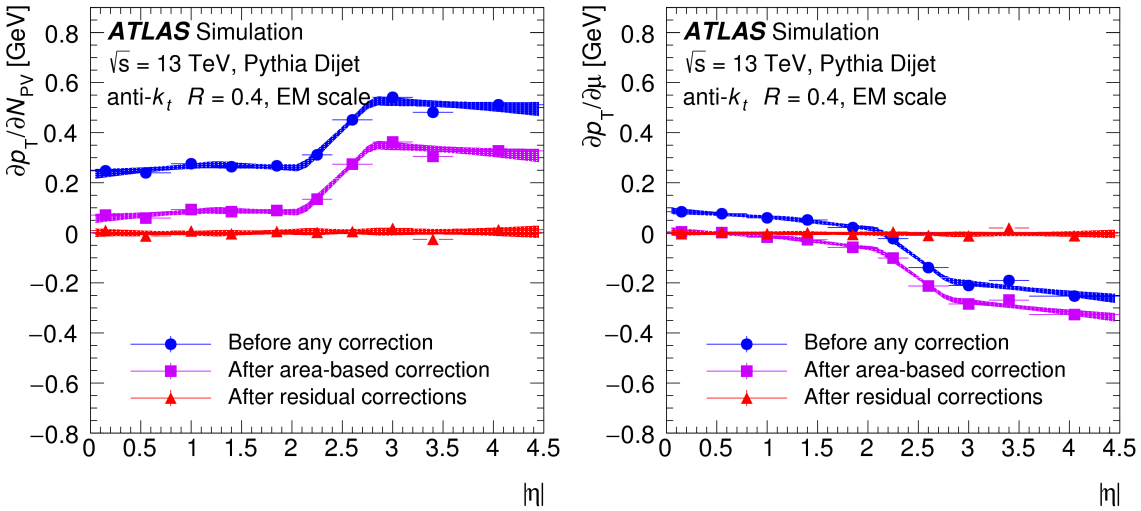
After the ρA correction, there is still some residual dependence on NPV and μ, particularly at high η. The dependence of the jet p_T on these two parameters is estimated with a linear fit and subtracted; the effect of this correction can be seen in Figure 5.19.

Following the pile-up correction, a correction is derived in simulation for the ratio between the overall reconstructed jet energy and the truth jet energy. In bins of η, the mode of the E^{reco}/E^{true} distribution (the energy *response*) in each E^{true} bin is found by fitting a Gaussian to the central peak of the distribution and taking the mean of the fitted Gaussian. As mentioned above, this process proceeds via numerical inversion (Chapter 8), by inverting this response function and applying it to the reconstructed energies. The dependence of the energy response on E^{true} and η can be seen in Figure 5.20.

Following the absolute scale correction, the global sequential calibration (GSC) implements a sequence of residual corrections on various features are applied to correct the jet p_T. The sequence of corrections corrects the dependence of p_T on each feature



(a)



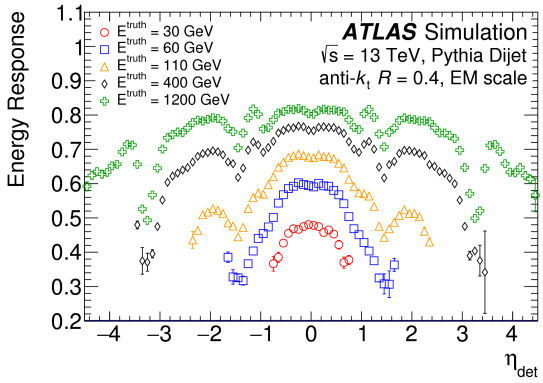
(b) (c)

Figure 5.19: (a) Distribution of pile-up density ρ for different values of NPV at fixed μ. (b) Effect of residual correction for NPV. (c) Effect of residual correction for μ. Figures sourced from [80].

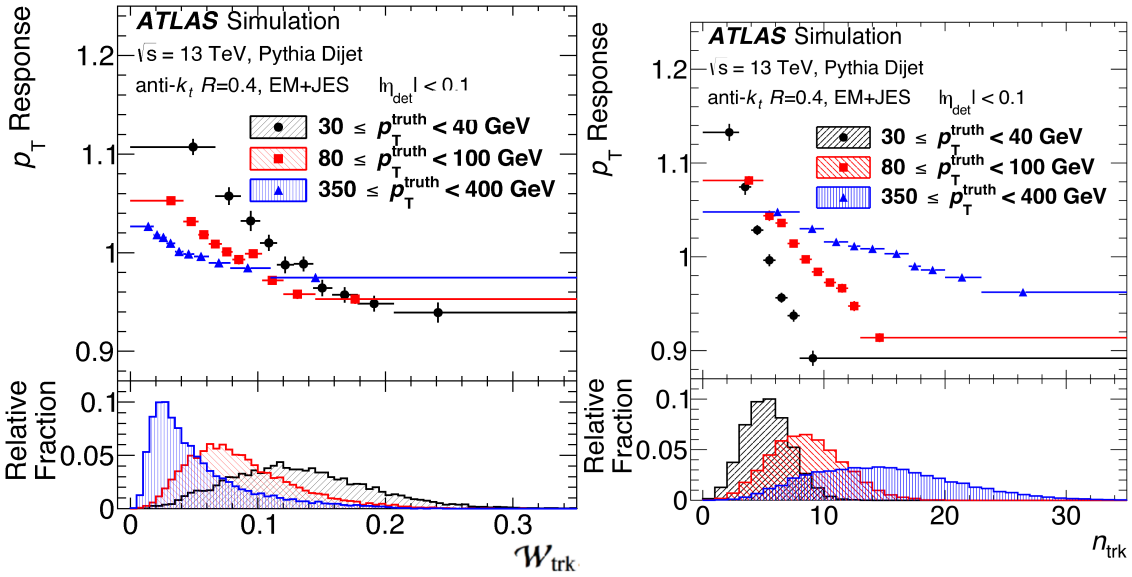
sequentially; since the dependence of the p_T response changes with p_T, these steps also proceed via numerical inversion. The features include those sensitive to the provenance of the jet (quark- or gluon-initiated) like the number of tracks in the jet and the width of tracks in the jet (the average p_T-weighted distance of tracks from the jet axis); removing this dependence makes the response more similar for quark and gluon jets and reduces the uncertainty due to jet fragmentation modeling for a given jet type. There are also two features to account for dead material between the electromagnetic and hadronic calorimeters (based on the fraction of jet energy in the final layer of the electromagnetic calorimeter and in the first layer of the hadronic calorimeters), and finally there is a feature to account for hadronic showers that punch through to the MS. The dependence of the p_T on some of the features that go into the GSC can be seen in Figure 5.20.

Chapter 9 proposes a new method of accounting for the dependence of the jet p_T on each of these auxiliary variables by deriving the correction simultaneously rather than sequentially. The new method is enabled by the use of neural networks to estimate the response dependence in the high-dimensional space formed by all the features, but retains the key properties of numerical inversion.

The above corrections are applied to all jets, both in simulation and data. There is a final step which compares the ratio of the jet response to a reference object in simulation and in data, and if there is a difference then the jets in data are corrected; this is called the *in situ* correction. As mentioned above, the reference objects include well-reconstructed $Z \to ee/\mu/mu$, γ, or multijet systems, taking advantage of conservation of momentum in the event. The correction using Z as the reference object cover the low p_T range $20 < p_T < 500$ GeV; the correction using γ as the reference object cover the medium p_T range $36 < p_T < 950$ GeV; and the multijet balance is used to extend the calibration up to 2 TeV. There is also an in situ η-intercalibration comparing jets at high η to more central jets. This correction also proceeds via numerical inversion. There can be uncertainties in the value of this correction due to mismodeling of physics effects; uncertainties in the kinematics of the reference object; and uncertainties in the p_T balance due to the event topology. These uncertainties are propagated at systematic uncertainties on the jet energy correction. The total



(a)

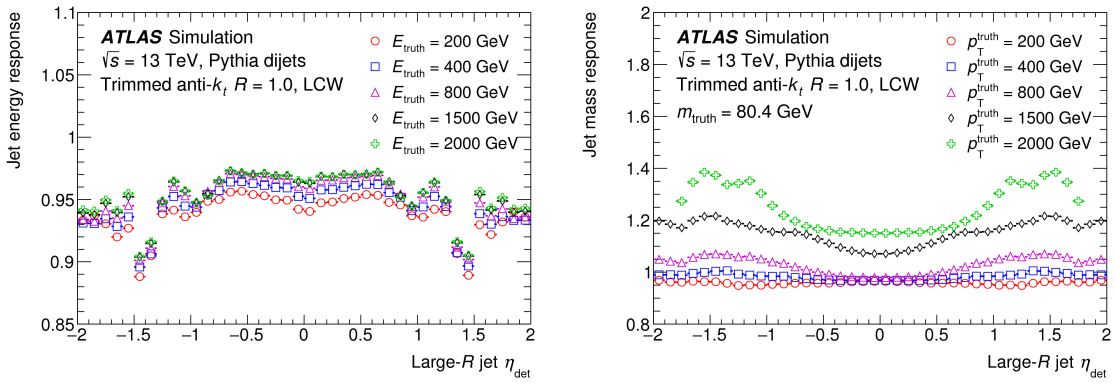


(b) (c)

Figure 5.20: Small-R jet calibration. (a) Energy response in bins of E^{true} and η for the absolute energy scale correction. (b) Dependence of p_T on track width for the GSC. (c) Dependence of p_T on number of tracks in jet for the GSC. Figures sourced from [80].

systematic uncertainty on the jet energy is around 1%, rising up to around 5% for low-p_T jets [291].

The calibration of large-R jets [82] does not include a pile-up subtraction step, as large-R jets are trimmed (Section 3.3) and are intended for high p_T massive objects for which pile-up has a small effect on the jet p_T. The energy correction for large-R jets in bins of η is derived in simulation similarly to that for small-R jets, again using numerical inversion. After the energy correction the large-R jet mass is also calibrated in bins of E and η, again using numerical inversion to prevent being biased by the distribution of the truth mass. The energy and mass response before calibration can be seen in Figure 5.21.



(a) (b)

Figure 5.21: Large-R jet calibration. (a) Energy response in bins of E^{true} and η for the absolute energy scale correction. (b) Mass response in bins of p_T^{true} and η for jets with truth mass around $m_W \approx 80$ GeV. Figures sourced from [82].

The Author has been involved in a project [292] intended to calibrate the jet energy and mass simultaneously rather than the sequential process outlined above. This project uses the simultaneous techniques enabled by neural networks introduced in Chapter 9.

There are two independent definitions of the jet mass [201]. The first is directly from summing together the four momenta of the constituent topoclusters (which are themselves assumed to be massless) and is called m^{calo}, or *calorimeter* mass. The second, called the *track* mass or m^{track}, is measured from tracks ghost-associated to

the jet. This measurement can provide a better measurement than the calorimeter mass due to the improved angular resolution of tracks compared to topoclusters; however, the tracks only comprise the charged particle component of the hadronic shower, so the overall scale of the mass will be off, even as the angular distribution is relatively accurate. To account for the absence of neutral particles in the track mass which are measured in the calorimeter, the track mass is scaled up by the ratio of the calorimeter and track p_T, forming the *track-assisted* mass [293] or $m^{TA} = m^{track} \frac{p_T^{calo}}{p_T^{calo}}$. Each of these two mass definitions are calibrated separately. The *combined* mass or m^{comb} is defined as a weighted sum between m^{track} and m^{TA} based on the (inverse) resolution of the truth jet mass under each definition. Since the sum of the weights is 1, m^{comb} is itself calibrated.

As with small-R jets, an in situ correction is applied to the large-R jet p_T in data, using as reference objects well-reconstructed $Z \to ee/\mu/m\mu$, γ, or multijet systems. The correction using Z as the reference object cover the low p_T range $200 < p_T < 500$ GeV; the correction using γ as the reference object cover the medium p_T range $500 < p_T < 1000$ GeV; and the multijet balance is used to extend the calibration up to 2.5 TeV. There is also an in situ η-intercalibration comparing jets at high η to more central jets. The in situ correction or large-R jets also proceeds via numerical inversion. The total systematic uncertainty on the jet energy rises from around 2% at low p_T to around 6% for high-p_T jets.

An in situ correction for the jet mass is also derived, though this correction cannot take advantage of p_T balance with a reference object. Two methods are employed to derive this correction. The first method compares the ratio of m^{calo} to m^{track} in data and simulation and provides a correction based on any differences, and uncertainties on this correction are applied as systematic uncertainties. The second method is called *forward folding* [294]. Both data and simulation are subjected to an event selection targeting $t\bar{t}$ events with high fidelity. This event selection takes advantage of events where one of the top quarks decays semileptonically ($t \to Wb \to l\nu b$) to unambiguously identify the top quark and the other decays hadronically ($t \to Wb \to qqb$), using the hadronic decay to calibrate the jet mass. In these events a correction to the mean and resolution of the mass response is derived by matching the entire

distribution seen in simulation to that seen in data (for the given choice of mass definition); this correction is taken as the in situ correction and uncertainties related to the measurement are taken as systematic uncertainties. A jet-topology dependence and uncertainty can also be derived from forward folding by requiring the tagged b-jet (Section 5.5.4.3) to be inside (for top jets) or outside (for W jets) the large-R jet. The jet mass response uncertainty depends on the jet p_T, but is generally less than 5%.

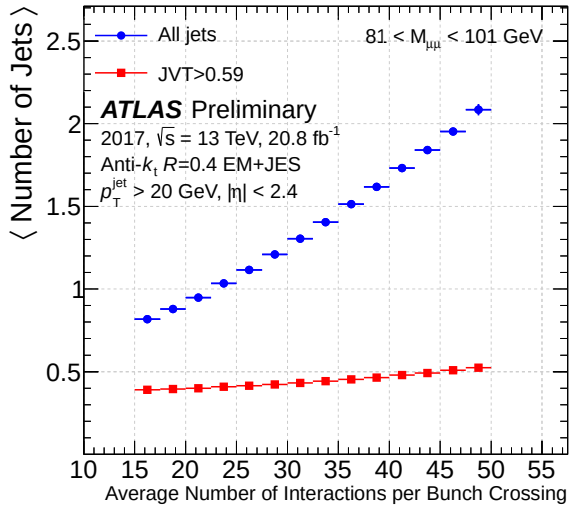
The tagging of jets based on their substructure is covered in Section 3.4. Briefly, small-R jets can be tagged as originating from quarks or gluons based on the number and distribution of tracks within the jet (gluons produce more particles and their showers are wider than those for quarks). Large-R jets can be tagged as fully hadronic decays originating from W/Z bosons (with two hard subprongs), from top quarks (with a W and a b-tagged subjet (Section 5.5.4.3)), or from Higgs bosons (with two b-tagged subjets). There are also decays involving leptons that can be used to tag large-R jets, e.g. $Z \to ee/\mu\mu$.

The following three Sections cover jet tagging not based strictly on the p_T and angular distribution of the tracks, topoclusters, and subjets within the jet: pile-up jet tagging (Section 5.5.4.2), b-tagging (Section 5.5.4.3), and τ-tagging (Section 5.5.4.4).

5.5.4.2 Pile-up Jet Tagging

While the pile-up contribution within small-R jets is subtracted out in the calibration process, there can be jets with low transverse momentum that originate entirely from pile-up. These jets can either be stochastic, meaning local fluctuations in the soft particle noise level from the sum of the other interactions in the event, or they can result from hard partons that originate from a single interaction other than the hard-scatter vertex (which presumably produced even harder objects to fire the trigger and have the highest $\sum p_T^2$.).

The dependence of the number of jets in the event on the average number of interactions per bunch crossing μ is shown in Figure 5.22. It can be seen that there is a linear component of the dependence on pile-up, corresponding to hard partons from a single interaction, and a superlinear component, corresponding to stochastic fluctuations.



(a)

Figure 5.22: Dependence of number of jets ($p_T > 20$ GeV) per event on the average number of interactions per bunch crossing. In blue, all jets. In red, after applying a selection of JVT> 0.59. Events from 2017 data-taking with a dimuon trigger and $81 < M_{\mu\mu} < 101$ GeV, corresponding to an event topology of Z+jets. Figure sourced from [295].

It is desirable to remove these pile-up jets from the event, since they are unrelated to the hard-scatter topology and would otherwise be a major source of noise in event selections. Jets are identified as *pile-up jets* [85, 296] (rather than *hard-scatter jets*) based on tracks ghost-associated with the jet and identified as originating from the hard-scatter vertex or a pile-up vertex.

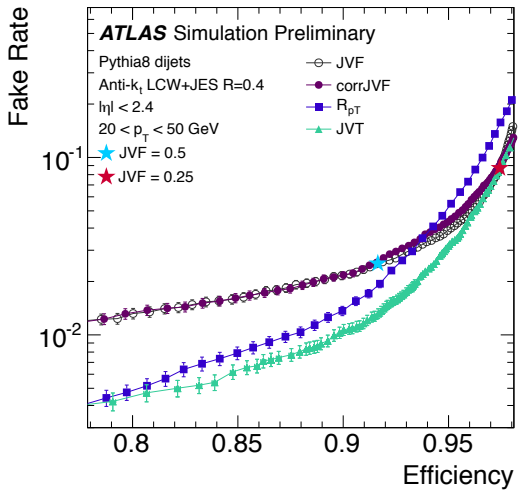
The *jet-vertex-fraction*, or *JVF*, is defined as the fraction of the scalar sum of associated track p_T coming from the hard-scatter primary vertex. While hard-scatter jets tend to have higher values of JVF than pile-up jets, the efficiency of a fixed cut on JVF degrades as the number of primary vertices increases. A *corrected JVF*, or *corrJVF*, variable is defined to account for this effect which compares the scalar sum of the associated track p_T coming from the hard-scatter primary vertex to the average (as opposed to total in the case of JVF) scalar sum of track p_T coming from pile-up vertices. Because of this correction, the efficiency of a fixed cut on corrJVF does not degrade with the number of primary vertices.

An additional variable, R_{pT}, is defined as the ratio of the scalar sum of associated track p_T coming from the hard-scatter primary vertex to the total p_T of the jet measured in the calorimeter, including the calibration with the pile-up subtraction step. This variable accounts for false positives with corrJVF in the case that a stray soft track from the hard-scatter vertex happens to land in a pile-up jet with no associated tracks. Hard-scatter jets tend to have a broad distribution in R_{pT} depending on the charged fraction of the shower, while pile-up jets have a sharp distribution near 0. As R_{pT} is defined in terms of only the hard-scatter vertex, it is to first order independent of the number of primary vertices.

A final variable, *jet-vertex-tagger* or *JVT*, is defined as a combination of these two variables. The two-dimensional likelihood distribution in hard-scatter jets and pile-up jets is formed using simulated dijet events. JVT is defined as the likelihood ratio between these two samples, estimated using a k-nearest neighbor algorithm [297] in $(\text{corrJVF}, R_{pT})$ space with a Euclidean metric and $k = 100$.

The curve for the rate of pile-up jets vs the efficiency on hard-scatter jets passing a selection on JVF, corrJVF, R_{pT}, and JVT is shown in Figure 5.23. It can be seen that JVT performs the best over all algorithms, and in particular that while JVT

and corrJVF perform similarly at high efficiency, JVT removes some false positives at lower efficiency due to the incorporation of R_{pT}.



(a)

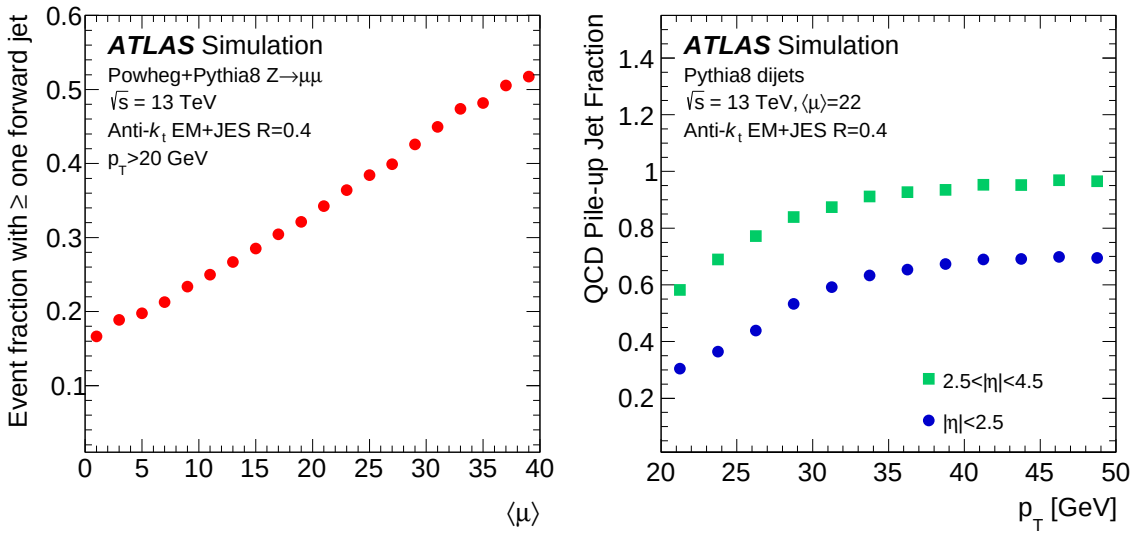
Figure 5.23: Fake rate (average number of pile-up jets per event) vs efficiency on hard-scatter jets as the selection on JVF, corrJVF, R_{pT}, and JVT is varied, in simulation. Figure sourced from [296].

A typical selection for jets in use in an analysis is JVT > 0.59, which has about a 90% efficiency on hard-scatter jets and a 0.02 pile-up jet rate. This selection is applied only to jets with $p_T < 60$ GeV, as the prior distribution of pile-up jets falls off rapidly with p_T and so a pile-up jet suppression is unnecessary (and would remove some high p_T hard-scatter jets). Pile-up jet suppression is not applied to large-R jets, as they are only calibrated down to 200 GeV and the rate of pile-up jets at that scale is negligible.

JVT of course relies on tracks and track-to-vertex association for pile-up identification, and so can only be applied to jets with $|\eta| < 2.4$ (*central jets*), the angular extent of the tracker (Section 5.2.1). For jets outside this range, or *forward jets*, a different pile-up tagger can be used [91]. The rate of forward pile-up jets is less than that for central pile-up jets, as can be seen in Figure 5.24. However, it is important to remove these forward pile-up jets for analyses that expect forward jets as part of

their signal selection - for example, the search presented in Chapter 6 targets Higgs bosons produced in the vector-boson-fusion mode, which tends to produce forward quark-initiated jets.

In contrast to central jets, forward pile-up jets are more likely to come from hard partons from a single pile-up interaction (*QCD pile-up jets*) than from stochastic fluctuations from multiple interactions (*stochastic pile-up jets*), as can be seen in Figure 5.24.



(a) (b)

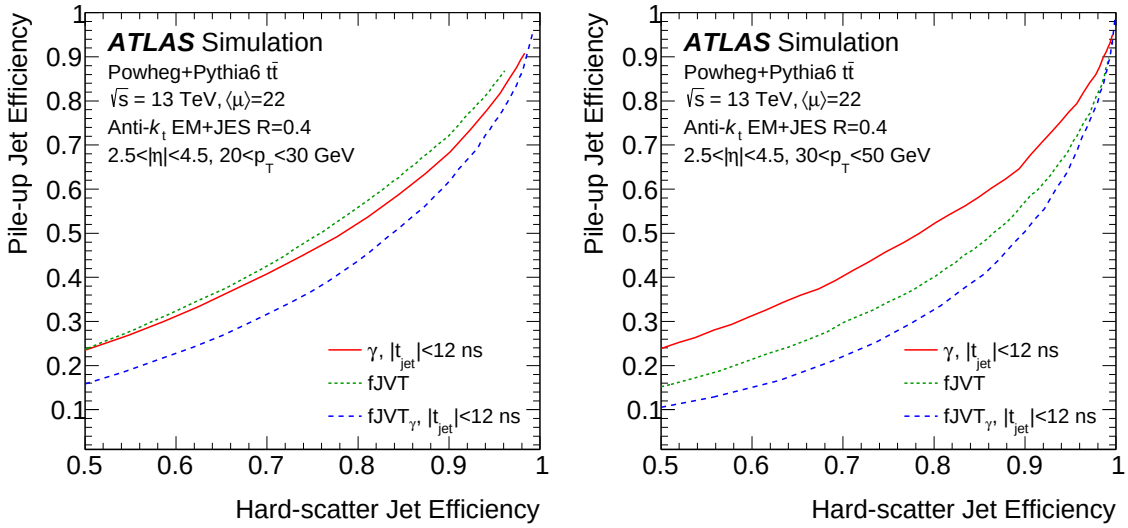
Figure 5.24: (a) Average number of events with at least one forward jet ($p_T > 20$ GeV) as a function of the average number of interactions per bunch crossing μ. (b) QCD pile-up jet fraction as a function of p_T for central and forward jets. Figures sourced from [91].

Stochastic pile-up jets are identified using the spatial width γ and temporal width of calorimeter towers in the catchment area of the jet, as hard-scatter jets tend to have a well-defined small core and stochastic pile-up jets tend to be diffuse.

QCD pile-up jets are identified taking advantage of the fact that the total momentum from each pile-up interaction must be conserved in the event. Therefore, events with forward QCD pile-up jets tend to have pile-up vertices with some missing vectorial transverse momentum in their associated tracks and jets, which matches

the vectorial transverse momentum present in the forward jet. Forward QCD pile-up jets can be identified taking advantage of this event topology; this algorithm is called *forward JVT* or *fJVT*.

The taggers based on γ and fJVT are combined together in a forward pile-up jet tagger called fJVT$_\gamma$, and the final tagger uses both fJVT$_\gamma$ and timing information to suppress forward stochastic and QCD pile-up jets. The performance of these taggers can be seen in Figure 5.25.



(a) (b)

Figure 5.25: Efficiency on pile-up jets vs efficiency on hard-scatter jets for taggers based on γ+timing information, fJVT, and fJVT$_\gamma$+timing information. (a) Jets with $20 < p_T < 30$ GeV; (b) jets with $30 < p_T < 50$ GeV. Figures sourced from [91].

The efficiencies of both JVT and the forward pile-up tagger are measured in data using $Z \rightarrow \mu\mu$+jets events and corrected for any data-simulation differences; the uncertainties on these corrections are applied as systematic uncertainties.

5.5.4.3 b-tagging

Jets originating from b-quarks (b-*jets*) can be identified taking advantage of the lifetime of B-hadrons, which is long enough for the decay to be measurably distant

from the primary vertex but short enough that the decay occurs in the inner detector ($\mathcal{O}(\text{mm})$)[14]. While the main background is jets originating from light quarks (*light-flavor jets*), jets originating from c-quarks (c-*jets*) also present a confounding background source, as hadrons containing c-quarks also tend to have a relatively long lifetime.

There are a variety of b-tagging algorithms [182–184], which combine low-level features related to the B-hadron decay into high-level multivariate taggers. The low level features fall into a few different categories: using the impact parameter of tracks matched to the jet directly (*IP2D* or *IP3D*) [298]; constructing secondary vertices using matched tracks (*SV1*) [299]; or fitting the entire B-hadron decay (*JetFitter*) [300]. These features are combined into two high-level multivariate tagging algorithms: *MV2*, which is trained using a boosted decision tree in TMVA [297], and *DL1*, which is trained using a deep feed-forward neural network in Keras [301]. The distributions of the scores from these high-level taggers can be seen in Figure 5.26. The background rejection rate, for light-flavor jets or for c-jets, as a function of the b-jet efficiency, for each of the above low-level and high-level taggers, can also be seen in Figure 5.26. DL1 does slightly better than MV2, with a rejection rate of light-flavor jets of about 10 at 70% efficiency on b-jets.

Typically analyses will use working points at the 60%, 70%, 77%, or 85% efficiency levels, depending on the needs of the analysis.

The efficiency of the b-taggers is measured in data in high-fidelity $\mathrm{t\bar{t}}$ events, where both of the top quarks decays leptonically $\mathrm{t} \to \mathrm{Wb} \to \mathrm{l\nu b}$. Any differences between data and simulation are corrected, and any uncertainties in this correction are applied as systematic uncertainties on the taggers.

5.5.4.4 Taus

Tau leptons decay hadronically about 65% of the time [107], in which case they are reconstructed as small-R jets (*tau-jets*)[15]. Of the hadronic decays, over 90% involve

[14] b-jets contain electrons or muons (+neutrinos) about 10% of the time each [107]; in those cases the leptons are used to identify the b-jet. This Section focuses on fully hadronic decays.

[15] Taus decay to electrons or muons (+neutrinos) about 17% of the time each [107]. Those cases are almost indistinguishable from prompt production of the (lighter) leptons, as the only other produced

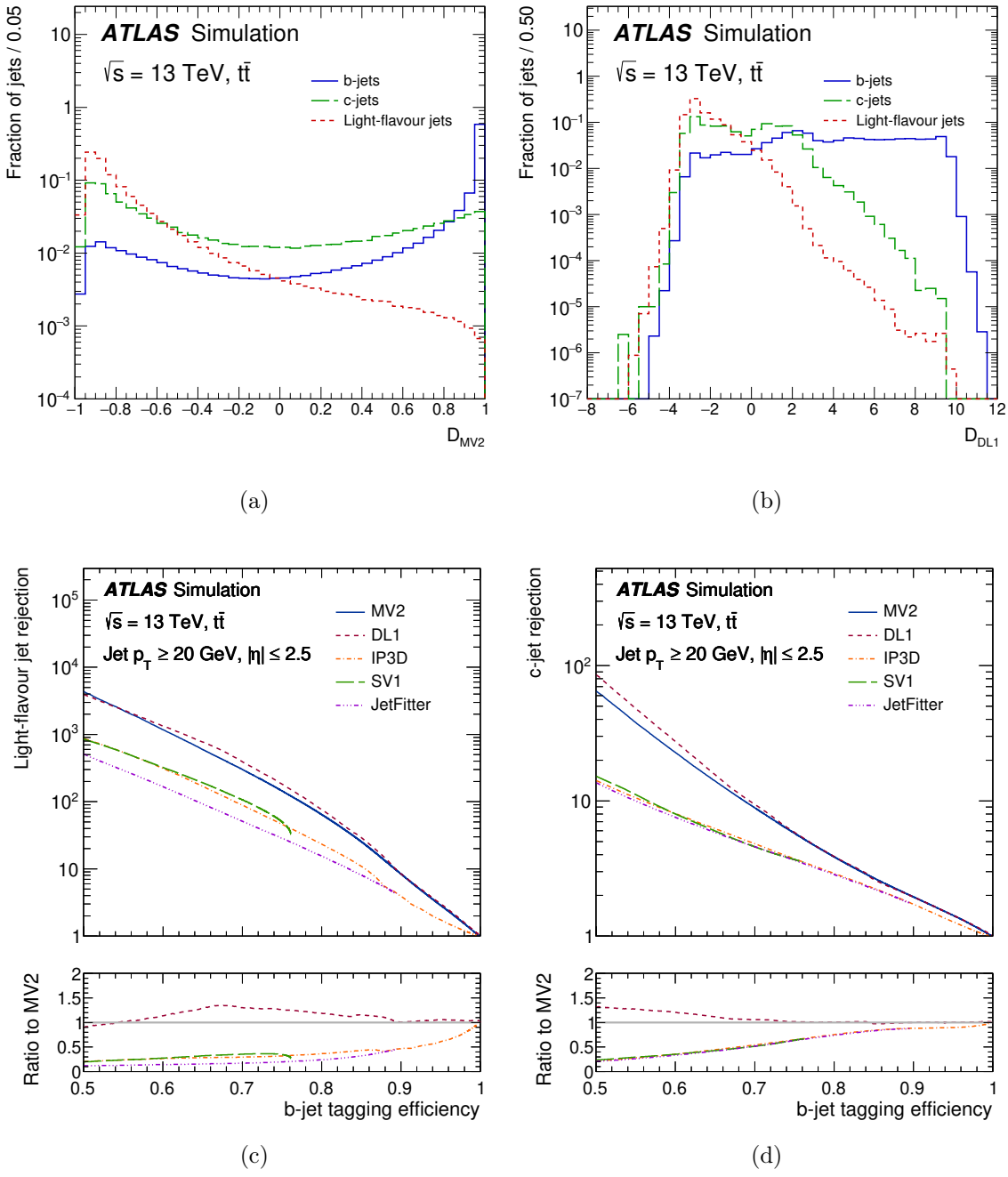


Figure 5.26: Distribution of scores from high-level b-taggers for light-flavor jets, c-jets, and b-jets: (a) MV2; (b) DL1. Background rejection rate of various b-taggers vs efficiency on b-jets: (c) light-flavor jets; (d) c-jets. Figures sourced from [184].

exactly 1 or 3 charged pions, up to two neutral pions, and a neutrino.

Tau-jets are identified [186, 302–304] given the number of tracks (1 or 3), taking advantage of the narrow shower shape relative to quark- and gluon-initiated jets. These low-level features are combined with a boosted decision tree into a final high-level tau-tagger. Electrons (Section 5.5.3) can pass the tau-jet identification with a single associated track; candidate tau-jets are removed if there is a nearby ($\Delta R < 0.4$) electron. Figure 5.27 shows the distribution of the high-level tagger for 1- and 3-prong jets initiated by quarks and gluons and for those initiated by taus. Also shown is the background rejection rate as a function of the tau-jet efficiency; for a tau-jet efficiency around 80%, the background rejection is around 30.

The energy of tau-jets is further calibrated to the truth visible energy (i.e. not including the neutrinos), in a manner similar to that for generic jets (Section 5.5.4.1).

The efficiency of the tau tagger and the energy calibration are measured in data with high-fidelity $Z \to \tau\tau$ events, where one of the taus decays to a muon to identify the event and the other decays hadronically. Any differences between data and simulation are corrected, and uncertainties on the correction are applied as systematic uncertainties.

5.5.5 Muons

Muons are reconstructed [305, 306] as charged particle tracks in the muon system. As almost all particles other than muons and neutrinos decay or shower in the calorimeter, muon identification has a very low background and events with muons in them can be used to identify specific physics topologies with high fidelity. Figure 5.28 shows an event display of an event likely corresponding to a Higgs boson (produced in the vector-boson-fusion mode) decaying to two muons.

Muons are reconstructed from hits in the inner detector just like any other charged particle. They are also reconstructed from hits in the various subsystems in the MS, using a Hough transform [309] to search for hits aligned along curved trajectories. Track segments are formed in each layer of the MS and combined to form track

particles are neutrinos. This Section focuses on fully hadronic decays.

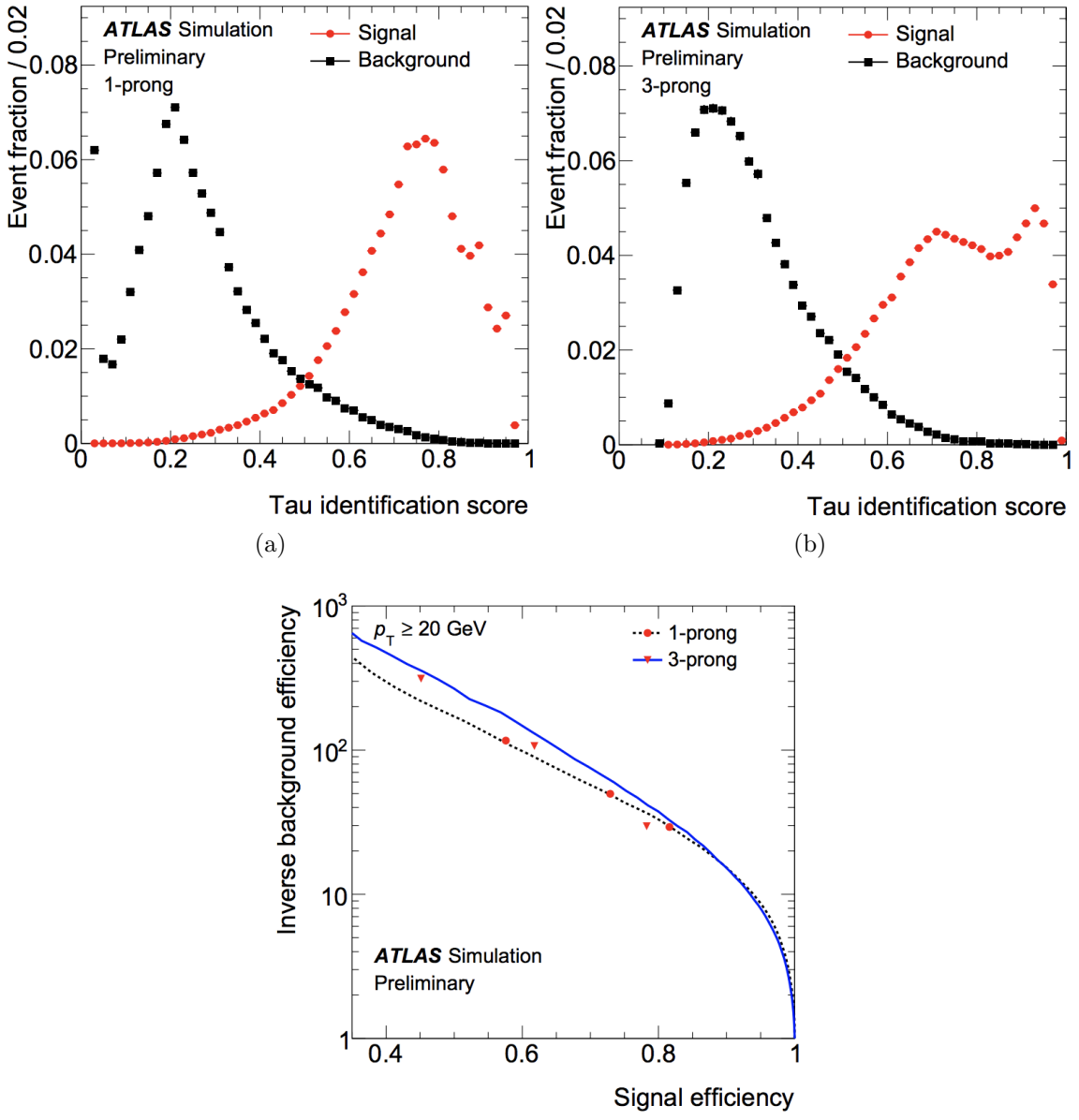
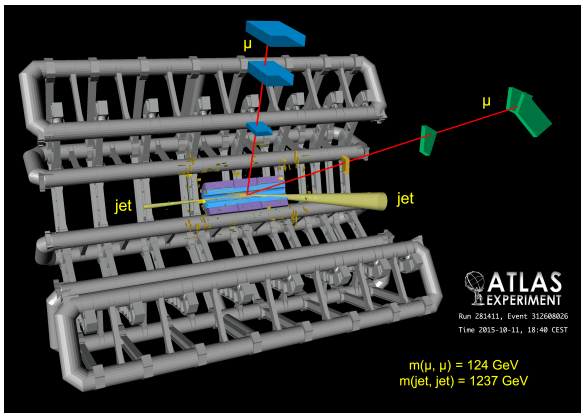


Figure 5.27: Distribution of scores from high-level tau-tagger for quark- and gluon-initiated jets (background) and for tau-jets (signal): (a) 1-prong jets; (b) 3-prong jets. (c) Background rejection rate vs efficiency on tau-jets. Some common working points are also included (not exactly on the lines due to implementing variable cuts intended to reduce the dependency of the efficiency on p_T). Figures sourced from [303].



(a)

Figure 5.28: Event display showing candidate Higgs boson produced in the vector-boson-fusion mode decaying to two muons. Figure sourced from [307], related to [308].

candidates, fitting the segments and tracks with a χ^2 fit to the hits. Tracks in the MS and in the inner detector are combined and fit again to form full muon candidates. Most muon candidates are formed outside-in (back-extrapolating MS tracks to inner detector tracks), but an inside-out method (extrapolating inner detector tracks to MS tracks) is used as a complementary approach. In regions that the MS covers but the inner detector does not ($2.5 < |\eta| < 2.7$), muon candidates can be formed from information only in the MS.

Fake muons can come from a variety of sources: from charged hadrons, primarily pions and kaons, decaying to muons, which produces a kink in the track in the inner detector; from charged hadrons punching through to the MS, which lose significant energy in the calorimeter relative to the track momentum; and from cosmic muons, which do not originate from the primary vertex. Each of these sources of background can be distinguished from prompt muons by testing the quality of the track in the inner detector and its compatibility with the track in the MS, especially the charge to momentum ratio. A few different working points are defined by selecting on these features - the `Medium` working point, which is the default for analyses, has about a 95% efficiency on prompt muons with $4 < p_T < 100$ GeV and $< 0.5\%$ efficiency on charged pions decaying mid-flight to muons.

Muons can also be required to be isolated using similar variables as for photons and

electrons (Section 5.5.3), depending on the signal topology targeted in an analysis.

The efficiency on muons for the identification and isolation is measured in simulation and in data in high-fidelity $Z \to \mu\mu$ events, where one muon and the invariant mass $M_{\mu\mu}$ are used to tag the event, and the efficiency on the other muon is tested. Differences in the efficiency between simulation and data are corrected, and uncertainties in this correction are applied as systematic uncertainties.

As muons are reconstructed as tracks in the MS, the charge-to-momentum ratio is measured very precisely. The muon momentum is also compared between simulation and high-fidelity $Z \to \mu\mu$ events in data. The differences between simulation and data in the momentum measurement are applied as corrections in the simulations and the uncertainties on this correction are applied as systematic uncertainties.

5.5.6 Missing Energy

Because of momentum conservation, the vectorial transverse momentum of all objects in the event should sum to 0[16]. Any missing transverse momentum in the event, or E_T^{miss}, could be due to invisible particles like neutrinos or new particles beyond the standard model.

E_T^{miss} is calculated [310–312] as the (opposite of the) vectorial sum of the transverse momentum from all other objects in the event - muons (Section 5.5.5), electrons and photons (Section 5.5.3), hadronically decaying tau jets (Section 5.5.4.4), and all other jets (Section 5.5.4) form the *hard term*, and all energy deposits in the event not matched to jets (Section 5.5.2) form the *soft term*.

For the hard term, objects are removed if they overlap with some better-measured object in order to avoid double counting; e.g., jets are removed if they overlap with a photon (which are reconstructed as jets), and only small-R jets are used in the jet term (as large-R jets are also reconstructed as small-R jets). Pile-up suppression in the central and forward region 5.5.4.2 has a major effect on the resolution of E_T^{miss}, as pile-up jets are unrelated to the hard-scatter event and only add an additional source of noise due to mismeasurements of their energies or incomplete event reconstruction

[16] This statement is not necessarily true for the longitudinal momentum, as the colliding protons could have a momentum imbalance with each other.

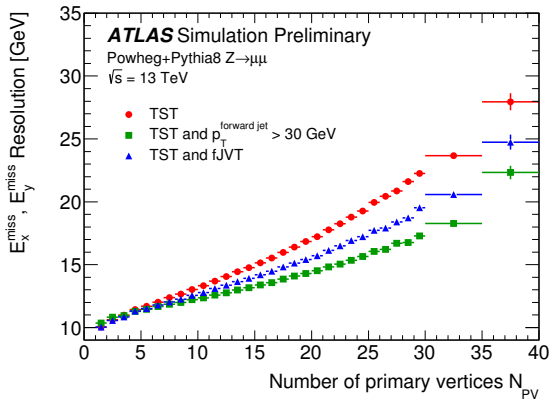
(even though in principle pile-up interactions should be p_T balanced themselves). The jet term can use either calorimeter topoclusters as defined in Section 5.5.2 or particle flow objects [174], which have been shown to improve the E_T^{miss} resolution.

The soft term is the hardest term to measure, as it accounts for energy deposits not reconstructed in other detector objects, which benefit from calibrations and pile-up suppression. Because of this, the standard definition of the soft term is the track-based soft term (though different definitions have been used in the past [310, 313]), which only uses all tracks with $p_T > 400$ MeV associated to the hard scatter primary vertex and not associated to any other object in the event, via ΔR matching or ghost association for jets. The gain in performance from excluding pile-up activity when using only tracks to measure the soft term makes up for the loss in performance due to not including neutral soft activity.

Despite this definition of the soft term, the E_T^{miss} resolution gets worse as the number of primary vertices (NPV) increases. Two working points are defined - a `Loose` working point using all jets with $p_T > 20$ GeV, requiring central jets to pass JVT, and a `Tight` working point which in addition removes all forward jets with $20 < p_T < 30$ GeV. There is in addition a middle working point that applies fJVT to forward jets. The resolution of E_T^{miss} in $Z \to \mu\mu$ events (which in the dominant Drell-Yan production mode has no true missing energy) as a function of NPV can be seen in Figure 5.29. The resolution clearly gets worse with increasing pile-up, but the dependence gets less as the tightness of the forward pile-up cut gets tighter.

As all the underlying objects that go into E_T^{miss} are themselves calibrated and have associated systematic uncertainties, the E_T^{miss} itself is in principle calibrated (given the same definition in simulation and data), and the underlying object systematic uncertainties as propagated up as uncertainties on E_T^{miss}. In addition, the E_T^{miss} distribution is compared between simulation and data in $Z \to ee/\mu\mu$ events, which can be identified with high fidelity, and any differences that exist are applied as systematic uncertainties.

There have been a variety of alternate definitions proposed for calculating E_T^{miss}. One proposed method is to incorporate the p_T resolution of the individual objects to get an E_T^{miss} significance [314]. An exciting alternative is to use a convolutional



(a)

Figure 5.29: The dependence of the E_T^{miss} resolution on the number of primary vertices (NPV). The red circles correspond to the `Loose` working point; green boxes correspond to the `Tight` working point; and blue triangles correspond to the middle working point of `Loose`+fJVT. Figure sourced from [312].

neural network [315–317] to predict E_T^{miss}, treating the entire event as an image in (η, ϕ) [318].

Chapter 6

A Search for a Beyond-the-Standard-Model Higgs Decay

This Chapter presents a search for exotic decays of the Higgs boson to a pair of new (pseudo)scalar particles, $H \to aa$, with a mass in the range 20–60 GeV, and where one of the a bosons decays into a pair of photons and the other to a pair of gluons. The results of this search are published in [319]. The search is performed in event samples enhanced in vector-boson fusion Higgs boson production by requiring two jets with large invariant mass in addition to the Higgs boson candidate decay products. The analysis is based on the full dataset of pp collisions at $\sqrt{s} = 13$ TeV recorded in 2015 and 2016 with the ATLAS detector at the CERN Large Hadron Collider, corresponding to an integrated luminosity of 36.7 fb^{-1}. The data are in agreement with the Standard Model predictions and an upper limit at the 95% confidence level is placed on the production cross section times the branching ratio for the decay $H \to aa \to \gamma\gamma gg$. This limit ranges from 3.1 pb to 9.0 pb depending on the mass of the a boson.

6.1 Introduction

The search for the Standard Model (SM) Higgs boson [21–24] has been one of the main goals of the LHC physics program. A new particle with mass of 125 GeV, and with properties compatible with those expected for the SM Higgs boson, has been discovered by the ATLAS [25] and CMS [26] collaborations. Since its discovery, a comprehensive program of measurements of the properties of this particle has been underway. These measurements could uncover deviations from expected SM branching ratios or allow for the possibility of decays into new particles.

Existing measurements constrain the branching ratio for such decays (B_{exotic}) to less than 34% at 95% confidence level [320]. Exotic decays are predicted by many theories of physics beyond the SM [55], including those with an extended Higgs sector such as the Next-to-Minimal Supersymmetric Standard Model [136–140], several models of dark matter [321–325], models with a first order electroweak phase transition [326, 327], and theories with neutral naturalness [328–330]. These theories are motivated by some of the most outstanding unanswered questions in physics, such as the gauge hierarchy problem [33], the nature of dark matter [28], and the strong CP problem [35].

One of the simplest possibilities is that the Higgs boson decays to a pair of new scalars or pseudoscalars, a, which in turn decay to a pair of SM particles. Several searches have been performed for $H \to aa$ in various final states [331–333].

The work presented here explores in particular the search for $H \to aa$, where the final state contains two photons (γ) and two gluons (g) ($H \to aa \to \gamma\gamma gg$). This decay mode becomes relevant in models where the $a \to$ fermion decays are suppressed and the a decays only to photons and gluons[1] [55,334]. The ATLAS Run 1 search for $H \to aa \to 4\gamma$ [335] sets a limit $\sigma \times B(H \to aa \to 4\gamma) < 10^{-3}\sigma_{SM}$ for $10\ \text{GeV} < m_a < 62\ \text{GeV}$. Before this work, there was no direct limit set on $B(H \to aa \to \gamma\gamma gg)$; however, in combination with $B_{exotic} < 34\%$, the $H \to 2a \to 4\gamma$ result sets indirect

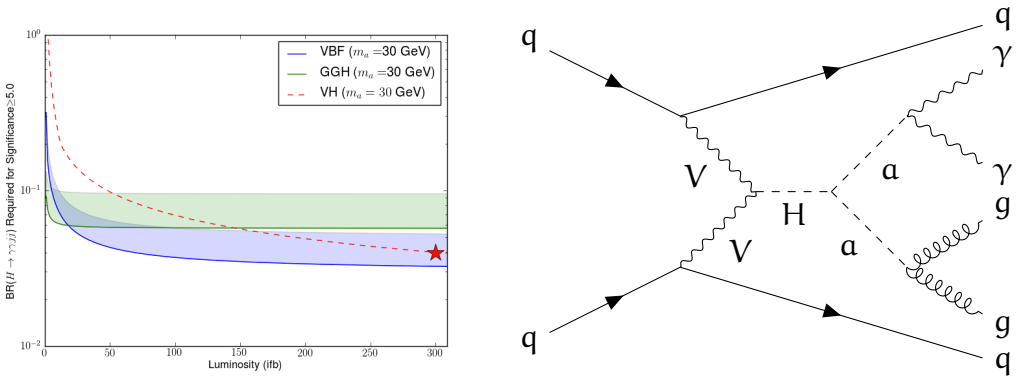
[1] This search is also sensitive to signal models of the type $H \to aa'$ ($a \neq a'$), where a decays primarily to $\gamma\gamma$, a' decays primarily to gg, and $m_a \approx m_{a'}$. For these models the $H \to aa \to 4\gamma$ search is not sensitive and there are no limits other than on the generic B_{exotic} rate, but these models are not as well-motivated.

limits on $B(H \to aa \to \gamma\gamma gg)$ to less than $\sim 4\%$ (Appendix A.1). Assuming a SM-like ratio of photon and gluon couplings, the $H \to 4\gamma$ decay occurs very rarely relative to the $H \to \gamma\gamma gg$ decay (a typical value of the relative ratio of branching ratios is 3.8×10^{-3} [334]) [2], making the $H \to \gamma\gamma gg$ an excellent unexplored final state for probing these fermion-suppressed models. The branching ratio for $a \to \gamma\gamma$ can be enhanced in some scenarios; the two searches are therefore complementary, where the $H \to \gamma\gamma gg$ final state is more sensitive to SM-like photon couplings with the new physics sector, while the $H \to 4\gamma$ final state is more sensitive to the enhanced photon scenarios.

The most common Higgs production modes at the LHC are through gluon fusion (ggF), vector boson fusion (VBF), or associated production with an additional vector boson (VH) [336]. The cross sections for each of these Higgs production modes is in a ratio of roughly 100:10:1 [337]; targeting any individual production mode requires a trade-off between higher production cross section and increased handles on distinguishing signal from background. Ref. [334] shows that the search for $H \to \gamma\gamma gg$ in the VH mode, where the additional vector boson decays leptonically, would require approximately 300 fb^{-1} of LHC data in order to be sensitive to branching ratios less than 4%. While the gluon fusion production has a larger cross section, it is overwhelmed by background. We have found (Figure 6.1a) that the search in the VBF production mode can achieve sensitivity competitive with or better than the above two production modes in the range 20 to 300 fb^{-1} of LHC integrated luminosity data, and in particular with the approximately 150 fb^{-1} of data at the end of LHC Run 2 [231].

In the VBF production mode, two extra quarks are produced along with the Higgs in the hard scatter interaction. The gluons from the Higgs decay and the quarks from the VBF production are reconstructed as jets (j) in the ATLAS detector; therefore, this search is focused on the $H \to 2\gamma 2j$ final state with two additional jets from the VBF production. Figure 6.1b shows a tree-level Feynman diagram of the VBF production and $\gamma\gamma gg$ decay of the Higgs boson, including the aa intermediate state.

[2] The $H \to 4g$ final state is the most common, but is unfortunately swamped by background.



(a) (b)

Figure 6.1: (a) Projected branching ratio for $H \to \gamma\gamma gg$ in order to make a discovery of new physics at a significance level of 5σ, as a function of integrated luminosity, when searching in the gluon fusion (GGH), vector boson fusion (VBF), and associated production (VH) production modes. The projection for GGH and VBF assumes a 20% systematic uncertainty due to jet reconstruction effects, while the projection for VH uses the estimation from [334], assuming the significance is statistics-dominated. With less than $\sim 20 fb^{-1}$, the gluon fusion mode is most sensitive, but with more statistics this production mode quickly becomes dominated by systematic uncertainties and is unable to provide better limits. Up to $\sim 300 fb^{-1}$, the VBF mode is most sensitive; in Run 2 the LHC has gathered $140 fb^{-1}$. With more than $\sim 300 fb^{-1}$, the VH mode likely provides the best sensitivity. (b) Tree-level diagram of production and decay of Higgs boson being searched for in this analysis.

6.2 Data and simulation

The search presented in this chapter is based on the 36.7 fb^{-1} dataset of proton–proton collisions recorded by the ATLAS experiment at the LHC at $\sqrt{s} = 13$ TeV during 2015 and 2016. As discussed in Chapter 5, the ATLAS detector [56] comprises an inner detector in a 2 T axial magnetic field, for tracking charged particles and a precise localisation of the interaction vertex, a finely segmented calorimeter, a muon spectrometer and a two-level trigger [258] that accepts about 1 kHz rate for data storage.

Monte Carlo (MC) event generators were used to simulate the $H \to aa \to \gamma\gamma gg$

signal (Section 5.3). Signal samples for the ggF and VBF processes were generated at next-to-leading order using POWHEG-BOX [249, 250, 338] interfaced with PYTHIA [245] for parton showering and hadronization using the AZNLO set of tuned parameters set [339] and the CT10 parton distribution function (PDF) set [252]. Samples were generated in the m_a range[3] $20\ \text{GeV} < m_a < 60\ \text{GeV}$, assuming the a boson to be a (pseudo)scalar. All MC event samples were processed through a detailed simulation [340] of the ATLAS detector based on GEANT4 [255], and contributions from additional pp interactions (pile-up), simulated using PYTHIA and the MSTW2008LO PDF set [341], were overlaid onto the hard-scatter events.

6.3 Event Selection

6.3.1 Trigger

Events are selected by requiring at least one of two diphoton triggers. One trigger path requires the presence in the electromagnetic (EM) calorimeter of two clusters of energy deposits with transverse energy above 35 GeV and 25 GeV for the leading (highest transverse energy) and sub-leading (second-highest transverse energy) clusters, respectively. In the high-level trigger the shape of the energy deposit in both clusters is required to be loosely consistent with that expected from an EM shower initiated by a photon. The other trigger path requires the presence of two clusters with transverse energy above 22 GeV. In order to suppress the additional rate due to the lower transverse energy threshold, the shape requirements for the energy deposits are more stringent. Using the ATLAS trigger nomenclature, these two trigger paths are called `HLT_g35_loose_g25_loose` and `HLT_2g22_tight`, respectively. Both of these high-level trigger are seeded by the Level 1 trigger `L1_2EM15VH`, which requires two Level 1 objects consistent with the EM shower initiated by a photon with transverse energy greater than 15 GeV. In 2015 and 2016 data these triggers were unprescaled. The efficiency of each of the triggers on signal events with $m_a = 30$ GeV is shown in

[3] The diphoton triggers considered for this search do not have acceptance for the lower mass range ($m_a < 20$ GeV), where the two photons are collimated.

Table 6.1: Efficiency of triggers on a signal sample with $m_a = 30$ GeV. The efficiencies are shown for the signal MC separated by production in the VBF and GGF modes, and for the combined production modes.

	Efficiency		
	VBF	ggF	Combined
`HLT_g35_loose_g25_loose`	0.158	0.161	0.160
`HLT_2g22_tight`	0.166	0.208	0.200
OR of triggers	0.201	0.242	0.234

Table 6.1. The efficiency of requiring at least one of the two triggers for other masses of m_a can be seen in Figure 6.9. The trigger efficiency notably goes down for $m_a < 20$ GeV; this limits the sensitivity of this search. The cause of this trigger inefficiency is discussed in Appendix B.1.

6.3.2 Photons

The photon candidates are reconstructed from the clusters of energy deposits in the EM calorimeter within the range $|\eta| < 2.37$. The energies of the clusters are calibrated to account for energy losses upstream of the calorimeter and for energy leakage outside the cluster, as well as other effects due to the detector geometry and response. The calibration is refined by applying η-dependent correction factors of approximately $\pm 1\%$, derived from $Z \to ee$ events [286]. As in the trigger selection, photon candidates are required to satisfy a set of identification criteria based on the shape of the EM cluster [283]. Two working points are defined: a *Loose* working point, used for the preselection and the data-driven background estimation, and a *Tight* working point, with requirements that further reduce the misidentification of neutral hadrons decaying to two photons. In order to reject the hadronic jet background, photon candidates are required to be isolated from any other activity in the calorimeter. The calorimeter isolation is defined as the sum of the transverse energy in the calorimeter within a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$ centered around the photon candidate, The transverse energy of the photon candidate is subtracted from

the calorimeter isolation. Contributions to the calorimeter isolation from the underlying event and pile-up are subtracted using the method proposed in [83]. Candidates with a calorimeter isolation larger than 2.2% of the photon's transverse energy are rejected.

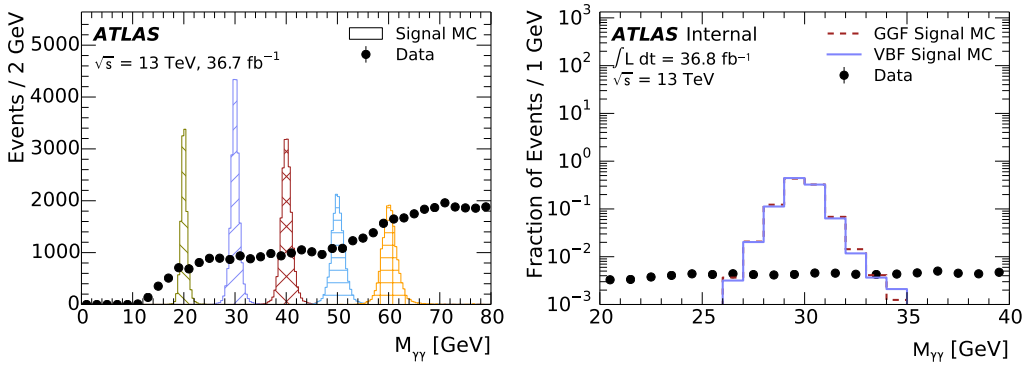
Events are required to have at least two photon candidates. The transverse energy requirements depend on the trigger path through which the event was recorded. For events passing the trigger with higher transverse energy thresholds (`HLT_g35_loose_g25_loose`), the leading photon is required to have $E_T > 40$ GeV, and the sub-leading photon is required to have $E_T > 30$ GeV. For events passing the trigger with lower thresholds (`HLT_2g22_tight`), both the leading and sub-leading photons are required to have $E_T > 27$ GeV. For events passing both triggers, the latter selection is applied.

The efficiency of requiring at least two photon candidates in the signal can be seen in Figure 6.9. This photon selection efficiency notably goes down for $m_a < 20$ GeV (beyond the inefficiency from the diphoton trigger); this further limits the sensitivity of this search. The cause of this inefficiency is discussed in Appendix B.1, but is mostly due to the calorimeter isolation on photons discussed above.

The invariant mass of the two leading photon candidates is denoted by $m_{\gamma\gamma}$. The $m_{\gamma\gamma}$ resolution is excellent, with $>\sim 97\%$ of signal events satisfying $|m_{\gamma\gamma} - m_a| < 2.5$ GeV, regardless of the value of m_a; this can be seen in Figure 6.2.

6.3.3 Jets

Jets are reconstructed from topological clusters [173] using the anti-k_t algorithm [171] implemented in the FastJet package [180] with a radius parameter of $R = 0.4$. Jets are calibrated using an energy- and η-dependent calibration scheme, and are required to have a transverse momentum (p_T) greater than 20 GeV and $|\eta| < 2.5$ or $p_T > 30$ GeV and $|\eta| < 4.4$. The jets with $|\eta| < 2.5$ are referred to as *central* jets and those with $2.5 < |\eta| < 4.4$ are referred to as *forward* jets. In this analysis both central and forward jets are used, however the requirements on these two categories of jets differ due to the treatment of pile-up jets in the central and forward region of the detector. Central jets with $p_T < 60$ GeV and $|\eta|<2.4$ are required to satisfy the *Default* Jet



(a) (b)

Figure 6.2: Distribution of $m_{\gamma\gamma}$. The quantities are plotted separately for signal MC produced in the VBF and GGF modes. (a) The distribution for some of the signal samples used in this analysis (each signal normalized to a branching ratio $BR(H \to aa \to \gamma\gamma jj) = 0.03$). (b) Zoomed in distribution, for a signal with $m_a = 30$ GeV.

Vertex Tagger selection [85]. Forward jets with $p_T < 50$ GeV are required to pass the *Tight* forward JVT cut [91]. Jets must have an angular separation of $\Delta R > 0.4$ from any *Loose* photon candidate in the event.

In the VBF production mode, the Higgs boson is produced in association with two additional light-quark jets with a large opening angle and a large invariant mass. Selected events are therefore required to have at least four jets.

6.3.4 Preselection

The preselection detailed in Sections 6.3.1, 6.3.2, and 6.3.3 is summarized in Table 6.2.

After the preselection, the expected dominant contribution to the background consists of events with two real radiated photons and additional radiated jets (photons + jets), as well as events with multiple jets with one or two of those falsely passing the photon identification requirements (fake photon + jets). Ultimately the background estimation is entirely data-driven, but the remaining selections are intended to enhance the signal presence while reducing the presence of this background component.

Table 6.2: Event preselection.

	Selection	
L1 trigger	`L1_2EM15VH`	
Primary HLT trigger	`HLT_g35_loose_g25_loose`	`HLT_2g22_tight`
Photon Selection	≥ 2 photons with $E_T > 30$ GeV ≥ 1 photon with $E_T > 40$ GeV	≥ 2 photons with $E_T > 27$ GeV
Jet Selection	≥ 4 jets, central or forward	

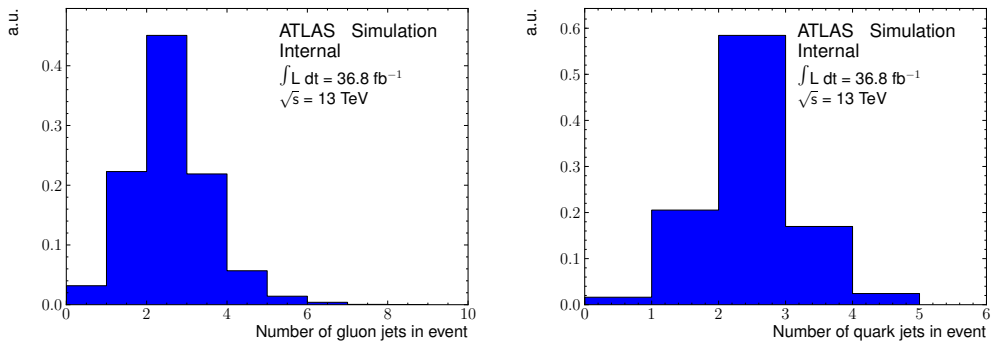
6.3.5 VBF Selection

The two additional light-quark jets that are produced in association with the Higgs boson in the VBF production mode tend to have a large opening angle and therefore a large combined invariant mass. The pair of jets with the highest invariant mass (m_{jj}^{VBF}, or VBF m_{jj}) are referred to as *VBF jets*. The two remaining highest-p_T jets are referred to as *signal jets*, with invariant mass m_{jj}. The jet assignment is designed to choose the correct VBF and signal jets for VBF signal events. The truth parton label (the truth ID of the highest p_T parton in the jet) is used to study the accuracy of this jet assignment in MC.

As seen in Figure 6.3, after the 4 jet preselection requirement, about 70% of VBF signal events contain at least 2 gluon jets; also, about 70% of VBF signal events contain at least 2 quark jets. This limits the possible accuracy of the jet assignment.

The best possible jet assignment would choose the two gluon jets with mass closest to m_a as the signal jets and two quark jets as the VBF jets, if such an assignment is possible. Figure 6.4 shows the results of this assignment (which can only be done at truth level) for a VBF signal sample with $m_a = 30$ GeV.

In the m_{jj} distribution there is a peak about m_a with width about $0.4m_a = 12$ GeV. Only about 35% of the events have two gluons with an invariant mass within this peak; this indicates that most of the time the two true signal gluons are either



(a) (b)

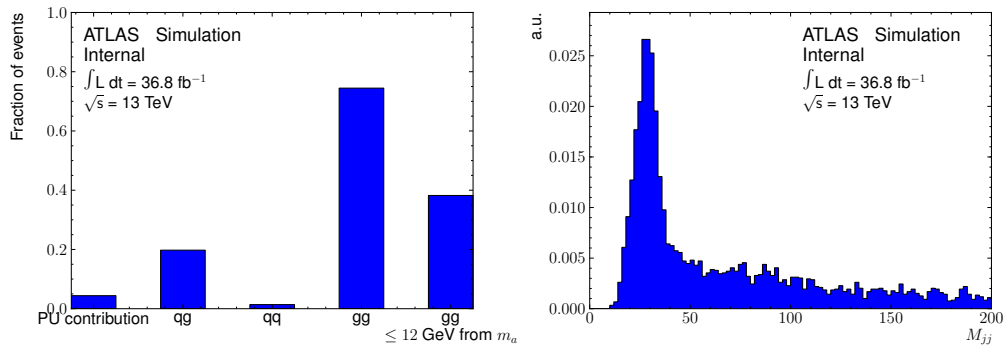
Figure 6.3: Jet multiplicity of (a) gluon and (b) quark jets in a VBF signal sample with $m_a = 30$ GeV.

not reconstructed or below the jet p_T cut. This 12 GeV width also informs the choice of a $0.4m_a$ cut on $|m_{jj} - m_{\gamma\gamma}|$ in the ABCD regions definition (6.3.6).

Figure 6.5 shows the results of the jet assignment used in this analysis for a VBF signal sample with $m_a = 30$ GeV. About 20% of the events choose two gluons with an invariant mass within 12 GeV of m_a as the signal jets. About 55% of the events choose two quarks correctly as the VBF jets.

Figure 6.6 examines the distribution of VBF m_{jj} with the jet assignment used in this analysis in a VBF signal sample with $m_a = 30$ GeV using truth information. The distributions are compared among all events, events where both VBF jets are truth identified as quarks, and events where at least one VBF jet is not truth identified as a quark. It can be seen that the VBF m_{jj} distribution is shifted upwards relative to the case where the truth quarks are not correctly selected, but the case where the truth quarks are always correctly selected is shifted to even higher values.

Figure 6.5 also shows the results of the jet assignment used in this analysis for a gluon fusion signal sample with $m_a = 30$ GeV and for a photons + jets background MC sample. In the gluon fusion signal sample, about 20% of the events choose two gluons with an invariant mass within 12 GeV of m_a as the signal jets. In these events the jets labeled the "VBF jets" in fact come from pile-up, underlying event, or initial or final state radiation.



(a) (b)

Figure 6.4: (a) Truth parton label of jets chosen as signal jets in a VBF signal sample with $m_a = 30$ GeV; the jet assignment is to choose the two truth gluons with mass closest to m_a as the signal jets and the two highest p_T truth quark jets as the VBF jets, if such an assignment is possible (i.e., this assignment can only be done at truth level). (b) The distribution of m_{jj} for the signal jets.

The VBF Higgs boson signal is further enhanced, relative to the dominant $\gamma\gamma$+multi-jet background, by requiring m_{jj}^{VBF} to be greater than 500 GeV and the p_T of the leading VBF jet to be greater than 60 GeV. The discrimination power of these observables can be seen in the difference in shape between the VBF signal and the data, shown in Figure 6.7.

This selection is summarized in Table 6.3.

Table 6.3: VBF Selection.

	Selection
VBF M_{jj}	>500 GeV
VBF Leading Jet p_T	>60 GeV

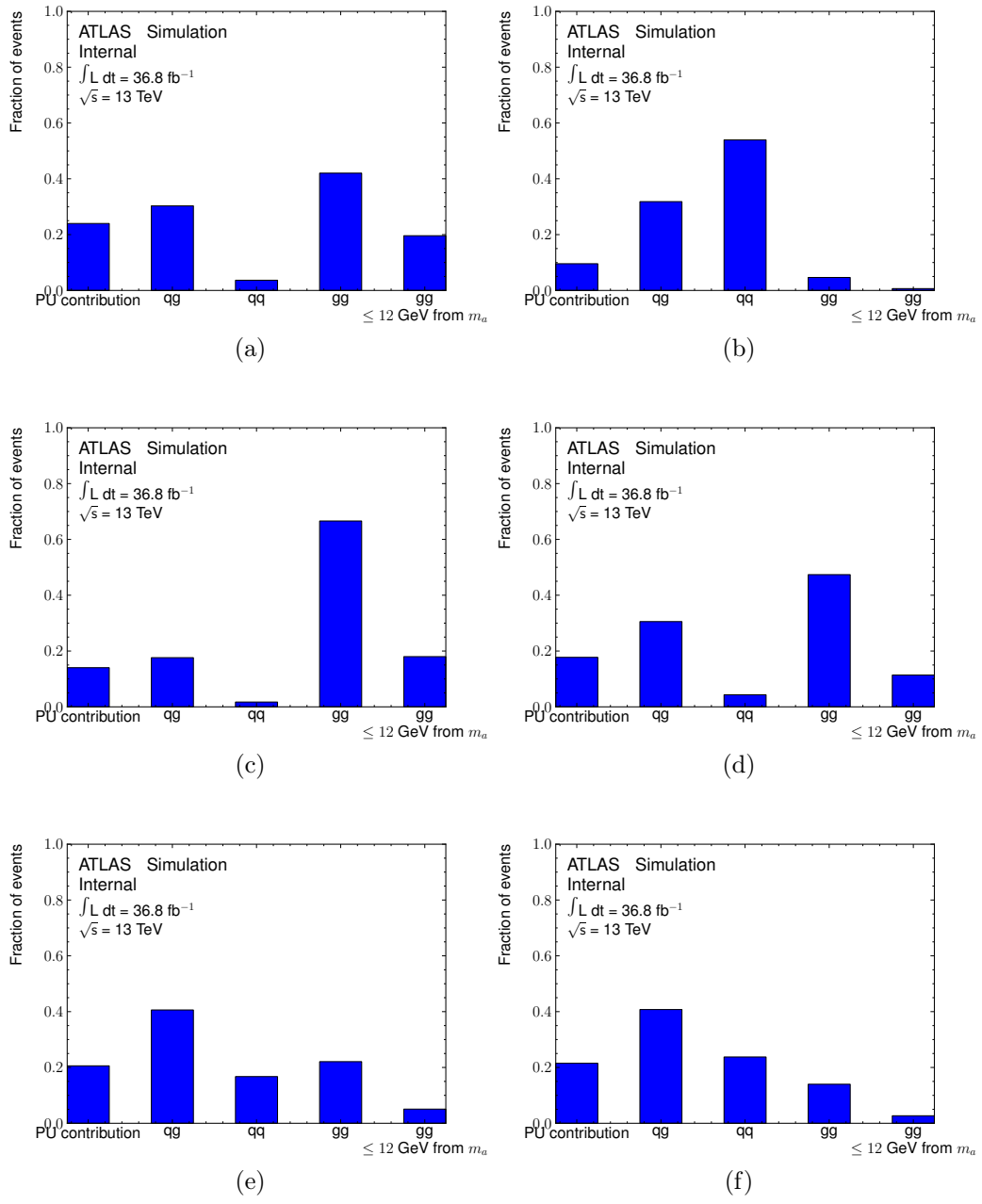
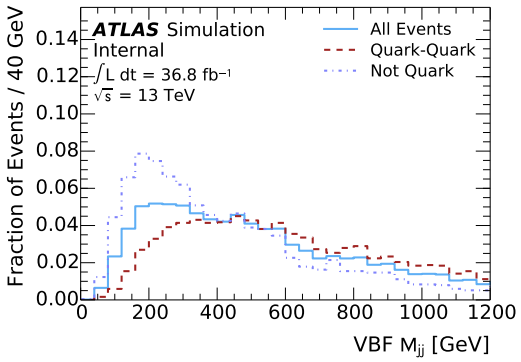


Figure 6.5: Truth parton label of jets chosen as (a) signal jets and (b) VBF jets in a VBF signal sample with $m_a = 30$ GeV; (c) signal jets and (d) VBF jets in a gluon fusion signal sample with $m_a = 30$ GeV; and (e) signal jets and (f) VBF jets in a photons + jets background sample.



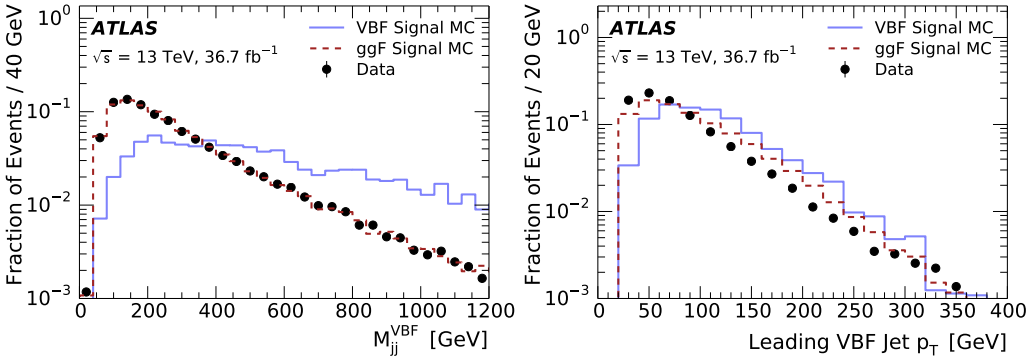
(a)

Figure 6.6: Comparison of distribution of VBF m_{jj} with the jet assignment used in this analysis in a VBF signal sample with $m_a = 30$ GeV. The comparison is between the distribution for all events; the distribution for events where both jets identified as VBF jets are truth labeled as quark jets; and the distribution for events where at least one jet identified as as a VBF jet is not truth labeled as a quark jet.

6.3.6 Combined Signal Selection

As mentioned in 6.3.5, the two remaining highest-p_T jets after the VBF jet assignment are referred to as *signal jets*, with invariant mass m_{jj}. The distribution of m_{jj} is shown in Figure 6.8; the resolution is quite broad, approximately 30% of the central value. The tail of the m_{jj} distribution comes from misidentifying the signal jets. In signal events it is expected that $m_{jj} \sim m_{\gamma\gamma} \sim m_a$ up to misidentifications of the signal jets and resolution of the signal objects. Because of this, the difference $|m_{jj} - m_{\gamma\gamma}|$ is expected to be smaller for signal events than for the background; the distribution of this quantity is also shown in Figure 6.8. This quantity is therefore used to define a signal-enhanced region for the data-driven background estimation (Section 6.4).

The two photon candidates and the two signal jets form the Higgs boson candidate with invariant mass $m_{\gamma\gamma jj}$. The requirement $|m_{\gamma\gamma jj} - m_H| \leq 25$ GeV, with $m_H \sim 125$ GeV the standard model Higgs mass [107], is imposed in order to select events consistent with a Higgs boson decay and is designated the *signal* region. The region $|m_{\gamma\gamma jj} - m_H| > 25$ GeV is designated the *validation* region and is used to validate the background estimation detailed in Section 6.4. Figure 6.8 shows that most of the selected signal events lie within the signal region, especially after the requirement of



(a) (b)

Figure 6.7: Distributions of kinematic observables before the requirements on m_{jj}^{VBF}, leading VBF jet p_T, $m_{\gamma\gamma jj}$ and $|m_{jj} - m_{\gamma\gamma}|$ for: (a) m_{jj}^{VBF}; and (b) leading VBF jet p_T. The quantities are shown separately for simulated signal events (with $m_a = 30$ GeV) produced in the VBF mode and compared with those produced in the ggF mode and the observed data.

$|m_{jj} - m_{\gamma\gamma}|$ used to define the signal-enhanced region, while the data have a broad distribution extending to higher values.

The distributions of m_{jj}, $|m_{jj} - m_{\gamma\gamma}|$, and $|m_{\gamma\gamma jj}$ for other values of m_a are included in Appendix A.2 - the conclusions for these other values of m_a are largely the same.

The definitions of the signal and validation regions are listed in Table 6.4.

Table 6.4: Signal and validation region definitions.

Region	Cuts
Signal	$\lvert m_{jj\gamma\gamma} - m_H \rvert \leq 25$ GeV
Validation	$\lvert m_{jj\gamma\gamma} - m_H \rvert > 25$ GeV

In order to take advantage of the very good $m_{\gamma\gamma}$ resolution to suppress the background with $m_{\gamma\gamma}$ far from the range of interest, five overlapping $m_{\gamma\gamma}$ regimes are defined as summarised in Table 6.5. The size of the $m_{\gamma\gamma}$ window in each regime is 15 GeV, in order to fully cover two signal samples separated by 10 GeV in m_a. Slightly different window sizes are used for the lowest and highest $m_{\gamma\gamma}$ regimes. Since the efficiency of the trigger and preselections is very low on signal with $m_a < 20$ GeV

(as can be seen in Figure 6.9), the lowest $m_{\gamma\gamma}$ regime starts at 17.5 GeV. Since the maximum possible value of m_a is $\frac{m_H}{2} \approx 62.5$ GeV, the highest $m_{\gamma\gamma}$ regime is limited to 65 GeV. For each $m_{\gamma\gamma}$ regime, the set of m_a values for which this requirement causes no significant signal acceptance loss is also indicated.

The efficiency of the various analysis cuts in the signal can be seen in Figure 6.9. The mass points with $m_a \leq 10$ GeV have little acceptance from the trigger selection. The $m_a = 20$ GeV mass point has almost an order of magnitude less overall efficiency than the other samples with greater values of m_a. The mass points with the highest efficiencies are the $m_a = 30$ GeV and $m_a = 40$ GeV samples, with the $m_a = 50$ GeV and $m_a = 60$ GeV samples having slightly worse efficiencies.

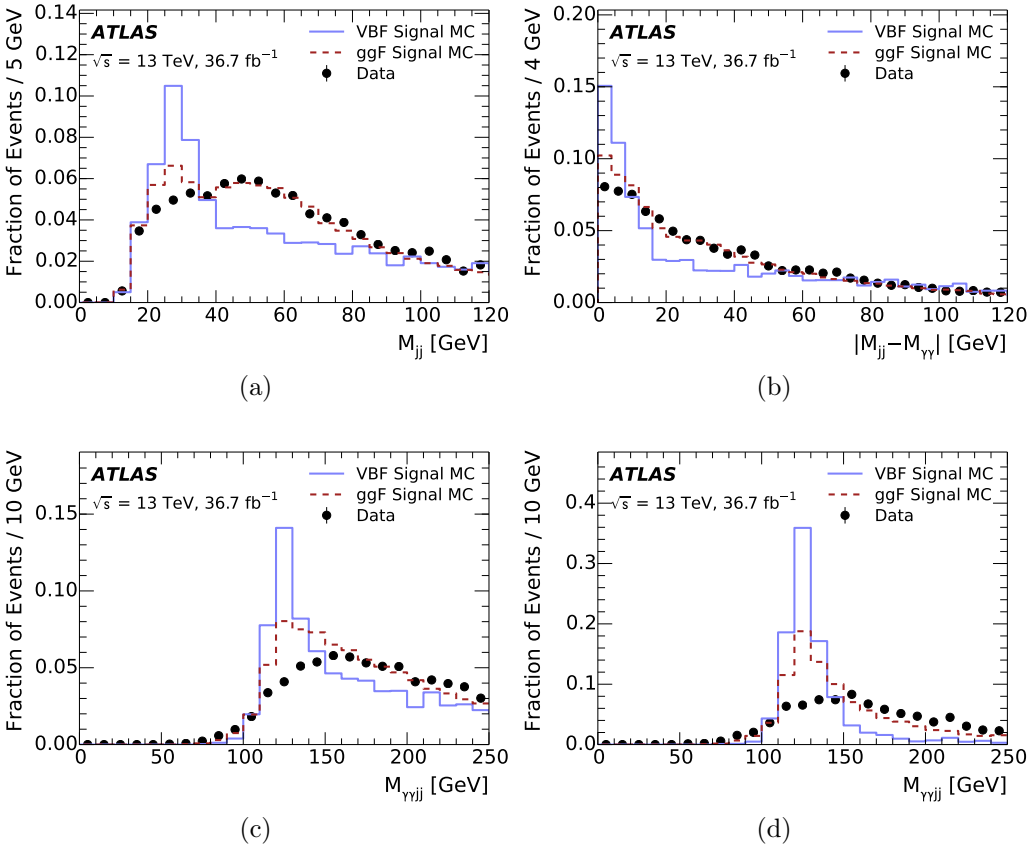


Figure 6.8: Distributions of kinematic observables before the requirements on m_{jj}^{VBF}, leading VBF jet p_T, $m_{\gamma\gamma jj}$ and $|m_{jj} - m_{\gamma\gamma}|$ for: (a) m_{jj}; (b) $|m_{jj} - m_{\gamma\gamma}|$; (c) $m_{\gamma\gamma jj}$; and (d) $m_{\gamma\gamma jj}$ (with the additional requirement $|m_{jj} - m_{\gamma\gamma}| < 12$ GeV that defines the signal-enriched region). The quantities are shown separately for simulated signal events (with $m_a = 30$ GeV) produced in the VBF mode and compared with those produced in the ggF mode and the observed data.

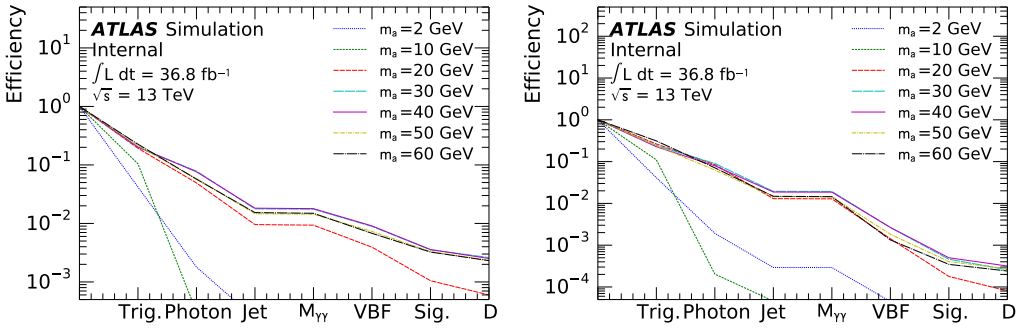


Figure 6.9: Efficiency of analysis cuts up to and including indicated cut on signal MC, for different mass points. The cut definitions are: Trig. - HLT trigger selection (Table 6.2); Photon - Photon preselection (Table 6.2); Jet - Jet preselection (Table 6.2); $m_{\gamma\gamma}$ - The highest efficiency $m_{\gamma\gamma}$ analysis regime cuts (Table 6.5); VBF - VBF cuts (Table 6.3); Signal - Signal region cuts (Table 6.4); D - D region cuts (as defined in 6.4). The signal is separated by production modes into (a) VBF and (b) ggF.

6.4 Background estimation

The $\gamma\gamma$+multi-jet background consists of multi-jet events with two reconstructed photon candidates, originating from isolated EM radiation or from jets. A data-driven estimation based on two-dimensional sidebands is used to predict the background yields. The method consists of using two uncorrelated observables to define four regions labelled A, B, C and D.

The first axis of the A/B/C/D plane separates events in regions C and D with both photons passing the *Tight* requirement from events in regions A and B with at most one photon passing the *Tight* requirement and at least one passing the *Loose* but not the *Tight* requirement. These regions are referred to respectively as *Tight–Tight* (C and D) and *Tight–Loose* (A and B).

The second axis separates events in regions B and D, satisfying $|m_{jj} - m_{\gamma\gamma}| < x_R$, from events in regions A and C, satisfying $|m_{jj} - m_{\gamma\gamma}| > x_R$. The value x_R depends on the $m_{\gamma\gamma}$ regime R to account for size of the resolution at higher mass (Appendix A.2). As mentioned in Section 6.3.6, the difference $|m_{jj} - m_{\gamma\gamma}|$ tends to be smaller in the signal than in the background. The signal events that lie outside of the range $|m_{jj} - m_{\gamma\gamma}| < x_R$ are due to poor m_{jj} resolution or to incorrect assignment of the jets corresponding to the gluons originating from the a boson decay. Specific x_R values are given in Table 6.5. In each $m_{\gamma\gamma}$ regime, the boundary for $|m_{jj} - m_{\gamma\gamma}|$ is 0.4 times the central $m_{\gamma\gamma}$ value. An exception is made for the lowest $m_{\gamma\gamma}$ regime, where x_R is larger in order to increase the signal efficiency.

The definitions of the A/B/C/D regions are shown in Table 6.6.

The distribution of events across $|m_{jj} - m_{\gamma\gamma}|$ (i.e., the second axis of the A/B/C/D method), separated into Tight–Tight and Tight–Loose events (i.e., the first axis of the A/B/C/D method) can be seen in Figure 6.10, for both the validation and signal regions. It can be seen that in the validation region, where the data statistics are higher and there is little signal contamination, the two variables corresponding to the two axes are uncorrelated, prompting the background estimation strategy outlined below (Equation 6.1). It can also be seen that the highest contribution of signal in the signal region occurs in region D.

Table 6.5: Definition of each $m_{\gamma\gamma}$ regime, the range of m_a values considered in the scope of this search with no significant signal loss acceptance due to the $m_{\gamma\gamma}$ requirement, and the corresponding boundary x_R for $|m_{jj} - m_{\gamma\gamma}|$.

$m_{\gamma\gamma}$ regime	Definition	Range of m_a values	x_R [GeV]
1	17.5 GeV $< m_{\gamma\gamma} <$ 27.5 GeV	20 GeV $\leq m_a \leq$ 25 GeV	12
2	22.5 GeV $< m_{\gamma\gamma} <$ 37.5 GeV	25 GeV $\leq m_a \leq$ 35 GeV	12
3	32.5 GeV $< m_{\gamma\gamma} <$ 47.5 GeV	35 GeV $\leq m_a \leq$ 45 GeV	16
4	42.5 GeV $< m_{\gamma\gamma} <$ 57.5 GeV	45 GeV $\leq m_a \leq$ 55 GeV	20
5	52.5 GeV $< m_{\gamma\gamma} <$ 65.0 GeV	55 GeV $\leq m_a \leq$ 60 GeV	24

Table 6.6: The A/B/C/D regions. A different x_R boundary is chosen for each analysis regime, as detailed in Table 6.5.

		Photon requirements	
		TightLoose	TightTight
$\lvert m_{jj} - m_{\gamma\gamma}\rvert$	$> x_R$	A	C
	$< x_R$	B	D

The efficiency in each of the A/B/C/D regions for the gluon fusion and VBF Higgs production modes in the signal region can be seen in Table 6.7. It can be seen that in region D, $\sim 60\%$ of the signal events are produced in the VBF mode and the remaining $\sim 40\%$ in the ggF mode - the ggF events are produced about $10\times$ as often as VBF events [337], and the VBF efficiency is about $20\times$ higher than the ggF efficiency. The efficiency of the event selection for the $pp \to H \to aa \to \gamma\gamma gg$ signal combining the two production modes in each of the A/B/C/D regions is shown in Table 6.8, assuming the SM cross-section and kinematics for the different production modes as described in [337].

Assuming the two observables used to define the A/B/C/D regions are independent in the background distribution, the number of background events in the signal region D (N_D^{bkg}) is related to the number of background events in the control regions

Table 6.7: Efficiency of analysis cuts for a sample of signal masses m_a in each of the ABCD regions. The analysis regime used is the one most efficient at that signal mass. The uncertainties account for the total effect of the systematic sources of uncertainty. (a) Simulated signal events produced in the gluon fusion mode. (b) Simulated signal events produced in the vector boson fusion mode.

(a)

m_a [GeV]	A ($\times 10^{-5}$)	B ($\times 10^{-5}$)	C ($\times 10^{-5}$)	D ($\times 10^{-5}$)
20	$0.25^{+0.17}_{-0.12}$	$0.26^{+0.20}_{-0.11}$	2.4 ± 0.4	2.4 ± 0.8
30	0.86 ± 0.24	1.39 ± 0.31	$4.4^{+0.8}_{-1.1}$	8.5 ± 2.0
40	0.67 ± 0.25	1.4 ± 0.5	$3.8^{+0.9}_{-1.3}$	10.3 ± 2.3
50	0.28 ± 0.14	1.6 ± 0.4	2.4 ± 0.6	9.0 ± 2.2
60	$0.16^{+0.19}_{-0.11}$	1.8 ± 0.4	1.8 ± 0.5	8.4 ± 2.2

(b)

m_a [GeV]	A ($\times 10^{-5}$)	B ($\times 10^{-5}$)	C ($\times 10^{-5}$)	D ($\times 10^{-5}$)
20	4.1 ± 1.6	14 ± 5	27 ± 13	60 ± 17
30	6.9 ± 2.1	31 ± 9	55 ± 18	260^{+60}_{-80}
40	8.6 ± 3.4	30 ± 9	68^{+14}_{-19}	250 ± 60
50	10.1 ± 2.8	44 ± 14	42^{+18}_{-10}	230^{+60}_{-70}
60	9 ± 4	34 ± 11	50 ± 16	230 ± 60

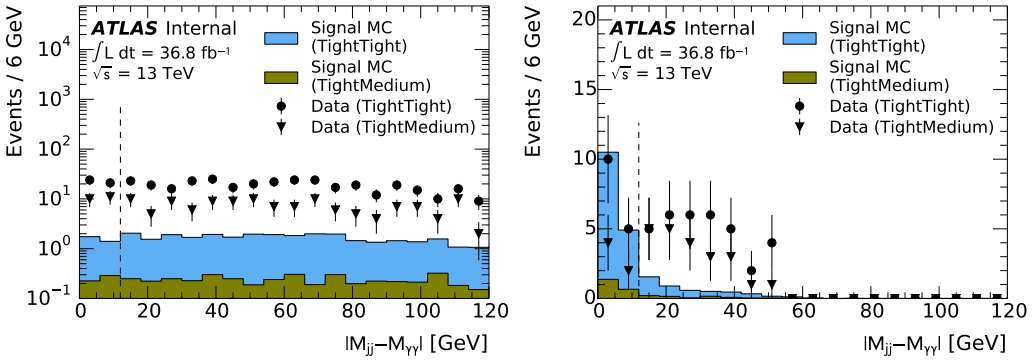


Figure 6.10: Distribution of data and signal MC with $m_a = 30$ GeV, normalized to a branching ratio BR($H \to aa \to \gamma\gamma jj$) = 0.03, in each region of the ABCD method, using analysis regime 2. (a) The validation region. (b) The signal region.

A, B and C, denoted by N_A^{bkg}, N_B^{bkg} and N_C^{bkg}, respectively, by the formula

$$N_D^{bkg} = \frac{N_B^{bkg} N_C^{bkg}}{N_A^{bkg}}. \tag{6.1}$$

In the following, the difference between the prediction N_D^{bkg} and the actual background yield in region D is referred to as *non-closure*. The non-closure results from residual correlations between the two observables used to define the A/B/C/D regions, and the uncertainty accounting for this effect is referred to as *closure uncertainty*. In order to quantify the non-closure, the data-driven estimation as described above is performed in the validation region. For each $m_{\gamma\gamma}$ regime, the closure uncertainty is defined to be the central value of the non-closure if it is found to be significant ($> 1\sigma$) in comparison with its statistical uncertainty; otherwise, the statistical uncertainty of its estimate is used. The events in region D of the signal region were blinded while the analysis selections were being optimized, and only unblinded after the entire analysis strategy was frozen. Figures 6.11 and 6.12 show the distribution of the data in each of the $m_{\gamma\gamma}$ analyses regimes, in the validation and signal regions, respectively, with an expected signal that has high efficiency in that analysis regime. The observed events in the signal D region are blinded to draw attention to the closure in

Table 6.8: Efficiency of event selection on the $pp \to H \to aa \to \gamma\gamma gg$ signal, assuming the SM Higgs boson production cross-section and kinematics, in each of the A/B/C/D regions, for different m_a mass hypotheses. For each m_a value, all $m_{\gamma\gamma}$ regimes in which there is no significant signal acceptance loss due to the $m_{\gamma\gamma}$ requirement are shown.

m_a [GeV]	$m_{\gamma\gamma}$ regime	Efficiency ($\times 10^{-5}$)			
		A	B	C	D
20	1	$0.50^{+0.16}_{-0.14}$	1.2±0.4	3.9±1.1	6.2±1.8
25	1	$0.67^{+0.27}_{-0.33}$	$2.6^{+0.5}_{-0.6}$	5.8±1.4	15±4
25	2	$0.67^{+0.27}_{-0.33}$	$2.6^{+0.5}_{-0.6}$	5.8±1.4	15±4
30	2	1.22±0.34	3.3±0.9	$7.6^{+1.4}_{-1.6}$	25^{+5}_{-6}
35	2	1.8±1.1	2.7±1.2	9.3±2.6	27±6
35	3	$0.53^{+1.20}_{-0.24}$	4.1±1.2	$6.1^{+1.2}_{-1.6}$	31±7
40	3	1.2±0.4	3.3±1.0	$7.9^{+1.7}_{-2.4}$	26±6
45	3	2.5±1.0	4.1±1.3	$7.7^{+1.7}_{-2.0}$	19±5
45	4	2.2±0.9	4.4±1.4	$5.9^{+1.5}_{-2.2}$	22±5
50	4	0.93±0.30	4.4±1.2	$5.0^{+1.3}_{-1.0}$	24±5
55	4	0.37±0.11	3.3±0.9	$5.4^{+1.3}_{-1.4}$	21±5
55	5	0.23±0.16	3.6±1.0	3.4±0.8	24±6
60	5	$0.77^{+0.32}_{-0.30}$	3.9±1.0	4.9±1.4	23±6

the validation region and the difference between the predicted contribution from the background closure and the signal contribution. The unblinded results can be found in Section 6.6.

In the validation region closure can be observed, to within the statistical uncertainties.

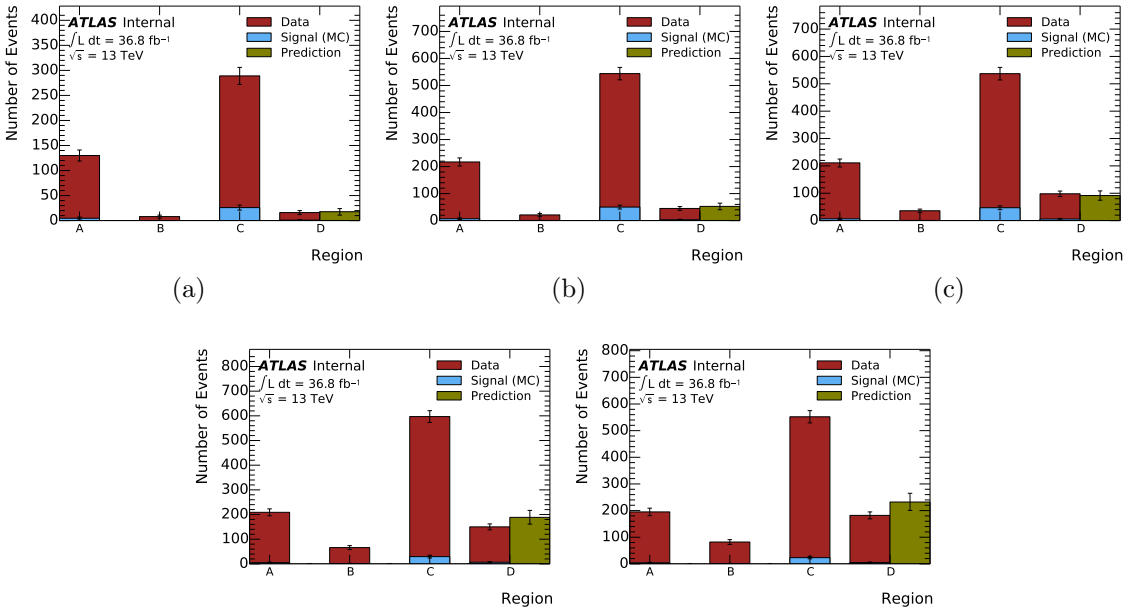


Figure 6.11: Number of events in data and signal MC in each region of the ABCD method, in the validation region. The analysis regimes are (a-e) 1-5. The data prediction assumes closure (6.1) in the background in the absence of signal. The signal is normalized to a branching ratio BR($H \to aa \to \gamma\gamma jj$) = 0.03. The errors shown are statistical.

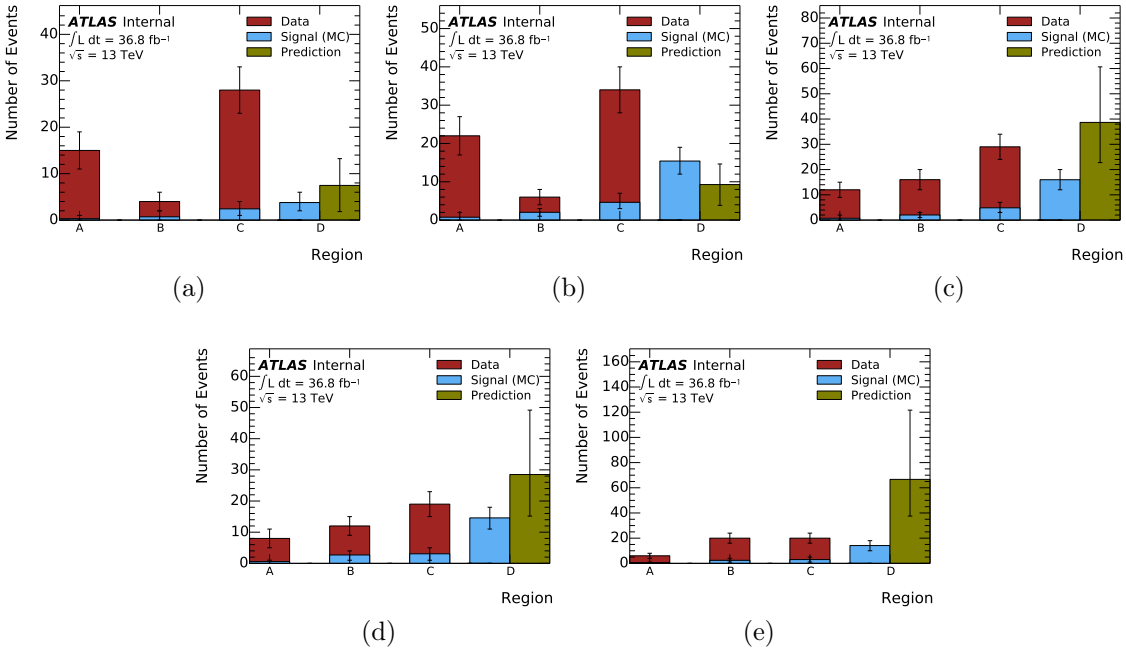


Figure 6.12: Number of events in data and signal MC in each region of the ABCD method, in the signal region. The analysis regimes are (a-e) 1-5. The data prediction assumes closure (6.1) in the background in the absence of signal. The signal is normalized to a branching ratio BR($H \to aa \to \gamma\gamma jj$) = 0.03. The errors shown are statistical. The observed events in the signal D region are blinded to draw attention to the closure in the validation region and the difference between the expected and the signal contribution.

6.5 Statistical Analysis

The statistical analysis is performed using an ABCD likelihood method which takes into account the likelihood of the data in each of the ABCD regions. This method is useful for dealing with situations in which the contamination of the signal in the control regions is non-trivial or the number of events in one of the control regions is small.

The likelihood function is defined as a product of Poisson likelihood functions over each of the ABCD regions:

$$\begin{aligned}\mathcal{L}(\mu_S, \mu, \tau_B, \tau_C | N_A, N_B, N_C, N_D) = & \frac{(\mu_S + \mu)^{N_D} e^{-(\mu_S+\mu)}}{N_D!} \\ & \times \frac{(b\mu_S + \tau_B\mu)^{N_B} e^{-(b\mu_S+\tau_B\mu)}}{N_B!} \\ & \times \frac{(c\mu_S + \tau_C\mu)^{N_C} e^{-(c\mu_S+\tau_C\mu)}}{N_C!} \\ & \times \frac{(a\mu_S + \tau_B\tau_C\mu)^{N_A} e^{-(a\mu_S+\tau_B\tau_C\mu)}}{N_A!} \end{aligned} \tag{6.2}$$

The parameters are defined as follows. μ_S and μ are the expected number of signal and background events, respectively, in the D region. τ_B and τ_C are the expected contamination of the background in the B and C regions respectively, so that the expected number of background events are $\tau_B\mu$ and $\tau_C\mu$, respectively. Closure in the signal region is assumed, so that the expected number of background events in region A is $\tau_B\tau_C\mu$ (i.e., the assumption of closure allows the background to be parameterized in terms of only 3 variables across the A/B/C/D regions). a,b, and c are the expected (from Monte Carlo) contamination of the signal in the A, B, and C regions respectively, so that the expected number of signal events are $a\mu_S$, $b\mu_S$, and $c\mu_S$, respectively (Table 6.8). Finally, N_A, N_B, N_C, and N_D are the number of data observed in each of the A, B, C, and D regions, respectively.

The statistical uncertainty on the closure in the validation region is assessed as a systematic uncertainty on the background prediction in the signal region. In case the non-closure is larger than a 1-σ statistical fluctuation, the size of the non-closure is

applied as a systematic on the background prediction.[4] This information is summarized in Table 6.9. The closure uncertainty, is included in the likelihood function by applying a Gaussian prior to the expected number of background events in region A. Note that this uncertainty is statistical in nature.

Table 6.9: Closure in each $m_{\gamma\gamma}$ analysis regime.

Regime	Closure	Background uncertainty syst.
1	1.11±0.50	0.50
2	1.17±0.32	0.32
3	0.93±0.20	0.20
4	1.26±0.21	0.26
5	1.28±0.20	0.28

The low number of observed events is the dominant source of uncertainty for this search. The second largest uncertainty is due to the closure uncertainty, also statistical in nature.

Other sources of systematic uncertainty only affect the overall signal normalisation and the amount of signal contamination in control regions A, B and C. Dominant sources of experimental systematic uncertainty arise from the calibration and resolution of the energy of the jets [80, 342]. Uncertainties associated with the photon energy calibration and resolution [286], as well as the photon identification and isolation efficiencies [283], are found to be negligible. Uncertainties associated with the estimation of the integrated luminosity and the simulation of pile-up interactions (*Lumi and Pile-up*) are found to be negligible. The systematic uncertainty associated with the modelling of the kinematics in signal events (*Modelling*) is evaluated by varying the choice of scales used in the generator program and assuming the SM Higgs boson production [343]. It is found to be similar in size to the experimental systematic uncertainty. The effect of these uncertainties on the best fit signal strength are included in Table 6.11.

[4]Note that the distribution of the non-closures in the validation region are consistent with that expected due to Poisson fluctuations; this larger uncertainty is applied conservatively.

Nuisance parameters corresponding to each source of uncertainty are included in the profile likelihood as Gaussian constraints.

The likelihood is maximized over all possible values of μ_S, μ, τ_B, and τ_C using the `MINUIT` migrad algorithm [344], marginalizing over the nuisance parameters. Then the maximum likelihood (MLE) value of $\hat{\mu} = \mathrm{BR}(H \to aa \to gg\gamma\gamma)$ is obtained by appropriately normalizing the signal Monte Carlo efficiency to the expected number of Higgs events in the data sample based on the total integrated luminosity.

The 95% confidence limit on $\hat{\mu}$ is obtained using the asymptotic limit of the CLs method [345, 346].

6.6 Results

The observed number of events in each of the A/B/C/D regions for each $m_{\gamma\gamma}$ regime is shown in Table 6.10 along with the predicted background in the signal region D, taking into account the closure uncertainty. This information is also presented in Figure 6.13.

Due to the low event counts in each of the A/B/C/D regions, the median expected background yield in region D estimated from pseudo-data experiments involving asymmetric Poisson uncertainties in the different regions slightly differs from the direct estimation from Equation (6.1). The prior distribution of the number of events observed in the signal D region is shown in Figure 6.14.

No large excess is observed in region D when comparing the data yield to the background predicted from the A/B/C regions assuming that the signal is absent in these regions.

In order to set limits, the likelihood function (Equation 6.2) is used to fit to all four A/B/C/D regions simultaneously.

The parameter μ_S can be expressed as the product of the total integrated luminosity, the signal cross-section $\sigma_H \times B(H \to aa \to \gamma\gamma gg)$, and the signal selection efficiency estimated in MC simulation and quoted in Table 6.8.

The effects of the uncertainties (Section 6.5) on the estimated number of signal events μ_S are studied using Asimov [346] pseudo-datasets generated for an expected

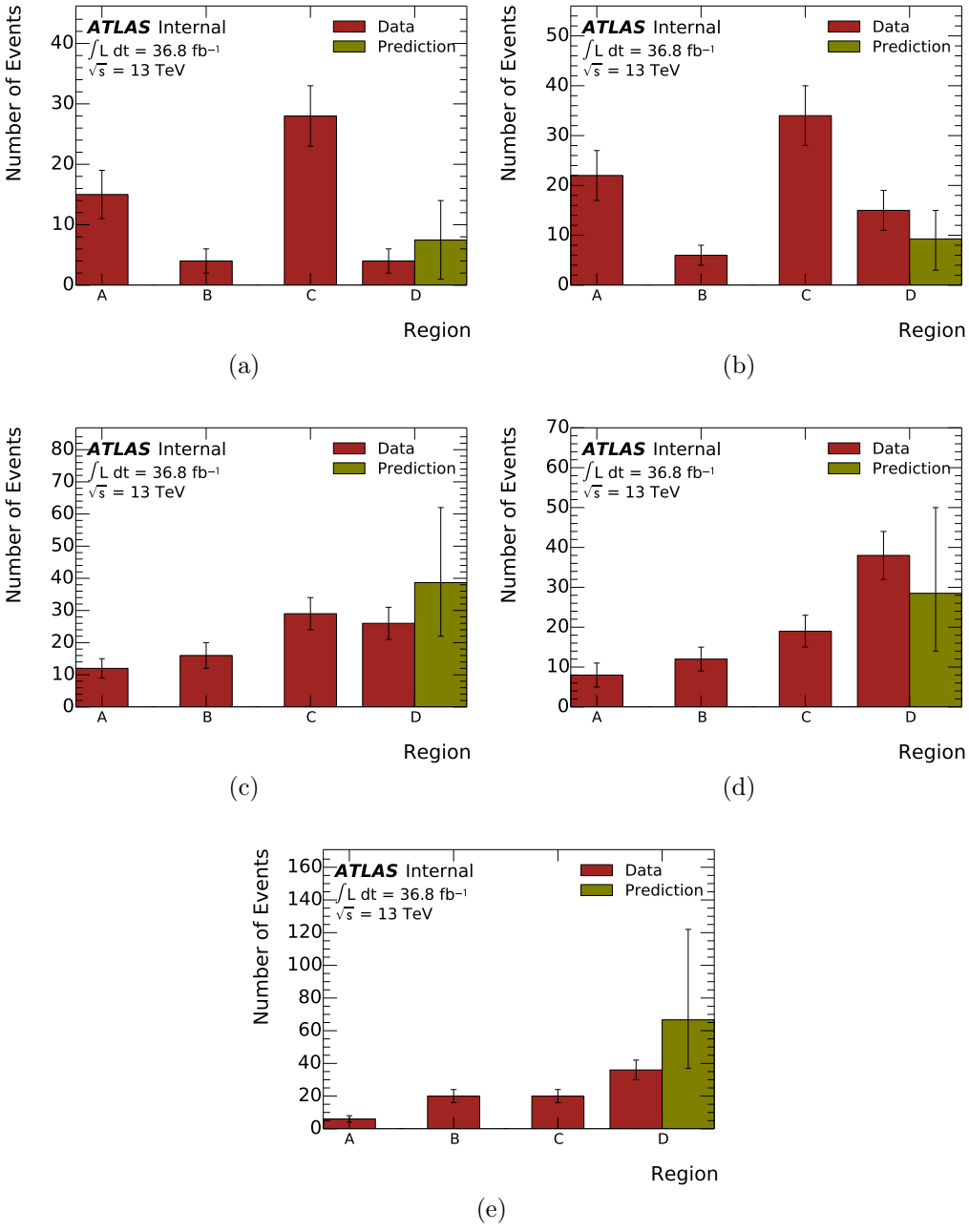


Figure 6.13: The observed number of events in each of the signal ABCD regions, as well as the predicted number of events in the D region under the background-only hypothesis. The error bars shown are purely statistical uncertainties (including the closure uncertainty). (a) Analysis regime 1. (b) Analysis regime 2. (c) Analysis regime 3. (d) Analysis regime 4. (e) Analysis regime 5.

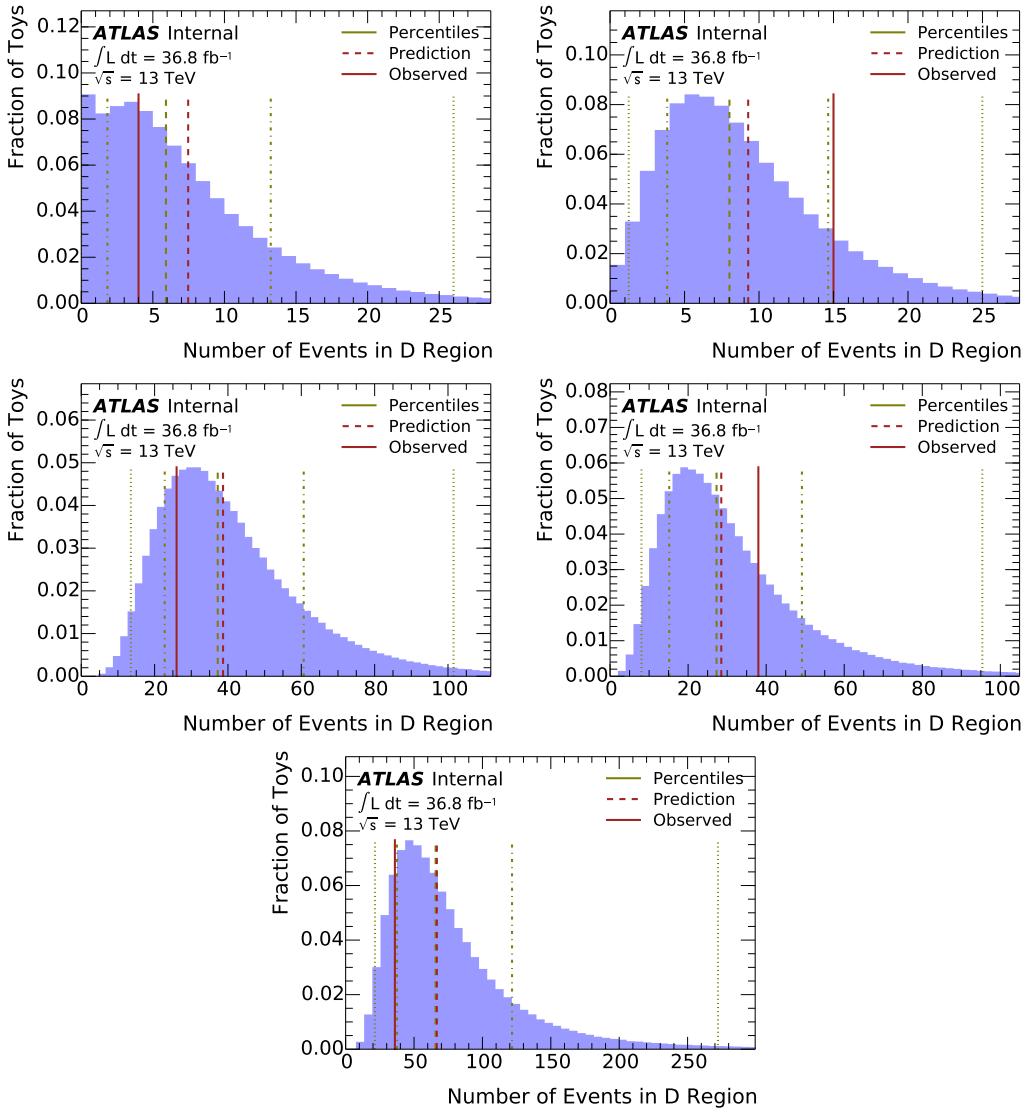


Figure 6.14: The blinded prior distribution of possible number of events in the signal D region, taking into account the Poisson uncertainty on the control regions A, B, and C. The distribution was generated with toys. In green, the percentiles in the distribution corresponding to the [-2,-1,0,+1,+2]σ points of a Normal distribution. In red, the predicted (dashed) and observed after unblinding (solid) number of events. (a) Analysis regime 1. (b) Analysis regime 2. (c) Analysis regime 3. (d) Analysis regime 4. (e) Analysis regime 5.

Table 6.10: Number of events observed in each of the A/B/C/D regions, the relative size of the closure uncertainty considered for each $m_{\gamma\gamma}$ regime, and the prediction for the number of background events in region D based on the control region yields. The median predicted background yield and its $\pm 1\sigma$ uncertainty in region D is also shown. The uncertainties in the prediction account for both the Poisson fluctuations of the number of events in the control regions and the closure uncertainty.

$m_{\gamma\gamma}$ regime	A	B	C	D	Relative closure uncert.	Predicted background yield
1	15	4	28	4	0.50	6^{+7}_{-4}
2	22	6	34	15	0.32	8^{+7}_{-4}
3	12	16	29	26	0.20	37^{+23}_{-14}
4	8	12	19	38	0.21	27^{+22}_{-12}
5	6	20	20	36	0.20	66^{+56}_{-28}

signal corresponding to the 95% CL upper limit obtained in this search (Table 6.13) and using the values of the background parameters maximising the likelihood in a fit to data which assumes no signal. Table 6.11 summarises the impact of each source of uncertainty varied by $\pm 1\sigma$ on the maximum-likelihood estimate for μ_S in each of the $m_{\gamma\gamma}$ regimes for an illustrative m_a hypothesis. The statistical uncertainty is the largest one for all regimes.

The best-fit values of the parameters of the likelihood function are given in Table 6.12. The probability that the data are compatible with the background-only hypothesis is computed for each $m_{\gamma\gamma}$ regime and no significant excess is observed. The smallest local p-value, obtained for the $m_{\gamma\gamma}$ regime 2 ($m_a \approx 30$ GeV), is approximately 4%. Since no significant excess is observed, an upper limit is derived at 95% CL. The expected and observed exclusion limits on μ_S are given in Table 6.13. This is related to the limit on the $pp \to H \to aa \to \gamma\gamma gg$ cross-section by appropriately normalising to the measured total integrated luminosity and selection efficiencies relative to the inclusive signal production obtained from the ggF and VBF MC samples (Table 6.8). The limit is also expressed relative to the SM cross-section for the Higgs boson, shown in Figure 6.15. Within a $m_{\gamma\gamma}$ analysis regime, limits are interpolated linearly in between simulated m_a values. Finally, for each mass point, the $m_{\gamma\gamma}$ regime that yields the best expected limit is used to provide the observed exclusion limit.

Table 6.11: Maximum fractional impact on the fitted μ_S from sources of systematic uncertainty estimated using Asimov datasets. The signal injected in the Asimov datasets corresponds to the observed upper limit quoted in Table 6.13.

	$m_{\gamma\gamma}$ regime (m_a)				
Source of Uncert.	1 (20 GeV)	2 (30 GeV)	3 (40 GeV)	4 (50 GeV)	5 (60 GeV)
Statistical	0.73	0.51	0.89	1.13	0.92
Closure	0.44	0.27	0.39	0.64	0.89
Modelling	0.35	0.34	0.46	0.42	0.65
Jet	0.58	0.38	0.25	0.90	0.71
Photon	0.06	0.05	0.10	0.12	0.13
Lumi and Pile-up	0.06	0.04	0.27	0.14	0.32

Table 6.12: Maximum-likelihood fit values for each of the free parameters of the likelihood function in each $m_{\gamma\gamma}$ regime for a relevant signal m_a hypothesis. The estimated uncertainties in the fit parameters assume that the likelihood function is parabolic around the minimum of the fit.

$m_{\gamma\gamma}$ regime	m_a [GeV]	μ_S	μ_{bkg}	τ_B	τ_C
1	20	-7±18	11±17	0.5±0.4	2.9±3.1
2	30	8±8	7±6	0.68±0.32	4.3±3.1
3	40	-30±80	60±70	0.35±0.19	0.67±0.33
4	50	22±28	16±23	0.5±0.4	0.9±1.0
5	60	-290±260	340±340	0.21±0.05	0.24±0.05

The limit is calculated using a frequentist CL_s calculation [345].

Table 6.13: Observed (expected) upper limits at the 95% CL, for each of the m_a values considered in the search. In each case, the $m_{\gamma\gamma}$ regime used to calculate the limits is also indicated. The uncertainties include both the statistical and systematic sources of uncertainty in the fit. $B = B(H \to aa \to \gamma\gamma gg)$ is defined as the branching ratio to the signal being targeted in this search.

$m_{\gamma\gamma}$ regime	m_a [GeV]	μ_S	$\sigma_H \times B$ [pb]	$\frac{\sigma_H}{\sigma_{SM}} \times B$
1	20	$10.8\left(10.4^{+4.6}_{-3.1}\right)$	$4.8\left(4.6^{+2.1}_{-1.4}\right)$	$0.086\left(0.082^{+0.037}_{-0.025}\right)$
1	25	$10.4\left(10.9^{+3.8}_{-2.5}\right)$	$1.9\left(2.0^{+0.7}_{-0.5}\right)$	$0.034\left(0.036^{+0.013}_{-0.008}\right)$
2	25	$28\left(25^{+8}_{-6}\right)$	$5.1\left(4.7^{+1.4}_{-1.1}\right)$	$0.092\left(0.084^{+0.026}_{-0.019}\right)$
2	30	$29\left(24^{+11}_{-6}\right)$	$3.1\left(2.6^{+1.1}_{-0.7}\right)$	$0.056\left(0.046^{+0.021}_{-0.012}\right)$
2	35	$27\left(22^{+9}_{-6}\right)$	$2.7\left(2.2^{+0.9}_{-0.6}\right)$	$0.049\left(0.040^{+0.016}_{-0.011}\right)$
3	35	$30\left(36^{+18}_{-9}\right)$	$2.7\left(3.2^{+1.6}_{-0.8}\right)$	$0.048\left(0.057^{+0.028}_{-0.014}\right)$
3	40	$31\left(39^{+19}_{-12}\right)$	$3.2\left(4.0^{+2.0}_{-1.2}\right)$	$0.058\left(0.073^{+0.035}_{-0.022}\right)$
3	45	$45\left(53^{+15}_{-20}\right)$	$6.3\left(7.5^{+2.1}_{-2.8}\right)$	$0.113\left(0.134^{+0.038}_{-0.050}\right)$
4	45	$74\left(68^{+16}_{-15}\right)$	$9.2\left(8.4^{+2.0}_{-1.9}\right)$	$0.166\left(0.152^{+0.036}_{-0.034}\right)$
4	50	$79\left(77^{+17}_{-16}\right)$	$9.0\left(8.8^{+2.0}_{-1.8}\right)$	$0.162\left(0.159^{+0.036}_{-0.032}\right)$
4	55	$73\left(69^{+11}_{-10}\right)$	$9.7\left(9.1^{+1.5}_{-1.2}\right)$	$0.173\left(0.163^{+0.026}_{-0.022}\right)$
5	55	$48\left(59^{+41}_{-19}\right)$	$5.5\left(6.8^{+4.7}_{-2.1}\right)$	$0.10\left(0.12^{+0.08}_{-0.04}\right)$
5	60	$67\left(81^{+24}_{-31}\right)$	$8.0\left(9.5^{+2.8}_{-3.6}\right)$	$0.14\left(0.17^{+0.05}_{-0.07}\right)$

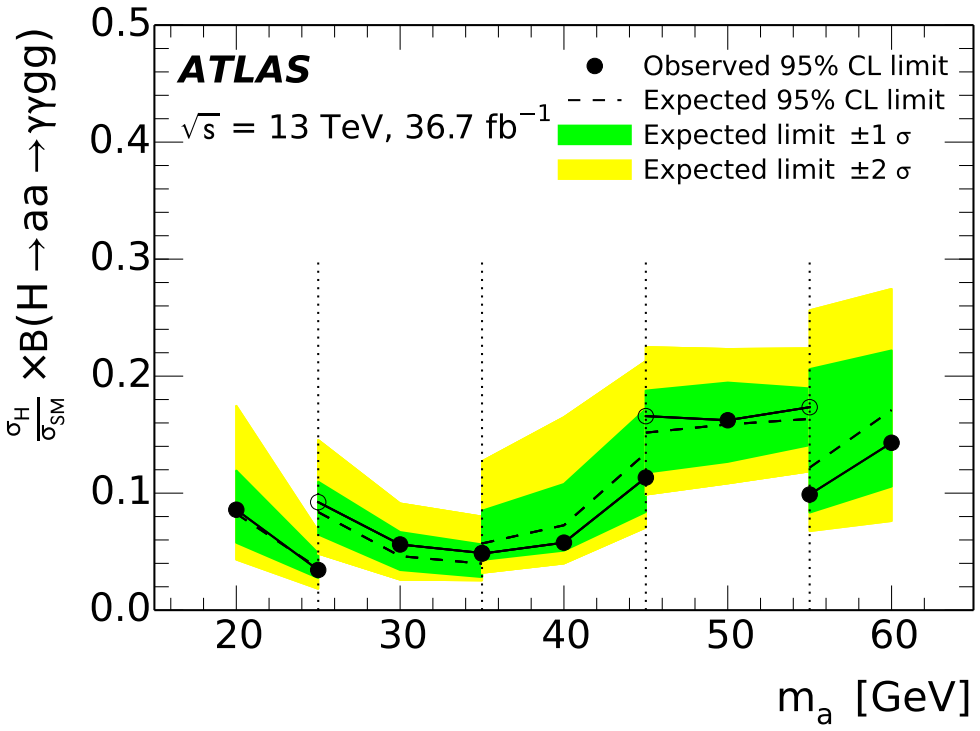


Figure 6.15: The observed (solid line) and expected (dashed line) 95% CL exclusion upper limit on the $pp \to H \to aa \to \gamma\gamma gg$ cross-section times branching ratio as a function of m_a, normalised to the SM inclusive $pp \to H$ cross-section [337]. The vertical lines indicate the boundaries between the different $m_{\gamma\gamma}$ analysis regimes. At the boundaries, the $m_{\gamma\gamma}$ regime that yields the best expected limit is used to provide the observed exclusion limit (filled circles); the observed limit provided by the regime that yields the worse limit is also indicated (empty circles).

6.7 Discussion

This analysis can be combined with the Run 1 search for $H \to 2a \to 4\gamma$ [335] to place a limit on the overall BR($H \to aa$). Figure 6.16 shows a quantitative comparison of the two analyses.

The two analyses have different sensitivities in different parts of the phase space, parameterized by $\rho = \text{BR}(a \to \gamma\gamma)/\text{BR}(a \to jj)$. The analysis presented in this Chapter is more sensitive for lower values of ρ, closer to SM-like couplings with photons and gluons; a typical value of ρ in a 2HDM model as discussed in [55] is shown for reference. The $H \to 2a \to 4\gamma$ search is more sensitive if there is a large BR to photons.

Both of these analyses are limited by statistics, therefore these bounds will improve with more data.

This search is notably limited to $m_a \geq 20$ GeV; the ATLAS Run 1 search for $H \to aa \to 4\gamma$ [335] similarly sets limits only for $m_a \geq 10$ GeV. As mentioned in Section 6.3.1 and Section 6.3.2, these lower bounds are due to the limitations set by using the diphoton trigger and the diphoton selection. For low m_a, the a particle is boosted and the decay products are collimated - this ruins the efficiency of the diphoton selection, both because of the way photon objects are constructed in the calorimeter at trigger level and because isolation is imposed on the photons in the offline object reconstruction. This inefficiency is discussed in detail in Appendix B. Thus, new techniques will be required to set limits on lower masses.

Reference [347] demonstrates a search for $X \to aa \to 4\gamma$ with X a high-mass scalar particle with mass between 200 and 2000 GeV (notably missing the Higgs mass $m_H \sim 125$ GeV) and a with mass between 100 MeV and up to between 2 and 10 GeV, depending on the X mass. There is a similar search [348] from Run 1 that sets limits for X with a mass between 110 and 150 GeV and m_a between 100 and 400 MeV. These searches are accomplished by taking advantage of the fact that in these regimes, the a is highly boosted, so that the photon decay products end up collimated in a single photon-like energy cluster and the trigger identities these as a single isolated photon. Substructure of this photon-like energy cluster is used offline

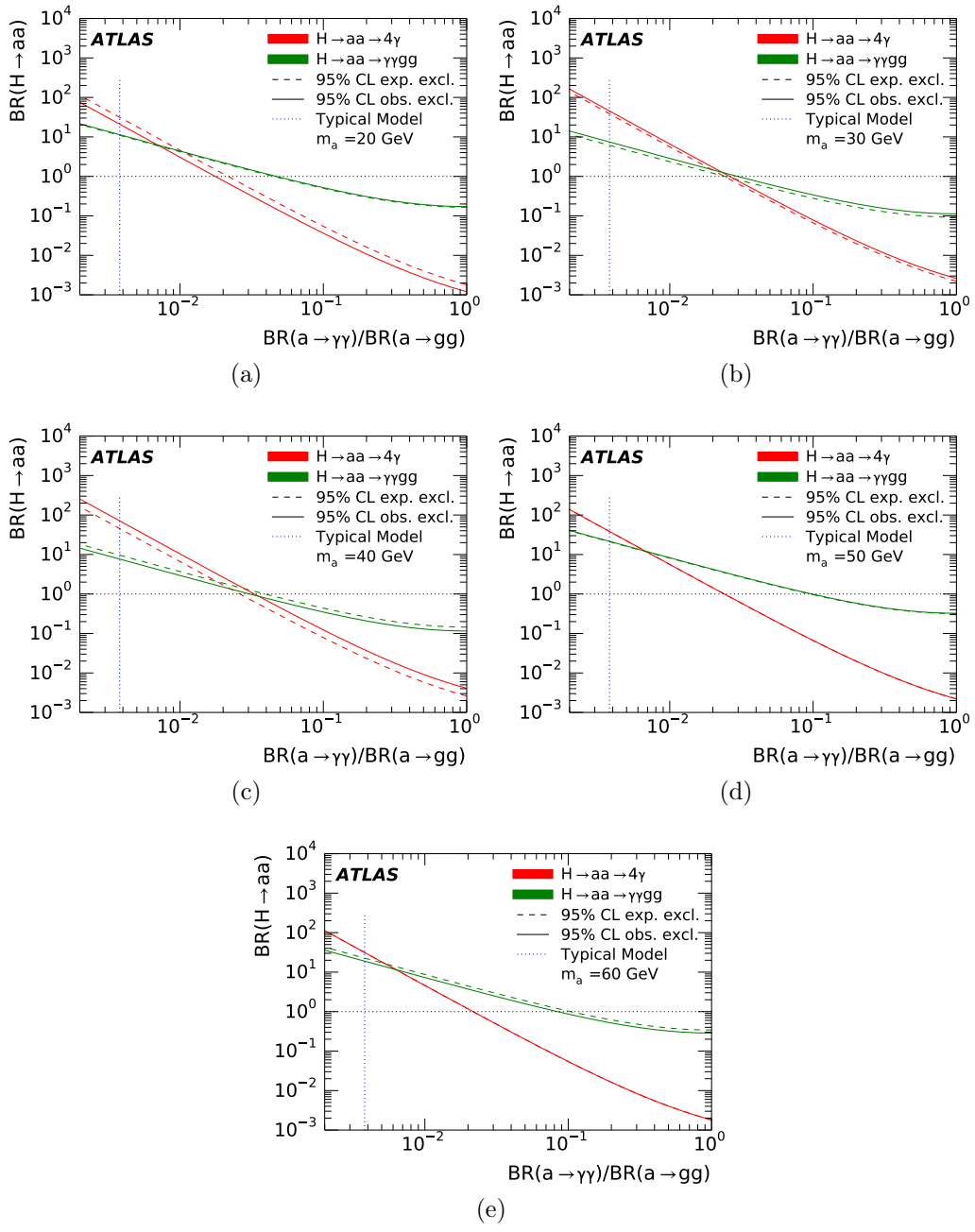


Figure 6.16: The exclusion limit placed on BR($H \to aa$) as a function of BR($a \to \gamma\gamma$)/BR($a \to jj$), and comparing the ATLAS 4γ analysis to this analysis. Also shown is a typical value for BR($a \to \gamma\gamma$)/BR($a \to jj$) in the NMSSM. (a) m_a =20 GeV. (b) m_a =30 GeV. (c) m_a =40 GeV. (d) m_a =50 GeV. (e) m_a =60 GeV.

to enhance the signal presence.

Appendix B proposes an already existing trigger and offline selection strategy that can be sensitive to $H \to aa \to gg\gamma\gamma$ with $m_a < 20$ GeV, and could enable a search for these low masses. Though not necessary to enable this search, the weak supervision analysis techniques used in Chapter 7 could be used to search for these signals while retaining limited dependence on the background simulation.

This search also presents many interesting areas of research in order to improve the sensitivity of the analysis for the resolved mass range $m_a \geq 20$ GeV. The event selection (Section 6.3.3) requires four jets, and this analysis is enabled by the use of these jets, in keeping with the overall theme of this Thesis. The main challenges present in the analysis include the correct identification of the VBF and signal jets; the VBF selection including the VBF m_{jj} selection and the leading VBF jet p_T; the resolution of the signal jet m_{jj}, which directly affects the resolution of $|m_{jj} - m_{\gamma\gamma}|$; and the final resolution of the signal $m_{jj\gamma\gamma}$.

The identification of the VBF and signal jets could be aided by using techniques intended to discriminate between jets initiated by quarks and those initiated by gluons [87–90].

Each of the remaining challenges would be better tackled by improving the energy and p_T resolution of the jets involved, since the invariant mass is a simple function of the individual jet energies and the angle between them[5]. When reconstructing the true energy of a jet, both the effects of the energy response of the calorimeter and the effect of other proton-proton interactions in the event (pile-up) must be taken into account. Chapter 8 presents work on studying in a rigorous framework the mathematical process of reversing the effects of the calorimeter response in order to access the originating parton energy [349]; and Chapter 9 presents work on using machine learning to further improve this process [350]. The Author has also worked on novel jet reconstruction algorithms which aim to reduce the effect of pile-up on the jet reconstructed energy [187]. By better reconstructing the energies of the jets that appear in the final state of this analysis, these efforts can improve the ultimate

[5]Actually, this would help with all the challenges mentioned, since the VBF selection also relies on the resolution of the VBF m_{jj}.

sensitivity of this search, in addition to being broadly applicable to other analyses than the one being studied here.

6.8 Conclusion

In summary, a search for exotic decays of the Higgs boson into a pair of new (pseudo)scalar particles, $H \to aa$, in final states with two photons and two jets is conducted using 36.7 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV recorded with the ATLAS detector at the LHC. The search for $H \to aa \to \gamma\gamma gg$ is performed in the mass range $20 < m_a < 60$ GeV and with additional jet requirements to enhance VBF-produced signal while suppressing the $\gamma\gamma$+jets background. No significant excess of data is observed relative to the SM predictions. An upper limit is set for the product of the production cross-section for $pp \to H$ and the branching ratio for the decay $H \to aa \to \gamma\gamma gg$. The upper limit ranges from 3.1 pb to 9.0 pb depending on m_a, and is mostly driven by the statistical uncertainties. These results complement the previous upper limit on $H \to aa \to \gamma\gamma\gamma\gamma$ and further constrains the BSM parameter space for exotic decays of the Higgs boson.

This search can be extended to the low-mass regime using novel techniques (Appendix B). In particular, the weak supervision analysis techniques used in Chapter 7 could be used to target this low-mass regime. The sensitivity of the search presented here, and the extension to the low-mass regime, can be improved utilizing the novel technique for jet calibrations proposed in Chapter 9, which builds upon the deep understanding of calibrations provided in Chapter 8.

Chapter 7

A Generic Data-Driven Resonance Search with Weak Supervision

This Chapter presents a search for dijet resonances enabled by new techniques with weak supervision. The results of this search are published in [351]. The search does not rely on a specific signal model hypothesis and is sensitive to generic final states of the form $A \to BC$, where all of A, B, C are massive and may be beyond the Standard Model, with $m_A \sim$ O(TeV) and $m_B, m_C \sim$ O(100 GeV). The B and C particles are each reconstructed as single large-R jets and the A particle is reconstructed as a resonance in the dijet invariant mass spectrum. The search uses a novel technique in weak supervision called classification without labels (CWoLa) to tag events with jets corresponding to possible signals with a neural network trained entirely in data. As the first demonstration of this technique, the neural network uses a reduced set of features of just the masses of the two jets to tag potential signals. As the neural network learns what to tag directly from data, this technique allows for avoiding the large trials factor associated with searching in the 3-dimensional feature space of $\{m_A, m_B, m_C\}$. The search uses the full Run 2 $\sqrt{s} = 13$ TeV pp collisions dataset of 139fb^{-1} gathered by the ATLAS detector at the LHC. Background-only p-values for the dijet invariant mass between 1.8 and 8.2 TeV are reported on and no significant evidence for a localized excess is found. In addition, new limits are set on the cross section of a variety of specific signal models with narrow-width B, C decaying

hadronically, with the limits ranging from 1-10 fb. These new limits correspond to improvements of up to 10 times existing limits from the inclusive dijet search at high m_B, m_C where the inclusive dijet search is less sensitive due to that search's use of small radius jets; limits at lower values of m_B, m_C where the inclusive dijet search is fully sensitive are improved by factors of up to 4.

7.1 Introduction

One of the simplest searches that can be done at any particle collider is to look for events in which two distinct objects are formed in the detector and search for excesses in the invariant mass spectrum formed by combining these two objects. In ATLAS the generic object formed in the detector is a jet (Chapter 3); almost all light standard model (SM) particles other than muons and neutrinos are reconstructed as jets in ATLAS, and more massive particles like the top quark, the vector bosons, and the Higgs boson have dominant or significant hadronic decays which are in turn reconstructed as jets, especially when boosted as is expected when decaying from a more massive resonant particle. ATLAS has an extensive history of searches for for these *inclusive* dijet resonances, including searches at $\sqrt{s} = 7$ [352–354], $\sqrt{s} = 8$ [355], and $\sqrt{s} = 13$ TeV [19, 267, 356, 357]. These searches are generically sensitive to beyond the SM (BSM) resonant decays to hadronic, electromagnetic, or electroweak objects reconstructed as jets in the detector, and therefore are one of the first searches performed when a collider reaches a new center-of-mass energy.

Dedicated searches for specific final states will always be more sensitive than the generic inclusive search, and ATLAS has an extensive program searching for di-τ [358, 359], di-b quark resonances [360, 361], di-top quark resonances [362], bt resonances [363], di-W/Z boson resonances [20, 364–367], VH resonances [368–370], di-Higgs resonances [371], and many more. These dedicated searches tag events based on the structure of the underlying jets in order to enhance the presence of the specific signal being searched for relative to the overwhelming background from events corresponding to SM QCD interactions. There are also complementary searches for the direct production of BSM particles decaying into SM particles and reconstructed as

a single large-R jet [267, 372, 373] using initial state radiation to boost the potential new particles. Searches for any combination of SM particles can be well-motivated by one or more standard (BSM) theory frameworks [374, 375], but not all combinations are currently covered by dedicated searches.

In addition, nearly all dijet resonance searches focus on decays to SM particles. Notable exceptions include the XH search [368,376], where the X is a Z-prime or H-like particle with unknown mass and the displaced jet search [377]. Of particular relevance to this Thesis is the photon-jet search [347] ($X \to aa, a \to$ photons), with the a particles boosted and reconstructed as single small-R jets; this search is sensitive to exactly the same class of models being targeted by the search presented in Chapter 6, and though that search does not exactly fall under the realm of dijet searches (since the decays of the a particles in that search are resolved into two separate small-R jets, and the X is constrained to be the Higgs boson), future iterations of that analysis will have exactly the dijet resonance topology (Appendix B.1). There is currently no search for $A \to BC$, where all of A, B, C can be BSM particles with possibly different masses.

This motivates the pursuit of a search which can be generically sensitive to these dijet topologies by tagging on the structure of the underlying jets which is complementary to these targeted efforts. A meta-search consisting of an ensemble of dedicated searches for each possible signal model would suffer from a very large trials factor due to the large space of possibilities. The trials factor accounts for the effect that since the probability of observing a large deviation from the background-only hypothesis increases with the number of orthogonal searches, a proportionally larger deviation is required from any one search in order to claim a new discovery - accounting for this effect (which is necessary) therefore makes the sensitivity of any one included search worse.

A novel analysis technique is proposed in [17, 18] based on weak supervision approaches called classification without labels (CWoLa) first introduced in [1] that can be sensitive to many particular topologies all at once without paying a large trials factor. The new technique trains a classifier directly on data, using features of the

jets within events, to distinguish between events at different values of the dijet invariant mass m_{JJ} which should be indistinguishable, if not for the presence of a signal resonant at one particular value of m_{JJ}. Since the classifier is trained directly on data, no specific signal model hypothesis is required, and the analysis is sensitive to a wide variety of new signal models. In addition, since the classifier is trained and tested on statistically independent datasets, there is no trials factor associated with scanning the feature space of the jets. This intuition is explored in detail and verified in Appendix C.1. The details of this technique and the application to this specific search are presented in later Sections.

This Chapter presents a search for a generic $A \to BC$ resonance where the B and C particles are each reconstructed as single massive large-R jets, and the A particle is reconstructed as a resonance in the dijet invariant mass spectrum, but other than that the properties of A, B, C are not specified, and in particular may or may not be BSM. Events collected by the ATLAS detector using the full Run 2 $\sqrt{s} = 13$ TeV pp collision dataset are used for the search, corresponding to 139fb^{-1} of data. The CWoLa technique is used by training a neural network to act as the classifier of potential signal. Though the original proposal for this kind of analysis [18] uses a broad class of features related to the substructure of the jet, this analysis uses a reduced feature space of just the masses $\{m_1, m_2\}$ of the two jets. As this is the first search of its kind, this reduction is intended to simplify the analysis and understand fully all the various effects. This simplification also allows for the setting of limits on specific signal models, since any limit setting requires the use of simulations of signal models, and the large-R jet mass is well-constrained and the uncertainties are well-understood [82] (a similar statement cannot be made about the general substructure of large-R jets). This search is therefore essentially a search in the 3-dimensional space of $\{m_A, m_B, m_C\}$, without paying the large trials factor associated with the scan in $\{m_B, m_C\}$.

The results of this search are compared to the results from the ATLAS inclusive dijet search [19], which as mentioned above is also generically sensitive to decays of the form $A \to$ jets, but is not specifically sensitive to massive decay products of the

A [1]. Also relevant are the results from the ATLAS all-hadronic diboson resonance search [20], which targets signal models of the form $A \to WW/WZ/ZZ$ and is thus sensitive to signals with $m_B \sim m_C \sim m_{W/Z}$ [2] , but has no sensitivity outside that range. There are also searches for direct production of B, C [267, 372, 373] when these are produced in association with photons or jets and their decay products are therefore collimated into a single large-R jet as in this analysis. However, those limits are orders of magnitude weaker than the ones studied here. This analysis does not present the limits on an exhaustive set of signal models, but rather shows the limits on a few specific models in order to demonstrate the sensitivity of the method. For some of these signal models the limits are improved by considerable factors, up to 10 times the existing limits at high m_B, m_C; for other signals, especially at low m_B, m_C, the technique sets about the same, worse, or no limits at all.

This Chapter is organized as follows. First, Section 7.2 introduces the data and simulated event samples. Section 7.3 gives a broad overview of the analysis, including the validation and blinding approach, and Section 7.4 goes over the steps of the analysis in detail. Section 7.5 demonstrates the full analysis pipeline using dijet Monte Carlo events to simulate the expected data. Section 7.6 demonstrates the full analysis pipeline using validation data with an inverted delta rapidity cut as a proxy for the expected data in the non-inverted signal region. The results of the unblinded analysis are presented in Section 7.7. Finally, some aspects of the analysis, including the future outlook, are discussed in Section 7.8 and the conclusions are provided in Section 7.9.

[1] In fact, the incusive dijet search uses small-R jets, so that the decay products of the B and C are not necessarily entirely contained within the jet. Because of this, as m_B and m_C get larger for fixed m_A, the limits from the incusive dijet search get worse.

[2] In fact this analysis is expected to be considerably more sensitive to these signals, since the substructure of the jets is used in addition to the masses in order to tag the W and Z particles.

7.2 Event Samples

7.2.1 Data

This analysis is performed using data from pp collisions provided from the Large Hadron Collider with $\sqrt{s} = 13$ TeV between 2015-2018, and collected with the ATLAS detector. The total integrated luminosity of this dataset is 139 fb^{-1}.

Data are collected using the lowest available unprescaled single large-radius jet trigger, which varies depending on the period during Run 2. In all years, the Level 1 trigger require a Level 1 jet with $p_T > 100$ GeV (`L1_J100`). In 2015 and 2016 the triggers in the high-level trigger fire on the untrimmed jet p_T; in 2015 the requirement was 360 GeV (`HLT_j360_a10_lcw_sub_L1J100`), and in 2016 this requirement was raised to 420 GeV (`HLT_j420_a10_lcw_L1J100`). From 2017 onward, the triggers in the high-level trigger use the trimmed jet p_T and apply a jet energy scale calibration, and the requirement on this p_T was 460 GeV (`HLT_j460_a10t_lcw_jes_L1J100`).

7.2.2 Simulation

As will be described in Section 7.3, this analysis is completely data-driven, both for training and testing. However, simulations of the expected background are used to validate the procedure and simulations of signals are used for setting model-dependent limits. Samples of Monte Carlo (MC) simulated dijet and multijet events are used to emulate the SM (Section 5.3). As the jet cross-section is orders of magnitude larger than electroweak processes, the consideration of these samples is sufficient to describe the data and all other processes are ignored.

Pythia v8.2 [244, 378] is used as the nominal MC generator for this analysis. Samples of $2 \to 2$ dijet events are simulated using the A14 tune [251] and NNPDF 2.3 [379] parton distribution function (PDF) set. The after-burner generator EvtGen [253] is used to model decays of heavy flavor hadrons. In order to fully populate a wide range of jet p_T, these samples are generated in slices of the particle-level $R = 0.6$ jet p_T.

Signal samples are generated using Pythia v8.2 and the process $W' \to WZ$,

where the W and Z masses are altered, and their widths are set to 0.1 GeV. These altered bosons are required to decay hadronically, except that top quark decays are switched off. As the parameter space of this simplified model is three-dimensional, we consider only a subset of all possibilities in order to demonstrate the method on some example signal models. The samples shown for the rest of this Chapter use $m_{Z'} \in \{3, 5\}$ TeV and $\{m_W, m_Z\} \in \{80, 200, 400\}$ GeV.

All simulation is reconstructed using a full detector simulation and superimpose simulated minimum-bias interactions to represent multiple pp interactions during the same or nearby bunch crossings (pile-up). The distribution of the average number of pile-up interactions in simulation is re-weighted during data analysis to match that observed in the Run 2 data.

7.3 Analysis Overview

7.3.1 Background: CWoLa and CWoLa Hunting

In a typical search for BSM at the LHC, some particular signal model is designated and an event selection is optimized to target this signal relative to the expected background using signal simulation and some proxy for the background (either simulation- or data-based). This is for example the strategy pursued in the search presented in Chapter 6, in which the event selection was chosen roughly manually to optimize the signal to background discrimination power. This strategy is also employed for example by the all-hadronic diboson resonance search [20], where dedicated W/Z hadronic jet taggers are trained as somewhat of a black box on simulation; the tagger is then calibrated in data, using a high-fidelity sample of W/Z jets (e.g. from $t\bar{t}$ events where one of the daughter W particles decays leptonically) to compare and correct the difference in estimated efficiency on W/Z jets in simulation and in data. This method of tagging signal can be broadly classified as "supervised learning", since the tagger is told what is signal and what is background at time of training.

There are two major drawbacks related to training a classifier in simulation and

using it in data. First, the simulation of the signal may not itself be entirely accurate, leading to a suboptimal selection even on the specific signal model in data. Data/simulation calibrations can remove bias in the tagger, in the sense that the efficiency on signal events estimated in simulation can be corrected to the efficiency estimated in data, but they do not in general remove this suboptimality[3]. Second, the trained tagger will only be sensitive to the specific signal model used for training, limiting sensitivity to other similar signals but with different values of the features used in training. There is therefore a motivation to train a classifier directly on data to tag whatever signals that may be found there.

There is of course the obvious issue that in data there are no labels on individual events telling the tagger whether the event comes from a signal or background process. However, a series of techniques using "weakly supervised learning" have been developed for high energy physics, which allow for learning even when these per-instance labels are not present [1, 380–382]. The basic idea for this method is shown in Figure 7.1. The setup of CWoLa is that there are two samples, each of which has a mixed composition of signal and background events. Crucially, the fraction of signal and background events between the two samples must be different; and while of course the signal and background events must differ from each other in the distributions of some features, the signal events and background events cannot differ among themselves between the two samples [4]. The insight of the CWoLa method is that a classifier can be trained to distinguish between the two mixed samples using any supervised learning technique. The resulting classifier learns to distinguish between the signal and background events. The main result of [1] is that it can be proved that this

[3] E.g., consider a case where a W jet tagger is being trained in simulation using the jet mass as a feature. Suppose further that the W jet mass is centered at 80 GeV in simulation but at 70 GeV in data (of course the jet mass is itself calibrated between data and simulation so this would not be the case, but the dedicated W/Z taggers do use certain features which are not themselves calibrated, and the jet mass is used as a simple tangible example feature to demonstrate the point). Then the tagger trained in simulation might derive an optimal selection centered around 80 GeV. When calibrating, the efficiency measured in simulation would be corrected to the actual efficiency in data on W jets with that selection, but the simulation-trained tagger would never find the true optimal selection, which would be centered around 70 GeV.

[4] Or rather, if there are differences between background events in Mixed Sample 1 and background events in Mixed Sample 2, those differences must be smaller than the difference in signal fraction between the two samples times the difference between signal and background.

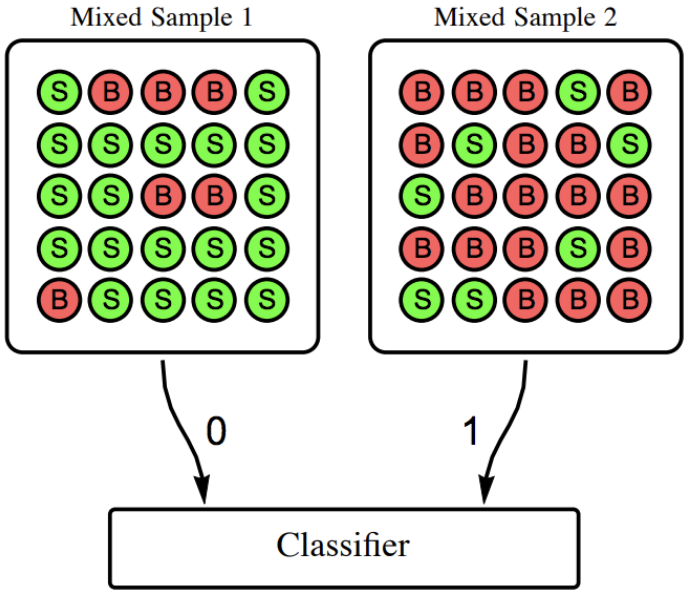


Figure 7.1: Schematic of basic idea behind classification without labels (CWoLa). There are two samples, each of which has a mixed but different composition of signal and background events. A supervised classifier is trained to distinguish between the two mixed samples. This same classifier is then used to distinguish between signal and background events. Figure sourced from [1].

procedure produces an optimal classifier for distinguishing signal from background in the asymptotic limit (enough data, flexible enough training model, etc.). The proof is straightforward and is provided in Appendix C.2.

It is worth noting the properties and limitations of this method. First, the features used to distinguish between signal and background events cannot or can only be very bad at distinguishing between background events in Mixed Sample 1 and Mixed Sample 2 (and similarly for signal events, though typically the signal presence is very small in one sample and basically zero in the other, so this is not as much of a concern). Second, the key insight is that it is okay if there is some signal in both mixed samples; however, learning is easiest when the fraction of signal between the two mixed samples differs the most.

Because the classifier is trained directly on data, it avoids the first major drawback of supervised learning mentioned above, since there is no possible simulation-data

difference, either in the signal events or in the background events.

As mentioned above, the second major limitation of supervised learning is that the trained classifier is only sensitive to one particular signal model, preventing the application to a generic search like the one being discussed here. The idea of "CWoLa hunting" [17, 18] is to use the CWoLa method to search for BSM physics without a particular signal model in mind, enabling a generic search for new physics. In fact, the method learns to tag whatever signal is present in the data, allowing the search to be generically sensitive to new signals that have different distributions of whatever features are used than the background.

The setup CWoLa hunting is shown schematically in Figure 7.2. There is some feature $\mathfrak{m}_{\text{res}}$ for which the background has a smoothly falling spectrum, while the signal is narrowly peaked (*resonant*) in that feature. Two mixed samples are formed: one with values of $\mathfrak{m}_{\text{res}}$ centered near the potential new resonance and one using events in neighboring regions of $\mathfrak{m}_{\text{res}}$. A neural network is trained to distinguish between these two mixed samples based on some other features, resulting in an overall score $\mathfrak{y}$. The first mixed sample clearly has a higher signal fraction than the second; and because the two mixed samples come from nearby regions of $\mathfrak{m}_{\text{res}}$, the difference in distributions of features in the background between the two mixed samples can be smaller than the fraction of signal times the difference in distributions between the signal and background. Therefore, because of the principles of CWoLa, the scores from the resulting network $\mathfrak{y}$ can distinguish between signal and background events. In the case that in fact there was no signal in the first mixed sample, the tagger will either learn to tag randomly or learn to tag based on the minute differences in the background between the samples. In either case, after tagging based on the output $\mathfrak{y}$, the combined spectrum of $\mathfrak{m}_{\text{res}}$ can be fit with a smooth function - if indeed there was a signal, there will be a bump in the $\mathfrak{m}_{\text{res}}$ spectrum; while if there was no signal, the background should remain smooth, even if the network learned to tag some differences in the background[5]. This process is then repeated for different regions of $\mathfrak{m}_{\text{res}}$ in order to be sensitive to a broad spectrum of BSM masses. One of the key aspects of this

[5]This is why it is key to use neighboring regions on both sides. The prevention of background sculpting is a major consideration in this analysis, as discussed in more detail in Section 7.3.2.

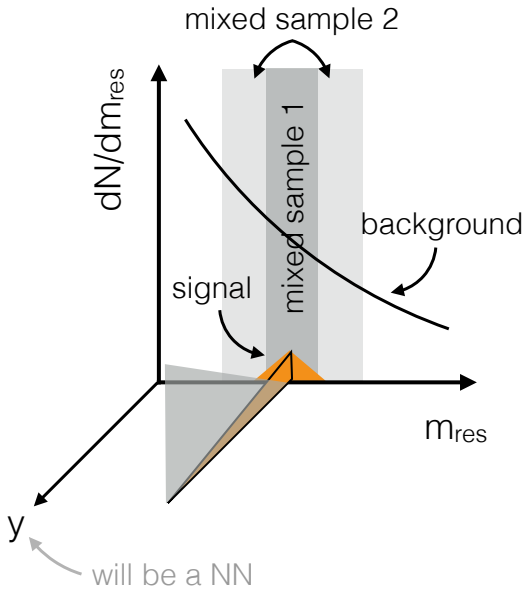


Figure 7.2: A schematic for the setup of CWoLa hunting. The background has a smoothly falling spectrum in some feature m_{res}, while the signal is narrowly peaked (resonant) in that feature. A network is trained to distinguish between a narrow region of m_{res} in which the signal lies from its neighboring regions. Because of the principles of CWoLa, the scores from the resulting network y can distinguish between signal and background events. Figure adapted from [2].

technique is a nested cross-validation procedure, which prevents overtraining in the case of no signal presence.

This and other key aspects are discussed in the following Section (7.3.2), which lays out the assumptions and specifics of this analysis based on the above ideas with $m_{res} = m_{JJ}$.

It should be noted that CWoLa hunting is not the only possible way of doing a model-independent bump hunt. A broad classification of existing and proposed searches based on their signal and background model independence is shown in Figure 7.3. As mentioned above, many analyses train a tagger to distinguish between signal and background simulations, and these analyses are therefore dependent on both signal and background modeling. There are a few analyses which train a tagger to distinguish between signal simulations and unlabeled data and are therefore background model independent, e.g. in the $\gamma\gamma$ channel of the observation of $t\bar{t}H$ [383],

using a validation region with low signal efficiency as the background sample for training. There are also some searches which directly test the compatibility of the observed data with the background-only hypothesis in a portmanteau test by comparing the observed data directly to the simulation of the background in ATLAS [5–7] and in CMS [3, 4]; these searches are therefore signal model independent but heavily dependent on background modeling. Recently, there have been proposals to enhance these searches with deep learning [384, 385].

CWoLa hunting falls into a final category which is both signal and background modeling independent. There are a few direct competitors to CWoLa hunting which also have these model independent properties. Many of these searches are based on autoencoders [8–13], but there are other unsupervised techniques based on nearest neighbor classification [386–388], probabilistic modeling [14], reweighted simulation [16], density estimation [15], and others [389]. These techniques must also be combined with a background estimation strategy, either in the form of a bump hunt, or using simulation, similarly to CWoLa hunting. The main difference between these techniques and CWoLa hunting is that they are *un*supervised, while CWoLa hunting is *weakly* supervised. I.e., the unsupervised methods simply find natural groupings or patterns in the data, and it may be the case that a new signal does not fit into those groupings or follow those patterns so that the unsupervised network can tag these as anomalies (however such an outcome is not guaranteed). The weakly supervised method, on the other hand, does see the signal during its training, if it exists, and so therefore can be biased towards better tagging the signal. Because of this difference, when signal rates are large enough such that the weakly supervised network can learn some information to tag the signal itself, the weak supervision method is superior. On the other hand, the unsupervised approaches can do better when there is very little signal.

It is important to point out that, while CWoLa hunting can be sensitive to a broad set of signal models, there are some assumptions made about the signal so that the method is not totally generic. First, there is an assumption that the signal is resonant in m_{res}, preventing sensitivity to wide or non-resonant signals. Second, there is a requirement that there are features of the signal events that can distinguish them

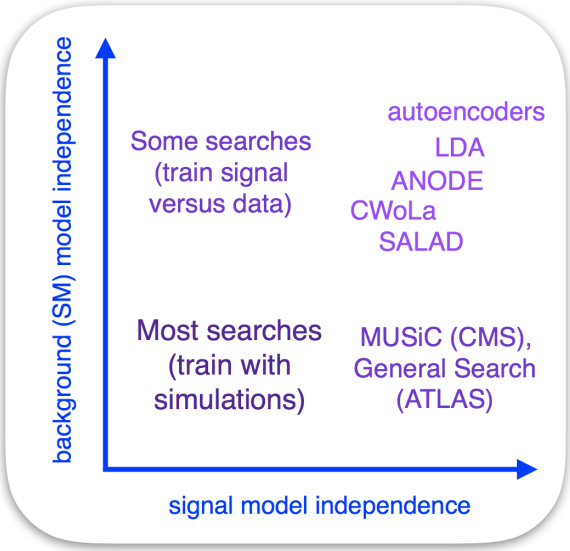


Figure 7.3: Classification of existing and proposed searches based on their signal and background model independence. The MUSiC (Model Unspecific Search for new physiCs) [3, 4] and General Search [5–7] are used in CMS and ATLAS, respectively. Methods based on autoencoders [8–13], LDA (Latent Dirichlet Allocation) [14], ANODE (ANOmaly detection with Density Estimation) [15], and SALAD (Simulation Assisted Likelihood-free Anomaly Detection) [16] can be considered to be direct competitors to CWoLa [1, 17, 18]. Figure sourced from [15].

from the background other than m_{res} (although if this were not the case the best search that could be done would be the inclusive fit in m_{res}, as there is no signal/background discrimination power). Third, as will be mentioned in Section 7.3.3, for the specific search using CWoLa hunting discussed here (with $m_{res} = m_{JJ}$), a non-trivial selection is placed on the rapidity difference between the two jets. This selection basically amounts to an assumption that the $A \to BC$ process being targeted by this search occurs via an s-channel process, and removes a large amount of the background due to t-channel QCD processes. It should be noted that this selection is also applied in the inclusive dijet search [19] and in the all-hadronic diboson resonance search [20], which are sensitive to similar signals and also aim to be inclusive in their sensitivity.

The search presented here, using CWoLa hunting with $m_{res} = m_{JJ}$, is the first application of these model-independent generic searches with LHC data.

7.3.2 Analysis Strategy

This analysis utilizes the concept of Classification Without Labels (CWoLa) described in Section 7.3.1 in order to be sensitive to a broad class of non-SM models of the form $A \to BC$, with A massive and on-shell, and the decay products of B and C reconstructed as large-R jets in the ATLAS calorimeter. In these models, the reconstructed dijet mass m_{JJ} will show a peak near the mass of the A, with width determined by the dijet mass resolution. There are two key assumptions that enable this analysis to be sensitive to these models.

1. **(Assumption 1)** One or both of the jets reconstructed from the B and C decays have features, e.g. in their substructure, which differ from the background composition (i.e., non-resonant dijet events from QCD and other SM processes), so that these jets can be tagged in order to increase the signal to background ratio of the signal m_{JJ} peak over the background shape.

2. **(Assumption 2)** After tagging on these features, the background m_{JJ} remains smooth, so that a fit to the background m_{JJ} spectrum does not indicate a potential discovery via a bump in the m_{JJ} spectrum.

In particular, clearly one of the features that differs between the signal and the background composition is m_{JJ} itself; if m_{JJ} is used as a feature, or if jet features that allow reconstruction of m_{JJ} are used, then Assumption 2 is violated, because after tagging on this feature the m_{JJ} spectrum in the background will no longer be smooth. Therefore some effort must be put into choosing features that will not violate Assumption 2, and into validating that the assumption is not violated once the features are chosen.

Under these assumptions, the outline of the analysis is as follows.

1. Select events in which at least two large-R jets are reconstructed, and in these events form the dijet invariant mass, m_{JJ}, and record some relevant jet features X.

2. Partition the m_{JJ} spectrum into discrete, monotonically increasing bins.

3. Label one of these bins the *signal region*, and the bins on either side in m_{JJ} space the *sideband regions*.

4. Train a neural network (NN) to distinguish between the signal region and the sideband regions based on the chosen features X. There are three possibilities of what the network learns, based on where the true signal lies relative to the signal region.

   (a) If the signal mass peak happens to lie in the signal region, then the signal fraction in the signal region will be higher than in the sideband regions, and the network will learn that it can identify some events in the signal region by tagging the true signal; i.e., the network will learn how to tag signal events. This process is what is referred to as CWoLa.

   (b) If the signal mass peak happens to lie in one of the sideband regions, then the signal fraction in the signal region will be lower than in the sideband regions, and the network will learn to identify some events in the sideband region by tagging the true signal; i.e., the network will learn how to anti-tag signal events.

(c) If the signal mass peak does not lie in the signal region or sideband regions, or if there is no signal at all, then the network will learn to distinguish between the background events in the signal and sideband regions based on the features X, as far as such correlations exist.

Note that, in Cases 4a and 4b, the network is sensitive to both the difference in the distribution of the features X between the signal and the background, and to the differences between the background in the signal region and the background in the sideband regions. In order to allow the network to focus on tagging the signal versus the background, the sideband regions are chosen to be adjacent to the signal region in kinematic space so that the difference in the features in the background between the signal and sideband regions is minimized.

5. Tag events by choosing events with high neural network output. In Case 4a, the neural network is a signal tagger, so the signal-to-background ratio increases, in particular increasing the significance of the signal peak over the background m_{JJ} shape. In Case 4b, the network anti-tags signal, and in Case 4c, the network tags only differences in the background between signal and sideband regions; in either case, the signal-to-background ratio after tagging is small or zero.

6. Fit to the m_{JJ} shape after tagging. Because of Assumption 2, the background spectrum is smooth after tagging. In Case 4a, there will be a bump in the m_{JJ} spectrum in the signal region due to the presence of signal, allowing the potential of a discovery of this signal. In Cases 4b and 4c, there will be little to no signal, so there will be no bump in the m_{JJ} spectrum, and the spectrum fit will not indicate the discovery of a signal.

7. Repeat Steps 3-6 with each possible signal region bin. If there is any true signal, then the signal will lie in one of the signal region bins, and Case 4a will be true in that bin, allowing discovery of this signal. If there is no true signal, then in each signal region bin Case 4c will be true, thus indicating no discovery over the whole m_{JJ} range.

Since there is a difference in the observed outcomes between the case where there is presence of signal and when there is no signal, limits can be set on specific signal models based on what would have been observed if those signals were present.

Since the network learns what features distinguish the signal from the background automatically, this analysis is sensitive to a wide range of possible signal distributions of these features, and thus the analysis can be sensitive to a wide range of signal models without paying the price of a trials factor that a dedicated search for each one of these signal models would entail (Appendix C.1).

Furthermore, since the network learns the features that distinguish the signal from the background directly from data, the tagger does not suffer from suboptimality due to differences between simulation and background, as could be the case with a tagger trained to distinguish between signal and background in simulation.

7.3.3 Blinding and Validation Procedure

Validating the analysis in a blinded region of the data is a challenge for this analysis. In a typical analysis, after running the analysis entirely in simulation, some validation region is defined with orthogonal cuts with low signal efficiency in order to validate the analysis. This validation region is used to either do data/MC comparisons to verify that the simulation is describing the background adequately or to perform in situ checks of the assumptions of the analysis. For example, in the search described in Chapter 6, the background is estimated in a data-driven way with an A/B/C/D matrix of regions; the highest signal efficiency is expected in region D. The assumption of the A/B/C/D method (that the ratio of events in B/A is the same as in D/C in the background) is verified both in simulations of the background and in a data validation region, inverting the $m_{gg\gamma\gamma}$ cut.

In this analysis, the original event selection (Analysis Step 1) is inclusive and minimal, in order to be sensitive to a broad class of models. The event selection learned by the NN depends on the data, so there is no way to construct an orthogonal signal-deficient region ahead of time. In Analysis Step 5, only events with some high NN score are chosen, so one idea is to use the events with lower NN scores for

validation. However, this does not clearly work - if selecting only events that the NN assigned *low* scores, then in Case 4b, where the network has anti-tagged the signal, this selection would actually have a high signal efficiency, defeating the purpose of the validation region. Another option is to select only events with *average* scores, i.e. close to the median NN output. However these events are exactly those for which the NN has decided it cannot distinguish between the signal and sideband regions, so that even if Assumption 2 were being violated in general by the NN selection, for these events the Assumption would *not* be violated, again defeating the purpose of a validation region. These ideas and others for constructing a validation sample are explored in Appendix C.3.

In light of these considerations, the event selection is given one additional non-trivial selection, a maximum rapidity difference between the two jets in the event; and the validation region is defined to be the inverse of this. This rapidity selection is also present in the inclusive dijet search [19] and in the all-hadronic diboson resonance search [20], two analyses which are sensitive to similar signals and also aim to be inclusive in their searches. This selection both suppresses contributions from t-channel QCD dijet production, while remaining efficient on $A \to BC$ signals produced in the s-channel. The distribution of the rapidity difference in the background simulation can be seen in Figure 7.4[6], and the distributions in a representative signal model can be seen in Figure 7.5.

The validation and unblinding procedure is then as follows:

1. Run the full method in simulation. These results are presented in Section 7.5. Two challenges with this validation are (i) we have less simulation than data and (ii) the simulated events are weighted. The latter can be challenging for our training procedure and is not an issue for the real data version of the analysis. This step is referred to as the "simulation" or "MC" analysis.

2. Proceed to run the full method on data where the rapidity gap cut is inverted. Any s-channel model will have a reduced cross-section while the background

[6]Note that there is a selection of $|\eta| < 2$ for large-R jets, so the maximum rapidity difference is $|y_1 - y_2| < 4$ (since $|y| \leq |\eta|$).

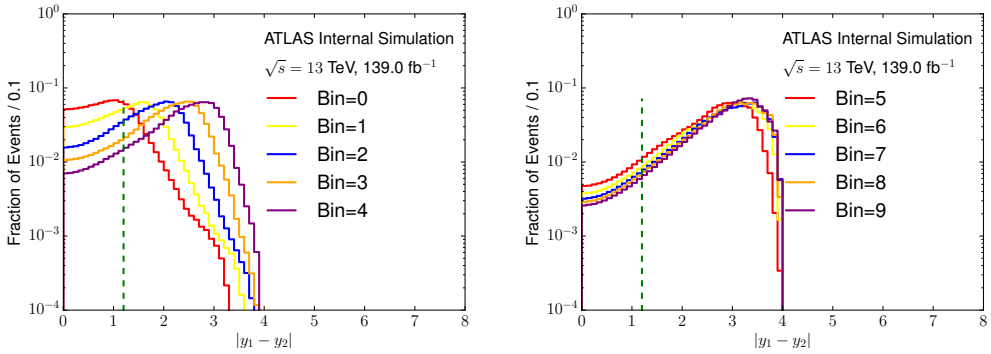


Figure 7.4: Distribution of $|y_1 - y_2|$ in the background simulation, broken down by the m_{JJ} region, as described in Section 7.4.2. (a) m_{JJ} regions 0-4; (b) m_{JJ} regions 5-9. The green line indicates the cut at 1.2 (Section 7.4.1).

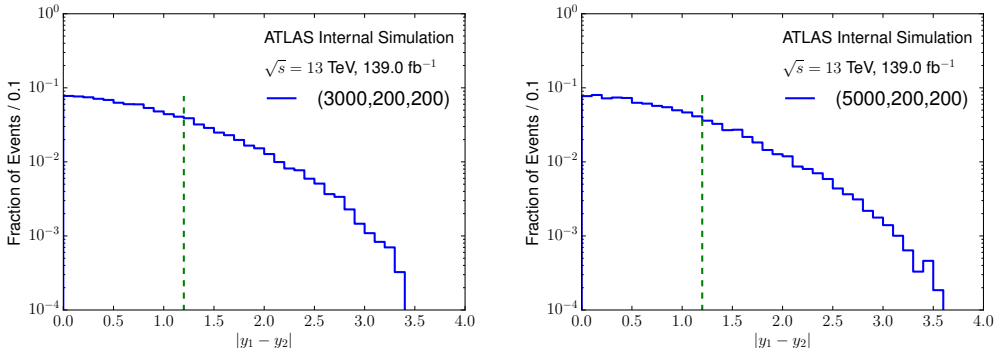


Figure 7.5: Distribution of $|y_1 - y_2|$ for a signal model with (a) $(m_A, m_B, m_C = 3000, 200, 200$ GeV); (b) $(m_A, m_B, m_C = 5000, 200, 200$ GeV). The green line indicates the selection at 1.2 (Section 7.4.1).

rate increases. This is discussed in Section 7.6 and allows us to run the full procedure on data where we do not expect to see any s-channel signal. This step is referred to as the "inverted rapidity cut", "data validation", or simply "validation" analysis, as it serves as the main validation in data before the full unblinded analysis.

3. Run the method on the full dataset (with the uninverted rapidity gap cut). The results of this are reported in Section 7.7. This step is referred to as the "uninverted", "unblinded", "signal selection", or simply "full" analysis.

The development of this analysis occurred by progressing through each of these steps in turn, and though overall most of the analysis remained the same between each of the steps, naturally some details changed after learning things at various steps of the validation. The details of the analysis outlined in Section 7.4 apply to each of these steps, but are clear about where they are different. In particular, the fitting process (Analysis Step 6) went through multiple iterations before settling on the final procedure. In general, doing a parametric background fit on large datasets tends to be a key challenge; this challenge was compounded due to the effective statistics of the background in each of these steps, which is expanded on below.

The effective statistics of the inverted sample compared with the nominal sample and the simulation are presented in Figure 7.6[7]. The bins presented there are those defined in Section 7.4.2, and given in Table 7.3. For the simulation, the effective statistics is defined as the inverse of the sum of the squares of the normalized event weights. For the central dataset, the MC statistics are relatively uniform in m_{JJ} and there are only more effective events past the third bin, with boundaries at $1.90 \leq m_{JJ} < 2.28$ TeV. Even in the fourth bin ($2.28 \leq m_{JJ} < 2.74$ TeV), there are events with a large weight, which makes the training and fitting unrealistic.

For the inverted dataset, the problem is much more severe because for a fixed m_{JJ}, lower p_T jets can contribute compared to the central dataset. For the full Run 2 statistics, the MC has the same number of weighted events only by the eighth bin

[7]Also given for good measure are the statistics for 3.2fb^{-1}, which corresponds to the 2015 LHC dataset.

$(4.73 \leq m_{JJ} < 5.68$ TeV).

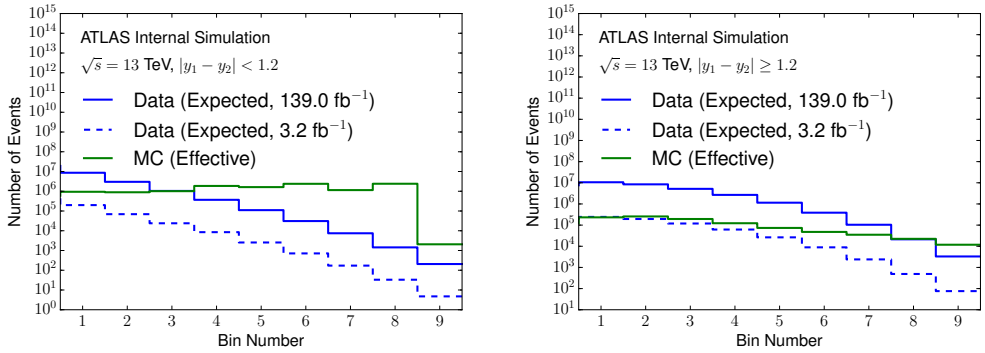


Figure 7.6: The effective statistics for the nominal rapidity difference (a) and the inverted one (b) for MC and data and for different integrated luminosities.

The inverted dataset has about ten times more events than the nominal rapidity difference dataset above about 3 TeV, which can be seen in Figure 7.7. In order for the inverted dataset to have comparable statistics to the nominal dataset for testing our methodology, random sampling without replacement is used to reduce the dataset. The random sampling rate is determined from a fit to the ratio of efficiencies for the two rapidity differences cuts (also shown in Figure 7.7) [8].

As will be mentioned in Section 7.4.6, each analysis imposes a cut of $m_{JJ} > 1.8$ TeV for the fitting. For the simulation analysis, as mentioned above, the effective MC statistics are only reliable above around that value. For the data analyses, this cut results mainly from the finding that the fitting procedure simply is not sufficient to describe the background shape for the high numbers of events seen below that cut[9]. However, it is also the case that the inverted rapidity selection provides multiple copies of independent datasets to test the analysis when the ratio of efficiencies is $< \frac{1}{2}$; in

[8] In principle, by looking at this distribution, we have unblinded the full dataset with no NN cut. However, we have not examined this distribution in detail and have not performed any statistical tests on the quality of the fit. It actually may be easier to do the bump hunt in this ratio, which is relatively flat, but this is left for future studies.

[9] A more sophisticated (possibly non-parametric) fitting procedure may be able to extend the search to lower values in the future.

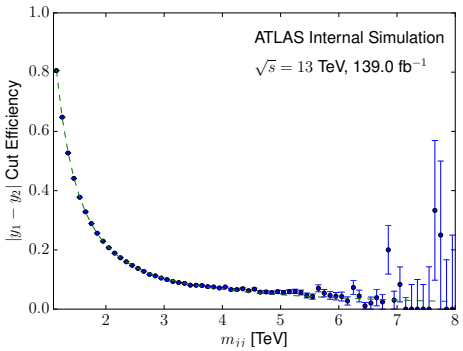


Figure 7.7: The ratio of efficiencies for the two rapidity differences cuts. A fit shown with a dashed line is the sum of two power law functions. Note that these are data plots and not simulation.

particular, the ratio of efficiencies is $< \frac{1}{3}$ (so 3 statistically independent copies) for $m_{JJ} >\sim 1.8$ TeV, which was useful for testing the statistical properties of the fitting.

7.4 Analysis Details

7.4.1 Event Selection

As mentioned in Section 7.2.1, data are collected using the lowest available unprescaled single-jet trigger.

This analysis uses anti-k_t, $R = 1.0$ jets reconstructed from topoclusters with the local cluster weighting scheme (LCW) [173]. The jets are trimmed [77] using using the parameters $f_{cut} = 5\%$ and $R_{sub} = 0.2$. A Monte Carlo-based particle-level calibration is then applied to the jets used in this analysis [82]. It corrects on average the reconstructed mass and p_T of the jets to their particle-level values. The *combined mass* measurement is used, which combines measurements from the tracking system and the calorimeter; including this information has been shown to have better mass resolution, especially at high jet p_T [201].

The offline selection that is made on top of the selection from the derivation is intended to be as fully inclusive as possible to prospective signal models, while remaining on the plateau of the turn-on for the trigger.

Each event is required to have at least two large-R jets with calibrated $p_T > 200$ GeV [10] and $|\eta| < 2.0$, and with at least one such jet with calibrated $p_T > 500$ GeV. This offline threshold has been shown to be fully efficient with respect to the trigger selection [390]. There is no explicit lepton overlap removal or veto.

Only the two jets with highest p_T in the event are used in this analysis, and the remaining jets are ignored. For each jet $j = 1, 2$, the four-momentum p_j^μ is recorded, in particular the jet mass $m_j^2 = (p_j^\mu)^2$. As mentioned in Section 7.3.3, a selection is placed on the rapidity difference, $|y_1 - y_2| < 1.2$.

The jets are ordered by their mass, so that $m_1 \geq m_2$. In order to suppress background and therefore increase sensitivity to massive signal objects, a minimum cut $m_1, m_2 > 30$ GeV is applied [11] . Also, in order to make the kinematic region in which the learning is applied definite, a maximum selection $m_1, m_2 < 500$ GeV is applied. The reasoning behind this selection is explained in more detail in Section 7.4.3.

[10] Jet calibrations for large-radius jets are only derived for $p_T > 200$ GeV.
[11] In addition, jet calibrations for large-radius jets are only derived for $m > 30$ GeV.

Typically the dijet invariant mass m_{JJ} would be defined as in terms of the sum of the four-momenta of the two jets, $m_{JJ}^2 = (p_1^\mu + p_2^\mu)^2 = m_1^2 + m_2^2 + 2(E_1E_2 - \vec{p}_1 \cdot \vec{p}_2)$, with $E^2 = |\vec{p}|^2 + m^2$ and $|\vec{p}| = p_T \cosh(\eta)$. In this analysis the dijet invariant mass is defined slightly differently: in order to reduce correlations between the features used in the neural network (m_1, m_2) and the final m_{JJ} value, the dijet invariant mass is formed by setting all m_i to zero: $m_{JJ}^2 \equiv 2(|\vec{p}_1||\vec{p}_2| - \vec{p}_1 \cdot \vec{p}_2)$. This new definition removes all correlations between m_{JJ} and m_1, m_2 except for those arising from indirect correlations between the jet m and the jet p_T and η. In practice this is a very small change, because the jets in this analysis typically have energies much larger than their masses, and so the new definition gives almost the same value for the invariant mass[12]. As will be discussed in Section 7.4.2, a selection is applied on the dijet invariant mass of $1.1 \leq m_{JJ} < 8.17$ TeV. As will be discussed in Section 7.4.6, effectively a selection of $1.8 \leq m_{JJ}$ TeV is applied as the fitting range.

These selections are summarized in Table 7.1.

Table 7.1: Jet selection. The m_{JJ} selection is indicated for the m_{JJ} bins (Section 7.4.2) and also for the fitting range (Section 7.4.6).

Observable	Selection
p_T	> 500 GeV (≥ 1 jet), > 200 GeV (≥ 2 jets)
$\|\eta\|$	< 2.0
$\|y_1 - y_2\|$	< 1.2
m	> 30 GeV, < 500 GeV
m_{JJ} (bins)	≥ 1.1 TeV, < 8.17 TeV
m_{JJ} (fitting)	≥ 1.8 TeV

The cutflow of these selections on the signal samples is given in Table 7.2.

[12]For example, an event with back-to-back jets and $m_{JJ} = 1.8$ TeV under the new definition would have $|\vec{p}_1| = |\vec{p}_2| = 0.9$ TeV. If each of these jets had a mass of 500 GeV, then the standard m_{JJ} would be ~ 2.06 TeV, a $\sim 15\%$ change. This is less than the dijet mass resolution of $\sim 20\%$ (Figure 7.11). In the background jets with that high mass are very rare (Figure 7.10), and with more typical values of < 200 GeV the change in m_{JJ} is $< 2\%$.

Table 7.2: Signal sample efficiency with selections up to and including given selection. All selections are given in Section 7.4.1. The jet selections are summarized in Table 7.1. The m_{JJ} selection is included for the binning selection of $1.1 \leq m_{JJ} < 8.17$ TeV and for the additional selection from the fitting of $1.8 \leq m_{JJ}$ TeV.

(m_A, m_B, m_C) [GeV]	Trigger	p_T	$\lvert\eta\rvert$	$\lvert y_1 - y_2\rvert$	m	m_{JJ} (bins)	m_{JJ} (fit)
(3000,80,80)	0.90	0.85	0.85	0.64	0.63	0.62	0.59
(3000,80,200)	0.91	0.85	0.85	0.63	0.62	0.61	0.58
(3000,80,400)	0.92	0.81	0.81	0.61	0.60	0.58	0.56
(3000,200,200)	0.93	0.87	0.87	0.64	0.63	0.63	0.59
(3000,200,400)	0.95	0.84	0.84	0.62	0.61	0.60	0.58
(3000,400,400)	0.95	0.82	0.82	0.62	0.60	0.59	0.57
(5000,80,80)	0.68	0.58	0.58	0.44	0.44	0.42	0.34
(5000,80,200)	0.70	0.59	0.59	0.45	0.44	0.42	0.34
(5000,80,400)	0.78	0.58	0.58	0.44	0.43	0.40	0.34
(5000,200,200)	0.75	0.63	0.63	0.48	0.47	0.45	0.37
(5000,200,400)	0.82	0.65	0.65	0.49	0.48	0.45	0.39
(5000,400,400)	0.88	0.68	0.68	0.50	0.48	0.47	0.43

7.4.2 Binning

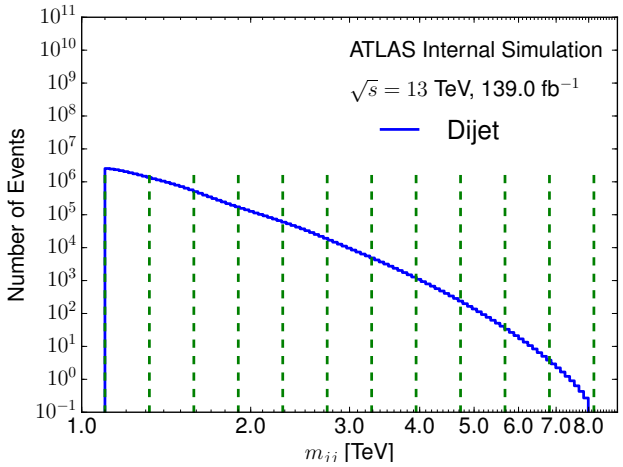
As mentioned in Analysis Step 2, the m_{JJ} spectrum is binned in order to derive the signal and sideband regions. The size of the binning is chosen to be 20%, guided by the dijet mass resolution of the signal models. The minimum m_{JJ} value is 1.1 TeV, due to the jet and trigger selection, and the maximum value is at 8.17 TeV (the right edge of bin 10 with 20% bin size), above which few events are expected to be observed. Since every event is required to lie in an m_{JJ} bin, effectively an event selection is therefore applied for $1.1 \leq m_{JJ} < 8.17$ TeV. The bin definitions are given in Table 7.3.

Table 7.3: m_{JJ} bin definitions.

Bin	Definition
0	$1.10 \leq m_{JJ} < 1.32$ TeV
1	$1.32 \leq m_{JJ} < 1.58$ TeV
2	$1.58 \leq m_{JJ} < 1.90$ TeV
3	$1.90 \leq m_{JJ} < 2.28$ TeV
4	$2.28 \leq m_{JJ} < 2.74$ TeV
5	$2.74 \leq m_{JJ} < 3.28$ TeV
6	$3.28 \leq m_{JJ} < 3.94$ TeV
7	$3.94 \leq m_{JJ} < 4.73$ TeV
8	$4.73 \leq m_{JJ} < 5.68$ TeV
9	$5.68 \leq m_{JJ} < 6.81$ TeV
10	$6.81 \leq m_{JJ} < 8.17$ TeV

The distribution of m_{JJ} in the background with these bin definitions is shown in Figure 7.8. It can be seen that the distribution of m_{JJ} in the background before any cuts is smooth, with no bumps indicating possible discoveries.

The distributions of m_{JJ} in a variety of signal models are shown in Figure 7.9. It can be seen that the mass of the signal object, m_A, is well-reconstructed as m_{JJ}, with significant tails at lower m_{JJ} values. The resolution of m_A with m_{JJ} is roughly 20%, which guides the sizing of the binning, as mentioned above.



(a)

Figure 7.8: Distribution of m_{JJ} in the background MC. The green lines indicate the bin edges of m_{JJ} regions 0-10.

One thing to note about this analysis is that the binning is fixed. A possible issue with this is that for a signal that is found close to the edge of a bin, the NN may not learn to tag that signal because it has presence in both that bin and its sideband. This effect is mitigated somewhat due to two factors: (1) Because the background is steeply falling, there is only a small region of m_A in which the signal fraction in both a signal bin and its sideband are comparable; and (2) the NN is trained on both sidebands simultaneously, so that the signal should still be tagged when distinguishing between the signal region and the combined sidebands even if it lies on the edge. A study of this effect is presented in Appendix C.6.

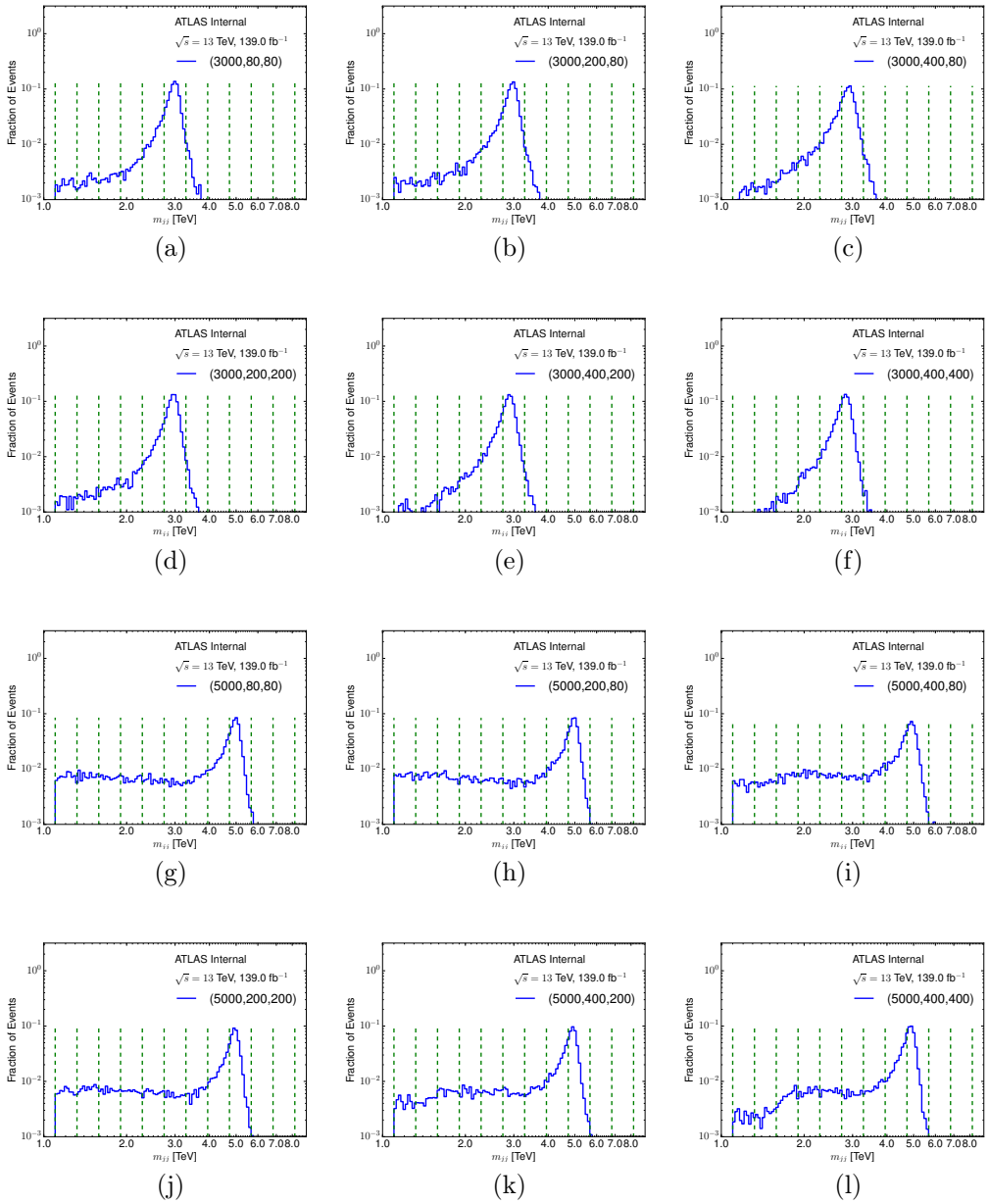


Figure 7.9: Distribution of m_{JJ} for a signal model with:
(a) $(m_A, m_B, m_C = 3000, 80, 80$ GeV); (b) $(m_A, m_B, m_C = 3000, 200, 80$ GeV);
(c) $(m_A, m_B, m_C = 3000, 400, 80$ GeV); (d) $(m_A, m_B, m_C = 3000, 200, 200$ GeV);
(e) $(m_A, m_B, m_C = 3000, 400, 20$ GeV); (f) $(m_A, m_B, m_C = 3000, 400, 400$ GeV);
(g) $(m_A, m_B, m_C = 5000, 80, 80$ GeV); (h) $(m_A, m_B, m_C = 5000, 200, 80$ GeV);
(i) $(m_A, m_B, m_C = 5000, 400, 80$ GeV); (j) $(m_A, m_B, m_C = 5000, 200, 200$ GeV);
(k) $(m_A, m_B, m_C = 5000, 400, 20$ GeV); (l) $(m_A, m_B, m_C = 5000, 400, 400$ GeV).

7.4.3 Features

In Analysis Step 4, some features X are designated which, as per Assumption 1, can be used to distinguish signals from the background. In this analysis, only the masses of the two jets, $X = \{m_1, m_2\}$, are used as features.

The distributions of m_1 and m_2 in the background using the nominal simulation are plotted in various m_{JJ} regions in Figure 7.10. The distributions of m_1 and m_2 in a variety of signal models are plotted in the m_{JJ} region which is most efficient on that signal in Figure 7.11. It can be seen that the masses of the signal objects, m_B and m_C, are well-reconstructed as m_1 and m_2.

Since m_1 and m_2 peak around the values of m_B and m_C in the signal, while they have smoothly falling spectra in the background, these features satisfy Assumption 1, and therefore can be used for the training.

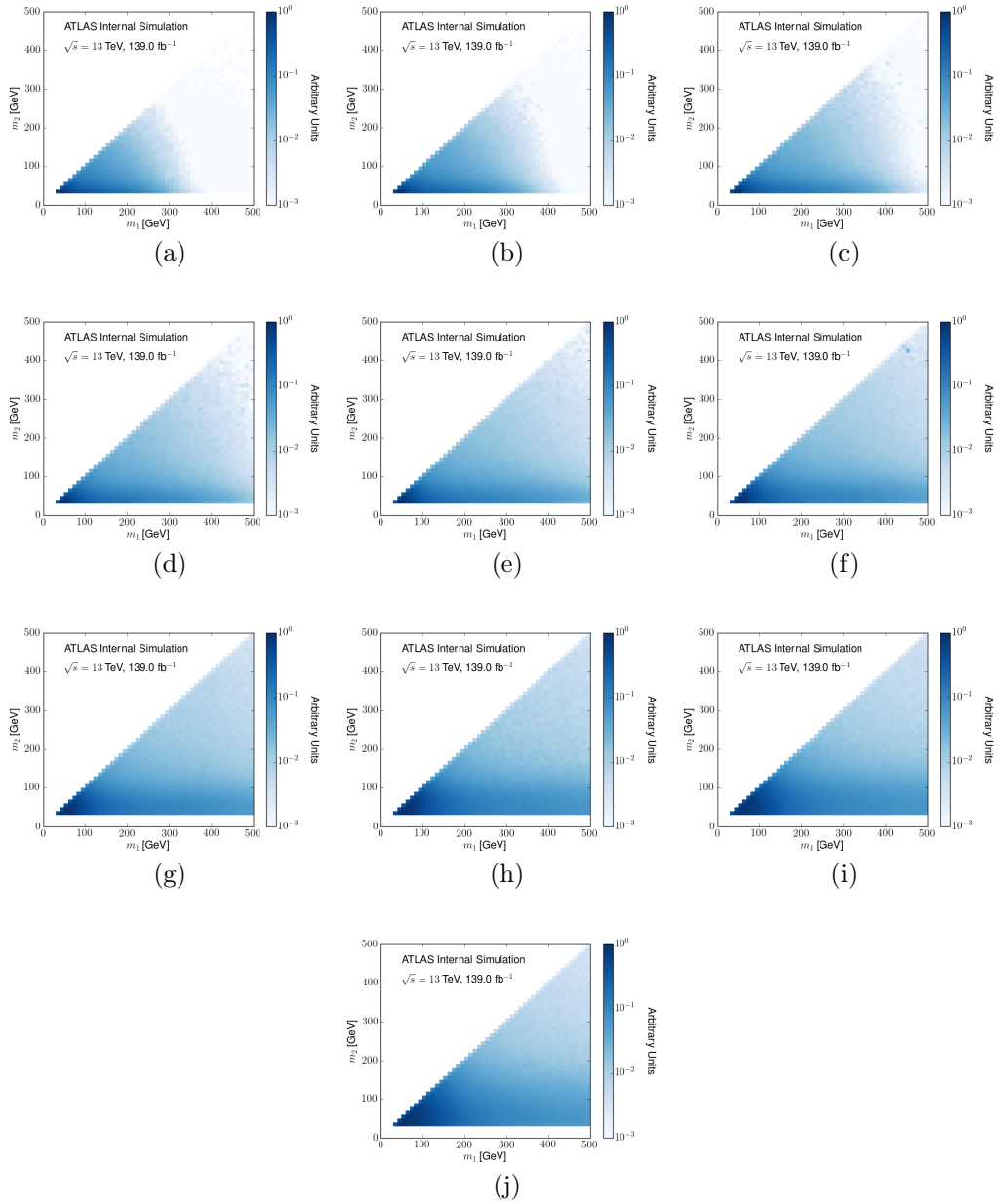


Figure 7.10: Distribution of m_1 and m_2 in the background simulation in m_{JJ} regions (a) 0; (b) 1; (c) 2; (d) 3; (e) 4; (f) 5; (g) 6; (h) 7; (i) 8; (j) 9.

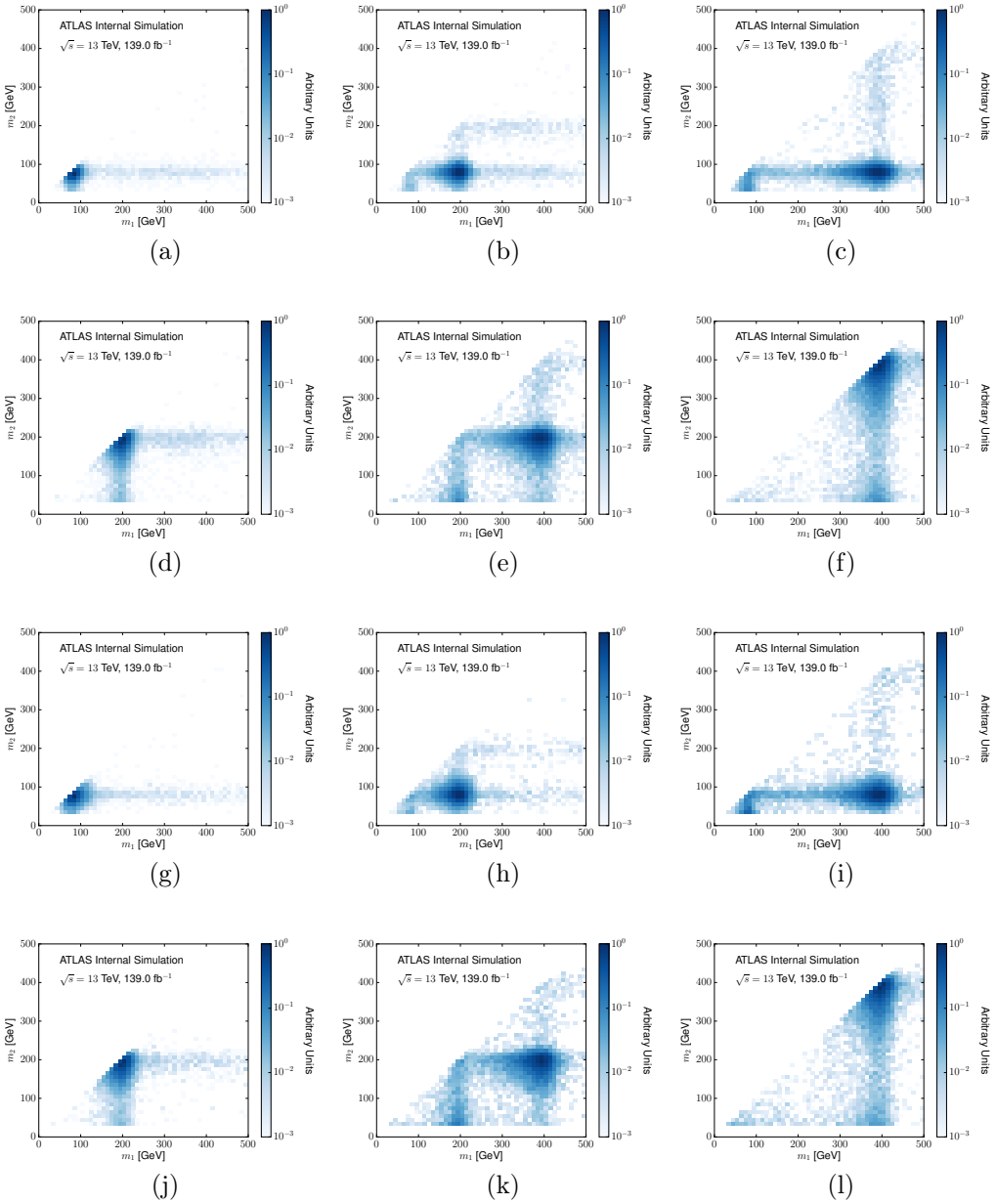


Figure 7.11: Distribution of m_1 and m_2 for a signal model with:
(a) $(m_A, m_B, m_C = 3000, 80, 80$ GeV); (b) $(m_A, m_B, m_C = 3000, 200, 80$ GeV);
(c) $(m_A, m_B, m_C = 3000, 400, 80$ GeV); (d) $(m_A, m_B, m_C = 3000, 200, 200$ GeV);
(e) $(m_A, m_B, m_C = 3000, 400, 20$ GeV); (f) $(m_A, m_B, m_C = 3000, 400, 400$ GeV);
(g) $(m_A, m_B, m_C = 5000, 80, 80$ GeV); (h) $(m_A, m_B, m_C = 5000, 200, 80$ GeV);
(i) $(m_A, m_B, m_C = 5000, 400, 80$ GeV); (j) $(m_A, m_B, m_C = 5000, 200, 200$ GeV);
(k) $(m_A, m_B, m_C = 5000, 400, 200$ GeV); (l) $(m_A, m_B, m_C = 5000, 400, 400$ GeV).

7.4.3.1 Mass Decorrelation

In Figure 7.10, it can be seen that there are true physical differences in the features in the background between a given signal region and sideband regions (as chosen in Analysis Step 3); in particular, the distribution gets more populated at higher m_1,m_2 as m_{JJ} increases. This therefore requires some modification of these features in order to satisfy Assumption 2.

The idea is to scale the 1-dimensional marginal distribution of the jet mass m (combining both m_1 and m_2 across all events) in the sideband regions to the signal region. This is accomplished by constructing the empirical cumulative distribution function (ECDF) in each region s:

$$\Phi_s(m) = \frac{1}{n_s} \sum_{j=1}^{n_s} \mathbb{1}(m_j \leq m) \tag{7.1}$$

Where n_s is the number of jets in m_{JJ} region s, $\mathbb{1}$ is the indicator function, and j goes over all jets (leading and subleading) in region s.[13] Note that the distribution of $\Phi_s(m)$ over all jets in m_{JJ} region s is by definition uniform; and that the distribution of $\Phi_s^{-1}(x)$ is exactly the distribution of m in m_{JJ} region s, if x follows a uniform distribution. The ECDF is a good approximation of the true CDF in regions where there are a sufficient number of samples. This motivates the choice to place a maximum cut on the jet mass at 500GeV, in order to restrict to the region in which the CDF can be approximated well.

Then, for a given signal region s, the masses of all jets in the sideband regions $s-1$ and $s+1$ are rescaled to the signal region:

$$\begin{cases} m \to \Phi_s^{-1}\left(\Phi_{s-1}\left(m\right)\right) & m \in \text{region } s-1 \\ m \to \Phi_s^{-1}\left(\Phi_{s+1}\left(m\right)\right) & m \in \text{region } s+1 \end{cases} \tag{7.2}$$

This ensures that the 1-dimensional distributions of m in the sideband regions are exactly the same, by construction, as in the signal region. Note that, if there is a

[13] Actually, $\Phi_s(m)$ is evaluated at all values of m that exist in the m_{JJ} region, and then is linearly interpolated for intermediate values.

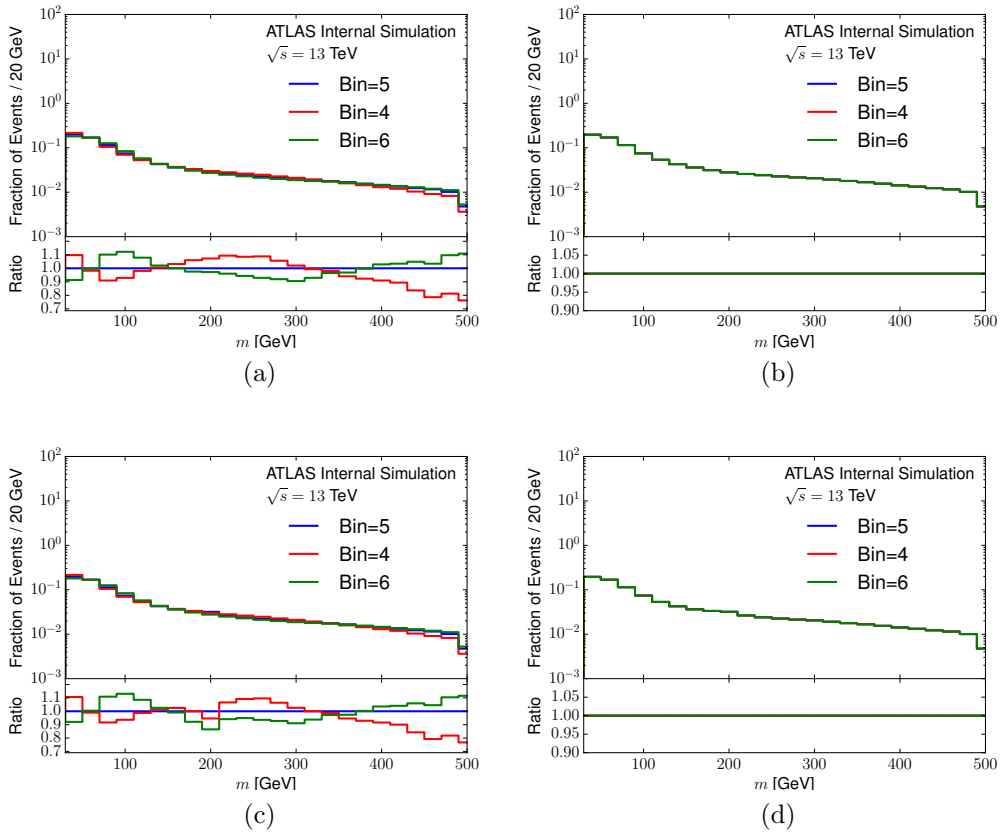


Figure 7.12: The distributions of m in simulation, comparing between signal region $s = 5$ and sideband regions $s-1 = 4$ and $s+1 = 6$. (a) Before any scaling. (b) After scaling via the empirical cumulative distribution function. (c) Including an injected signal sample with $m_A, m_B, m_C = 3000, 200, 200$ GeV, which lies mostly in signal region 5, and with $\frac{S}{\sqrt{B}} \sim 2$ in that region. (d) After scaling, with the presence of the signal sample.

signal present in the signal region, that the scaling will be slightly biased by the presence of this signal, depending on the signal fraction in that region. This scaling is demonstrated for an example signal and sideband regions in Figure 7.12, including an example with an injected signal.

After the 1-dimensional scaling the only differences that can exist are in the 2-dimensional distribution m_1 and m_2, which can arise due to differences in the correlation between m_1 and m_2 between the regions. It is expected that in the background

these differences are small, and this is supported by examining the distributions of m_1 and m_2 in simulation, as can be seen in Figure 7.13.

In order to additionally remove correlations between the features and m_{JJ}, the left and right sidebands are combined with equal total weight for the training (Section 7.4.4). Then the distribution of the features in the combined sidebands is exactly the average of the distribution in the left and right sidebands, removing first-order dependencies of the features on m_{JJ}.

The presence of a signal in one of the regions is exactly such a difference that can exist in the correlations between m_1 and m_2. Therefore, even though the scaling can be biased by the presence of a signal in the signal region, it is expected that there will still exist differences in the 2-dimensional distribution between the signal and sideband regions, as can be seen in Figure 7.13.

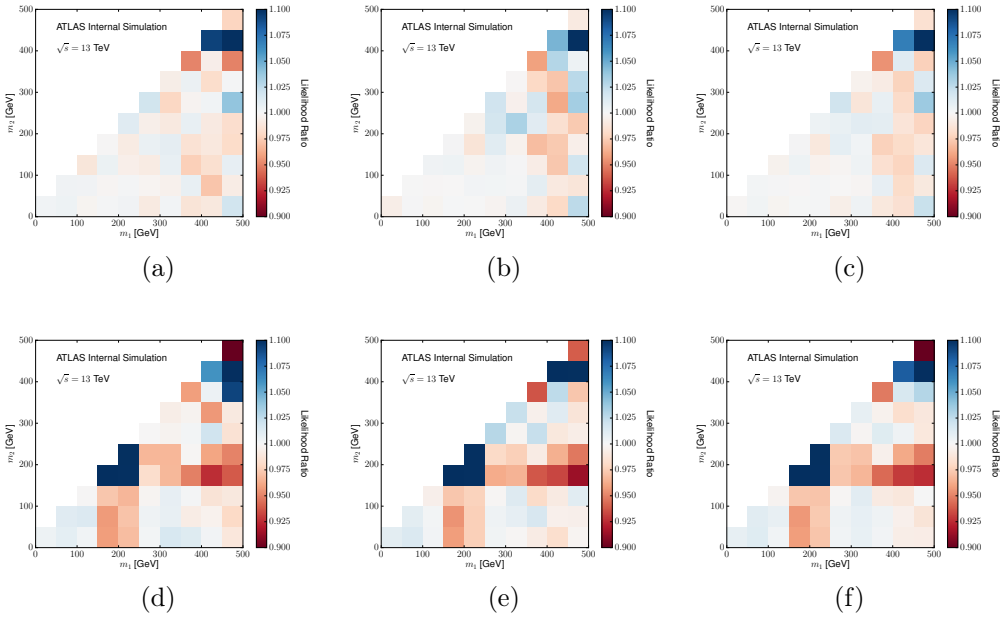


Figure 7.13: The 2-dimensional likelihood ratio in m_1 and m_2 after scaling, comparing between signal region $s = 5$ and (a) sideband region $s - 1 = 4$; (b) sideband region $s + 1 = 6$; and (c) combining sideband regions 4 and 6. (d,e,f) Including an injected signal sample with $m_A, m_B, m_C = 3000, 200, 200$ GeV, which lies mostly in signal region 5, and with $\frac{S}{\sqrt{B}} \sim 2$ in that region. N.B. This Figure may appear fuzzy if viewing in macOS Preview; using a different PDF viewer, e.g. Google Chrome, seems to fix the problem.

7.4.4 Neural Network Architecture

The neural network which determines the final score for each event, indicating more or less signal-like, is derived in multiple stages. Neural networks have been shown to have excellent performance on a wide variety of training tasks in high dimensions with limited training data [391–396]. Since there are only two features being used in this search, it may seem like overkill to use a NN, as opposed to simply examining the likelihood ratio by using two-dimensional histograms, or using a boosted decision tree [397–400].

The first response to this is that, while the NN is a monotonic function of the likelihood ratio in the asymptotic limit, the NN also involves certain regularization

constraints which allow the estimation of the likelihood ratio to be smooth. This smooth estimation is actually crucial for allowing CWoLa hunting to be generically sensitive without paying a large trials factor - this claim is explored in detail in Appendix C.1. The multiple stages in the derivation of the final NN score are intended to maximize the sensitivity to new potential signals, while remaining robust to statistical fluctuations and maintaining or enhancing this smooth behavior.

The second response is that the natural extension of this search is to involve a large number of features of the jets (Section 7.8). In that case, a NN is the natural choice to handle this high-dimensional classification problem. In order to set up this search (the first of its kind) to be able to be extended as straightforwardly as possible, a NN is used here in anticipation of the next iteration of this search.

In Analysis Step 3 (Section 7.3.2), some signal bin s is designated. The process below can then be considered an enumeration of the steps in Analysis Step 4 (Section 7.3.2). See also Figure 7.14 for an overview.

1. Designate the sideband regions $s+1$ (the "upper" sideband) and $s-1$ (the "lower" sideband).

2. All events in the entire dataset are separated randomly into 5 equally-sized *cross-validation* sets $i \in [0-4]$.

3. Designate one of the cross-validation sets i_t as the *test set*.

4. Designate a different cross-validation set $i_v \neq i_t$ as the *validation set*. The remaining 3 cross-validation sets are designated as the *training set*.

5. A neural network is trained using only the training set, with features $X = \{m_1, m_2\}$, rescaled as described in Section 7.4.3.1. The events are labeled with $Y \in [0, 1]$, with 1 indicating an event in the signal region, and 0 indicating an event in the sidebands. The events in each of the sideband regions are weighted uniformly in the training so that the sum of weights in each of the sideband regions is equal.

The details of the NN are given below[14]. The network is a Sequential Neural Network with 4 layers of size 64,32,8,1, and activation functions `ReLu`,`ReLu`,`ReLu`,sigmoid, respectively, as implemented in `Keras` [301]. The loss used is the binary cross-entropy, and the training loss is minimized using the `Adam` [401] optimizer. The loss is evaluated on the validation set.

Total number of epochs is 1000, with an early stopping with patience of 100 on the validation loss. The batch size is 1% of the total.

6. Repeat Step 5 with 3 different random initial configurations of the neural network weights, and choose the network with the lowest validation loss. The other 2 networks are discarded.

   Finally, this network is evaluated on the test set, and only the scores on the test set are recorded.

   Scaling

   The score from the neural network output is scaled monotonically as a quantile between 0 and 1 on the events in the signal region test set (note that the scores for all events in the test set are recorded, not just those in the signal region). That is to say, after the scaling, an event with a score of $0 \leq \epsilon \leq 1$ received a higher score from the neural network than exactly a fraction ϵ of the events in the signal region test set. In particular, a high NN score corresponds to a low ϵ. E.g., a selection choosing the 10% most signal-like events would place a minimum on the NN score, which because of the monotonic rescaling corresponds to a maximum on ϵ, with exactly the selection $\epsilon < 0.1$. The scaling is done in this way so that different networks can later be combined by averaging their scores; since the network output score is standardized, this averaging is meaningful.

7. Repeat Steps 4-6 by varying i_v to each of the remaining cross-validation sets in turn, keeping the test set i_t fixed. I.e., there are 4 different validation set choices for the given test set, and these steps are repeated for each one, resulting in 4

[14] For an excellent introductory explanation of machine learning and neural networks, see [396].

different networks. The scores from these 4 networks are then averaged, and the average score is further rescaled as described in Step 6.

8. Repeat Steps 3-7 by varying the chosen test set i_t, for a total of 5 networks. The total sample is then combined together; since the score from each network is scaled to be an efficiency on its respective test set, and the test sets are equally sized, the score for each event remains an efficiency over the entire dataset.

Ultimately, there are $3 \times 4 \times 5 = 60$ neural networks trained for each signal region s. A flowchart of the process described above is given in Figure 7.14, and the neural network outputs are shown at each step in the process for a sample with injected signal.

The events are separated into training, validation, and test sets as described above in order to reduce the effect of statistical fluctuations in the training. Since the validation set is statistically independent from the training set, choosing the network with the lowest validation loss, as described in Step 6, ensures that the network which learns true correlations the best is chosen. Since the test set is statistically independent from both the training and validation sets, the network cannot do artificially better by being biased by statistical fluctuations in the training or validation set. This is actually a crucially important point for removing the look-elsewhere effect in this analysis. If the network were trained and tested on the same dataset, one would expect a high rate of false positives in terms of being able to separate the signal region from the sidebands due to correlated (actually, the same) statistical fluctuations between the train and test sets, which results directly in a bump in the m_{JJ} spectrum. Because the train and validation sets are statistically independent from the test set, the statistical fluctuations are uncorrelated, and in the case that there is no true signal the rate of false positives is no more than that expected due to a random classifier.

The choice to train 4 different networks with each non-test set chosen to be a validation set in succession, as described in Step 7, is again intended to reduce the sensitivity to statistical fluctuations in the training sets, since for each validation set the training set is (somewhat, but not entirely) different. The averaging over the

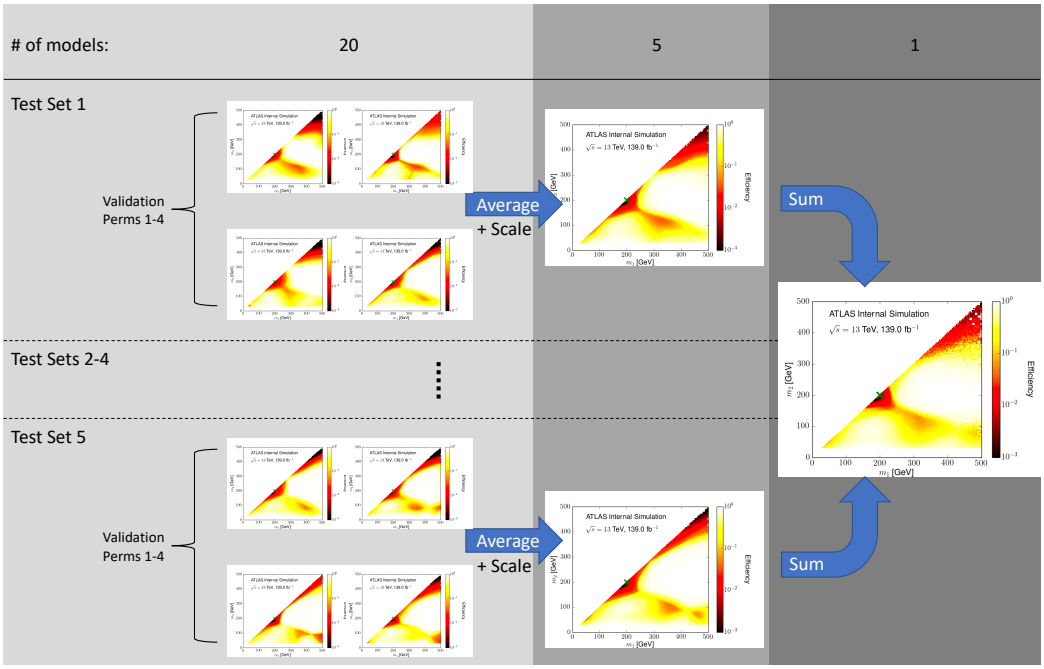


Figure 7.14: Flowchart of steps in derivation of final network scores, as described in Section 7.4.4. The networks in the left-most column have already been chosen to have the lowest validation loss, as described in Step 6. The networks are represented by a 2D plot showing the neural network output in the m_1,m_2 plane, as expressed as an efficiency on events, as described in Step 6. In this particular example, a signal was injected at $m_B = 200$ GeV, $m_C = 200$ GeV. All plots in this figure use simulation. Note that the amount of effective data in the left parts of the plot is actually $1/5$ of the total.

validation sets can be seen when going from the first column of Figure 7.14 to the second column.

Note that, because of the fact that the networks between different test sets never interact, as described in Step 8, two events with exactly the same features could in principle have different scores, if they happened to lie in different test sets. However, because of the extensive measures taken to reduce the sensitivity to statistical fluctuations in the training set, it is expected that the final 5 networks should mostly be the same; this can be seen for example in the penultimate column of Figure 7.14.

Details about the software used to execute the analysis pipeline up to this point (Sections 7.4.1, 7.4.2, 7.4.3, 7.4.4) can be found in Appendix C.4.

7.4.5 Systematic Uncertainties

As will be described in Section 7.4.6, the background is estimated in a fully data-driven manner, and the only uncertainties are associated with the background fit. The uncertainties related to the signal simulation are only relevant for setting cross-section limits (Section 7.4.7.2). For the signal, there are uncertainties on the jet mass scale and resolution as well as on the modeling of jet fragmentation; there is also an overall luminosity uncertainty.

The former set of uncertainties use the prescriptions of the Jet/MET group [402].

The luminosity uncertainty for the full Run 2 dataset is 1.7% [403]. It is derived from the calibration of the luminosity scale using x-y beam-separation scans, following a methodology similar to that detailed in Ref. [404], and using the LUCID-2 detector for the baseline luminosity measurements [226]. The total integrated luminosity is 139 fb^{-1}.

The NN is not retrained for every systematic variation. Instead, the NN trained with the nominal signal and then applied to the events with the kinematic properties of the jets varied according to the uncertainty. In principle this is a conservative treatment of the uncertainties, since the NN learns what the actual signal looks like and tags the kinematic space accordingly; training the network on the nominal signal and then applying to the varied kinematic properties can therefore lead to artificially low NN signal tagging efficiencies. In practice, though, the uncertainties are small enough such that the total efficiency on the signal for some fixed NN selection is about the same in the nominal and varied samples.

7.4.6 Fitting

After the application of the NN selection, a standard parametric background fit is performed to estimate the background contribution in the signal region. As mentioned in Section 7.3.3, the fitting procedure is relatively simple for the simulation analysis, but more complex for the data validation and full unblinded analyses. There are in addition some small differences in the fitting procedure between these latter two analyses. Because of this, the fitting procedure for each of these analyses is broken

out into separate subsections.

Details about the software used for the fitting are given in Appendix C.5.

7.4.6.1 Simulation Analysis

The parametric fit is performed using 100 GeV bins from 1.8 TeV to 8.2 TeV[15]. These bins are finer than the bins used for learning as described in Section 7.4.2 [16].

The selected fit function is the same as the one used by the all-hadronic diboson resonance search [20]:

$$\frac{dn}{dx} = p_1(1-x)^{p_2-\xi p_3}x^{-p_3}, \tag{7.3}$$

where $x = m_{JJ}/\sqrt{s}$, p_1 is a normalization parameter, p_2 and p_3 are dimensionless shape parameters, and $\xi = 4.2955$ is a constant (this value chosen by the all-hadronic resonance search to remove correlations between p_2 and p_3).

The parameter p_2 is initialized to 10 and restricted to the range $[-100, 100]$ and the parameter p_3 is initialized to -30 and restricted to the range $[-100, 100]$ as well. The signal region is not masked during this fit. After the fit, the prediction is normalized to have the same integral as the data over the entire range. As there are two free parameters, the fit produces a 2×2 covariance matrix.

The background fit is decoupled from the signal strength scan; the signal strength is used as a POI to set limits, as discussed in Section 7.4.7.2. A background prediction is generated as a histogram from the central values of the background fit. There is one nuisance parameter for the background fit systematic uncertainty. The "up" and "down" variations are created by taking the sum in quadrature of the uncertainties from the two parameters from the original background fit (p_2 and p_3). This nuisance parameter can be profiled in the second fit used to fit the POI.

[15] The 8.2 TeV maximum bin edge contains the 8.17 TeV maximum m_{JJ} from the event selection (Section 7.4.1); this value is essentially arbitrary as there are expected to be < 1 events at that high m_{JJ}.

[16] In the corresponding Section (7.5), some example fits are shown using the m_{JJ} bins themselves for the parametric fit; these are not actually used for the final results.

7.4.6.2 Validation Analysis

As in the fitting for the simulation analysis (Section 7.4.6.1), the parametric fit uses bins of size 100 GeV, spanning 1.8 TeV to 8.2 TeV. Unlike in that fit, **the signal region is masked for this fit**. In order to remove the dependence of a potential signal, the masked region for each bin is enlarged to include half of both the left and right neighboring bins. In particular, the bins used for training and the windows used for masking in the fit are presented in Table 7.4.

Table 7.4: m_{JJ} bin definitions and the mask regions for the background fit.

Bin	Definition	Mask
5	$2.74 \leq m_{JJ} < 3.28$ TeV	$2.5 \leq m_{JJ} < 3.6$ TeV
6	$3.28 \leq m_{JJ} < 3.94$ TeV	$3.0 \leq m_{JJ} < 4.3$ TeV
7	$3.94 \leq m_{JJ} < 4.73$ TeV	$3.6 \leq m_{JJ} < 5.2$ TeV
8	$4.73 \leq m_{JJ} < 5.68$ TeV	$4.3 \leq m_{JJ} < 6.2$ TeV
9	$5.68 \leq m_{JJ} < 6.81$ TeV	$5.2 \leq m_{JJ} < 7.5$ TeV

The fitting procedure then proceeds as follows:

1. Perform a fit using the sidebands using Equation 7.3. Compute the χ^2 in the sideband:

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}, \tag{7.4}$$

   where O_i is the number of events observed in data in bin i, E_i is the expected number of events from the fit function in bin i, and the sum runs over all N sideband bins. The parameter p_1 is fixed by the normalization in the sidebands. The sideband p-value is then computed using $N - 3$ degrees of freedom, since there are 3 fit parameters. If this p-value is greater than 0.05, move on to step 5, else:

2. Try an extended fit function:

$$\frac{dn}{dx} = p_1(1-x)^{p_2-\xi_1 p_3} x^{-p_3+(p_4-\xi_2 p_3-\xi_3 p_2)\log(x)}. \tag{7.5}$$

As before, compute the sideband p-value (now with $N-4$ degrees of freedom). If this p-value is greater than 0.05, move on to step 5, else:

3. Try with the UA2 fit function [405]:

$$\frac{dn}{dx} = p_1 x^{p_2-\xi_1 p_3} e^{-p_3 x+(p_4-\xi_2 p_3-\xi_3 p_2)x^2}. \tag{7.6}$$

As before, compute the sideband p-value with $N-4$ degrees of freedom. If this p-value is greater than 0.05, move on to step 5, else:

4. Reduce the sideband window size and repeat steps 1-3 until the sideband p-value is above 0.05 or the range used to fit in the sideband is smaller than 800 GeV (in which case, the fit fails). The sideband is range is reduced as follows. If the right sideband is bigger than the left one, the rightmost 400 GeV is removed from the right sideband. If the left sideband is larger than the right one, the leftmost 400 GeV is removed from the left sideband. If both sidebands are the same size, then 200 GeV is removed from both.

5. After a fit is found with p-value greater than 0.05, the ξ_i parameters are optimized to reduce correlations between parameters in order to improve the quality of the uncertainties used for the statistical analysis (Section 7.4.7). For all of the fits, the ξ_i are initialized to zero. Then, whatever fit setup was found in the previous steps is repeated after iteratively adjusting the ξ_i as follows. For the three parameter fits, we set

$$\xi = \text{correlation}(p_2, p_3)\frac{\sigma_{p_2}}{\sigma_{p_3}}. \tag{7.7}$$

This can be viewed as a Gram-Schmidt orthogonalization, treating the random variables p_2 and p_3 as vectors in an inner product space with the inner product between two vectors given by their covariance.

The ξ setting is iteratively repeated automatically until the residual correlation is less than 0.25. In practice, we find that after one iteration, the correlation converges to 10^{-5} or smaller. For the four-parameter fit, the pairwise correlations are removed in a similar fashion:

$$\xi_1 = \text{correlation}(p_2, p_3)\frac{\sigma_{p_2}}{\sigma_{p_3}} \tag{7.8}$$

$$\xi_2 = \text{correlation}(p_3, p_4)\frac{\sigma_{p_3}}{\sigma_{p_4}} \tag{7.9}$$

$$\xi_3 = \text{correlation}(p_2, p_4)\frac{\sigma_{p_2}}{\sigma_{p_4}}. \tag{7.10}$$

This procedure converges slower than the three-parameter fit and sometimes does stop at close to (but less than) 0.25 correlation between some pair of two variables.

6. Finally, the local p-value in the masked region is quoted to give a sense of the deviations from the background expectation in the signal region. The χ^2 is calculated including the signal uncertainties:

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i + \sigma(E)_i^2}, \tag{7.11}$$

where O_i is the number of events observed in data in bin i, E_i is the expected number of events from the fit function in bin i, $\sigma(E)_i$ is the uncertainty on the mean value of the fit function in bin i (as described in Section 7.4.7), and the sum runs over all N bins in the masked region. Since the masked region is not used to derive the fit, the number of degrees of freedom used to calculate the p-value is equal to N.

7.4.6.3 Unblinded Analysis

The fitting procedure is very similar to the procedure in the data validation analysis (Section 7.4.6.2). In particular, bins of size 100 GeV are used for the fitting, spanning 1.8 TeV to 8.2 TeV. Like the all-hadronic diboson search, and unlike the fitting for the validation analysis described in Section 7.4.6.2, **the signal region is not masked for this fit**, and the likelihood of the fit function is minimized across the entire range. However, in order to fit the background when there is the presence a potential signal, there is still a masked region defined, and **the fit quality is evaluated on only the sidebands outside the masked region**. This change was applied as it was found that the fit function was sometimes performing poorly in the masked regions[17]. The masked regions are the same as for the validation analysis as described in Section 7.4.6.2 (Table 7.4); in particular, the masked region for each bin is enlarged to include half of both the left and right neighboring bins.

The fitting procedure is exactly the same as described in Section 7.4.6.2, other than that, in Fit Step 1, the fit includes the masked region. In particular, the fit is still evaluated on the sidebands outside the masked region and continues through the steps until the χ^2 p-value in the sidebands is greater than 0.05. The local p-value in the masked region is quoted to give a sense of the deviations from the background expectation in the signal region. Since the signal region is used in the overall fit, the χ^2 is calculated using only the statistical uncertainties on the background expectation, not including the uncertainties on the fit:

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}, \tag{7.12}$$

where O_i is the number of events observed in data in bin i, E_i is the expected number of events from the fit function in bin i, and the sum runs over all N bins in the masked region. Since the masked region is used to derive the fit, the number of degrees of freedom used to calculate the p-value is equal to $N-3$ or $N-4$, depending on the ultimate fit function used.

[17] In the validation data fits.

Since the fit includes the masked region as well as the sidebands, in the presence of a signal the background expectation can be biased upwards, leading to a negative bias on the fitted $\hat{\mu}$ when setting limits (Section 7.4.7.2). This effect is small, though, when the signal presence is at or less than the limits that are actually set; the results of a signal injection test can be found in Appendix C.8.

Detailed tests on the validation dataset indicated that the $\epsilon = 1.0$ and $\epsilon = 0.25$ background spectra were not adequately described by the fitting functions used. The results of these fits can be found in Section 7.6. Therefore, the final analysis only includes the results with $\epsilon = 0.1$ and $\epsilon = 0.01$.

For these values also it was found in the validation dataset that the fitting function was inadequate to describe the background spectra - the fit function tends to underestimate the data at low m_{JJ} and overestimate at high m_{JJ}, leading to a non-closure in the significances (Equation 7.13) of the data with respect to the background (meaning a mean significance of < 0). However, this effect is quite small, so it can be corrected. This correction is described in detail in Appendix C.10. A linear correction in m_{JJ} is applied to the final fit values. This correction is derived in the data validation dataset, and then validated by testing for closure in the sidebands of the unblinded dataset. An uncertainty on this correction is derived in the validation dataset and applied as an additional uncertainty on the background expectation.

7.4.7 Statistical Analysis

7.4.7.1 Background Compatibility

The background compatibility is presented in two ways, depending on the precision of the result required.

For the simulation analysis and data validation analysis, the background compatibility is quantified as the (Data-Fit)/Uncertainty, i.e. the (signed) χ in the respective χ^2 calculation.

For the final fit results, a more precise p-value calculation is included. Where indicated, the *significance* of the data with respect to the background fit is shown. The significance S is calculated as the inverse Gaussian CDF of the p-value of the

data under the background-only hypothesis:

$$S = \Phi^{-1}\left(\sum_{k<O_i} P(E_i; k)\right), \tag{7.13}$$

where O_i is the observed count in bin i, E_i is the expected count in bin i (background fit value), and Φ^{-1} is the inverse Gaussian CDF; the p-value is the argument of Φ^{-1}. When $k = 0$, the p-value is 0, so the significance is in principle $-\infty$; in these bins the significance is simply quoted as 0. As mentioned in Section 7.4.6.3, since each bin is included in the fit in the full unblinded analysis, the uncertainties on the background fit itself are not included when calculating the significance. However, the additional uncertainty on the background fit due to the fit correction derived in the validation selection (Appendix C.10) *is* included in the background-only significance calculation by allowing E_i to vary within Gaussian constraints when calculating the p-value. This significance is directly used as the test of the background-only hypothesis; a significance > 5 would indicate a "discovery" of a new signal.

7.4.7.2 Setting Limits

The observed results are interpreted using a frequentist statistical analysis when setting limits. The parameter of interest is the signal strength, μ, defined as a scale factor on the total number of signal events expected relative to some benchmark, so that $\mu = 0$ corresponds to no signal, and $\mu = 1$ corresponds to the benchmark. As there is no specific signal model, the couplings and thus cross-sections are free parameters. Therefore, the benchmark is defined to be such that, in the benchmark model, the total number of events expected to be produced is exactly 1: $\sigma \times B \times \mathcal{L} = 1$, with σ the signal cross section for A production, BR the signal branching ratio to BC which then decay hadronically, and $\mathcal{L}$ the data luminosity. Then μ is exactly the total number of signal events expected to be produced.

A likelihood function $\mathcal{L}(\mu, \theta)$ is defined, with $\theta = \{\theta_s, \theta_b\}$ going over the nuisance parameters in the signal (θ_s) and background (θ_b). The likelihood function is defined

as follows:

$$\mathcal{L}(\mu, \theta) = \prod_i P_{pois}(n^i_{obs}|\mu s_i + b_i) \times \mathcal{G}(\theta_b) \times \mathcal{G}(\theta_s), \tag{7.14}$$

where n^i_{obs} is the number of events observed in bin i, s_i is the fraction of signal expected in bin i (including the event selection due to the NN selection), θ_s are the nuisance parameters corresponding to the systematic uncertainties on the signal shape[18] (Section 7.4.5), b_i is the total number of background expected in bin i, θ_b are the nuisance parameters corresponding to the uncertainty on the background fit[19], as derived from the fitting and described in Section 7.4.6, $P_{pois}(n|e)$ is a Poisson likelihood of n events given e expected, and $\mathcal{G}$ is a Gaussian. Note that, since θ_s only affect the shape of the signal, these uncertainties cannot be profiled.

The test statistic $\lambda(\mu)$ based on the profile likelihood ratio is defined using the lowest order asymptotic approximation [346]. Exclusion limits at the 95% confidence level, μ_{95} are also set following the CL_s prescription [345].

In this analysis, the cuts are not set in advance, and are rather determined by the number and nature of a potential signal. In order to remain agnostic to the number and nature of a potential new signal, rather than optimizing the efficiency of the NN cut for a particular signal model, a few different NN cut efficiencies ϵ are chosen in order to scan the space of possible NN cuts. These chosen efficiency values are listed in Table 7.5[20]. Each choice of ϵ is treated as a separate statistical analysis; since the

Table 7.5: Chosen values of NN cuts with efficiency ϵ for analysis.

	Values
ϵ	[1.0,0.25,0.1,0.01]

choices of ϵ differ by factors of more than 2, the results with a given NN at different values of ϵ should be mostly independent.

[18] I.e., $s_i \to s_i(1 + \theta_s \sigma_{s_i})$ with the size of the variation σ_{s_i} derived by varying the systematic uncertainties in simulation.

[19] I.e., $b_i \to b_i(1 + \theta_b \sigma_{b_i})$ with the size of the variation σ_{b_i} the uncertainty on the background fit in bin i.

[20] In the final unblinded analysis, only the values $\epsilon = 0.1$ and $\epsilon = 0.01$ are used.

For deriving the 95% confidence exclusion limits μ_{95}, it is necessary to inject signal on top of the background expectation in order to evaluate the profile likelihood $\lambda(\mu)$. The NN output depends on the presence of the signal, and in particular gets better at tagging signal the more signal there is (Section 7.5.2). Therefore, the μ_{95} values are evaluated by first injecting a certain amount of signal μ, running the NN training, and then performing the statistical analysis, deriving a limit $\mu_{95}(\mu)$ which is a multiple of the injected signal strength while keeping the NN fixed. In particular, the values $\mu_{95}(\mu)$ are functions of the injected signal strength μ, derived from performing the statistical analysis after such a signal is injected. Ideally, μ would be scanned until $\mu_{95}(\mu) = \mu \equiv \hat{\mu}_{95}$ - with this strength of signal present, that exact signal would be excluded.

However, it is expensive to scan finely over the injected μ (because of the requirement to retrain the NN at each value), so upper limits on $\hat{\mu}_{95}$ are derived by injecting a coarse grid of signal strengths μ; for each analysis (validation in simulation (Section 7.5.4), validation in data (Section 7.6.4), and unblinded (Section 7.7.5)) this grid is provided in the respective section. In order to simulate a true signal, for each given injected signal strength μ, exactly μ events are chosen from the given signal MC sample, with probability proportional to the MC weight. These events are then injected into the data sample with weight 1 and included with all the rest of the data when being passed through the steps of the analysis, in particular the event selection (Analysis Step 2), the cross-validation splits (Section 7.4.4), and the mass decorrelation and training (Analysis Step 4). However, for the tagging (Analysis Step 6), the derived NN is applied to the entire signal sample with weights in order to get the full signal m_{JJ} histogram after tagging. The signal is injected in this way with weights 1 for the training in order to simulate as close as possible for the NN what a true signal would look like. The tagging is done with the full sample with the MC weights in order to derive limits as a multiple of the injected μ with the NN fixed, as described above. For each of the $3 \times 4 \times 5 = 60$ (Section 7.4.4) iterations of the signal training, the chosen signal events injected for the NN training are held fixed.

However, for the final unblinded analysis (Section 7.7.5), this entire process is repeated for 5 different random samplings of the signal MC as a smoothing procedure,

and the expected limits that are used for the given μ value (median and bands) are the limits derived from the sampling which gives the *median* median expected limit over the 5 random samplings. The 5 random samplings indicate an uncertainty in the output of the NN, based on the specific presence of the signal the NN is trained on. This uncertainty is added to the bands of the expected limits: the (unbiased or Bessel-corrected) variance of the median expected limits is added in quadrature to the $\pm 1, 2\sigma$ bands relative to the median. The observed limits that are used for the given μ value are the limits derived from the samplings with gives the median (over the 5 random samplings) observed limit.

Once a limit $\mu_{95}(\mu)$ is derived for each injected μ, the final limit overall is derived as described below.

A crucial observation is that $\mu_{95}(\mu_1) \geq \mu_{95}(\mu_2)$ if $\mu_1 < \mu_2$; this is because, when more signal is injected, the NN only gets better at discriminating that signal from the background, and the exclusion limit therefore only gets better (recall that the total efficiency in the signal region ϵ is fixed). We then derive the following results:

$$\mu > \hat{\mu} \rightarrow \mu_{95}(\mu) \leq \mu_{95}(\hat{\mu}) = \hat{\mu} < \mu \tag{7.15}$$

$$\mu < \hat{\mu} \rightarrow \mu_{95}(\mu) \geq \mu_{95}(\hat{\mu}) = \hat{\mu} > \mu \tag{7.16}$$

Therefore, if $\mu > \mu_{95(\mu)}$, then μ is an appropriate upper limit on $\hat{\mu}$; and if $\mu < \mu_{95(\mu)}$, then $\mu_{95}(\mu)$ is an appropriate upper limit on $\hat{\mu}$. In summary, with a signal injection of μ, the upper limit on $\hat{\mu}$ that is set is $\max(\mu_{95}, \mu)$. Since it is desirable to derive an exclusion limit as close as possible to $\hat{\mu}$, a few different values of μ are injected (depending on the signal region), and the value of μ that leads to the lowest median expected upper limit is used.

For some signals, at low μ values, the rule $\mu_{95}(\mu_1) \geq \mu_{95}(\mu_2)$ if $\mu_1 < \mu_2$ can be violated; this is due to the NN learning to tag basically a random region of the parameter space and just happening to tag the signal. For these signals, the limits at those μ values are simply not used.

7.5 Simulation Analysis

7.5.1 Event Selection

In this section, the analysis is validated on a MC sample with the same event selections as used in the full unblinded data analysis, as detailed in section 7.4.1.

7.5.2 Neural Network Output

Some thorough studies are performed to test the sensitivity of the neural network to various signals.

A study is performed by training the neural network with different injected signals in different part of the m_A, m_B, m_C kinematic space, for fixed μ.

For $m_A = 3000$ GeV, this signal lies in signal region 5. In this region, there are roughly 1.0×10^5 background events, and μ is set to 1500; after all event selections (45% efficiency), there are roughly 650 signal events in signal region 5, for a signal fraction of $\sim 0.7\%$ and an estimated significance $\frac{S}{\sqrt{B}} \sim 2$. The results of this study can be seen in Figure 7.15. The network successfully learns to tag the kinematic region near the true signal as being signal-like.

A second study is performed with a fixed signal model ($m_{JJ} = 3000$ GeV; $m_B, m_C = 200, 400$ GeV), varying μ in order to test the sensitivity of the neural network to μ. This signal lies in signal region 5; the results when training with signal region 5 as the signal region can be seen in Figure 7.16. There are roughly 1.0×10^5 background events in this signal region; the signal fraction and estimated significance varies from $\sim 0.3\%$, ~ 1 (at $\mu = 750$), respectively, to 0.7%, ~ 2 (at $\mu = 1500$), respectively. For low μ, the network is unable to detect the presence of the signal, and it learns to tag the same kinematic regions as in the background-only case. As μ increases, the network becomes more confident about tagging the kinematic region near the true signal, while still tagging the kinematic regions that are present in the background-only case.

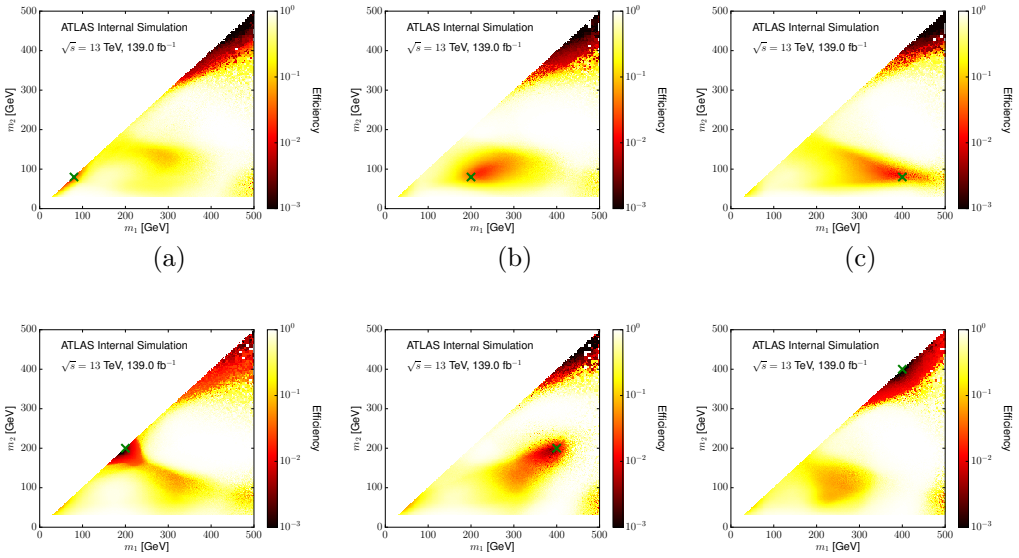


Figure 7.15: Neural network output for signal region 5 with $\mu = 1500$ for a signal with $m_A = 3000$ GeV and (a) $(m_B, m_C = 80, 80$ GeV); (b) $(m_B, m_C = 200, 80$ GeV); (c) $(m_B, m_C = 400, 80$ GeV); (d) $(m_B, m_C = 200, 200$ GeV); (e) $(m_B, m_C = 400, 200$ GeV); (f) $(m_B, m_C = 400, 400$ GeV).

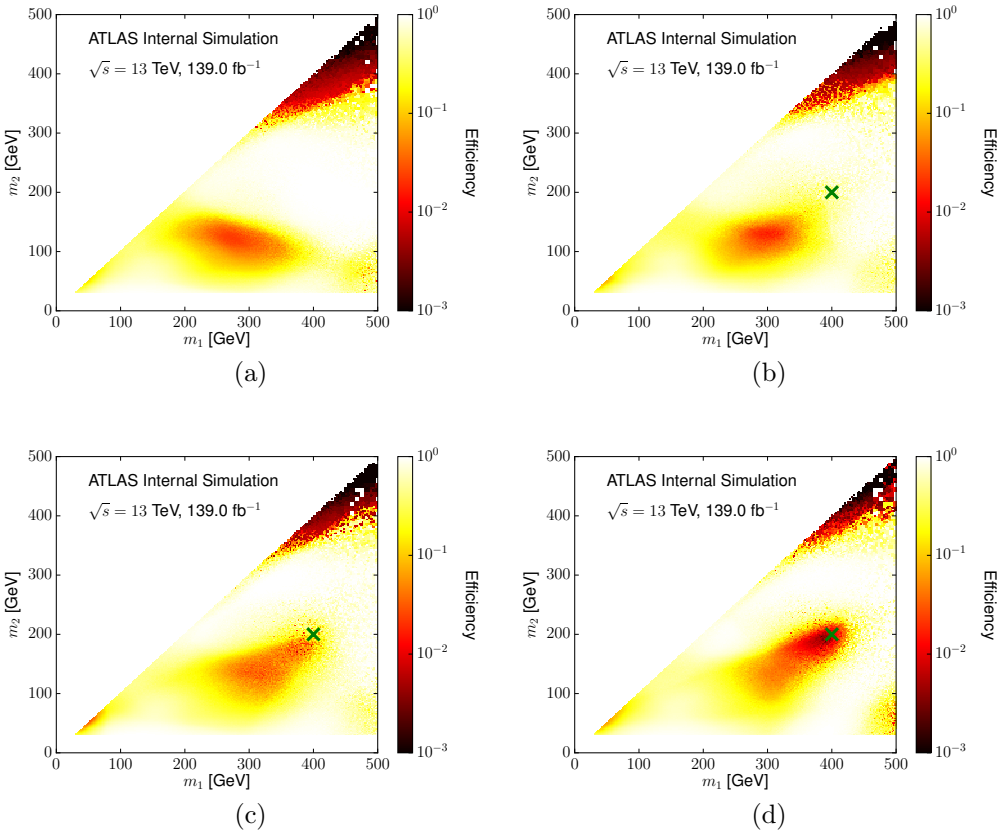
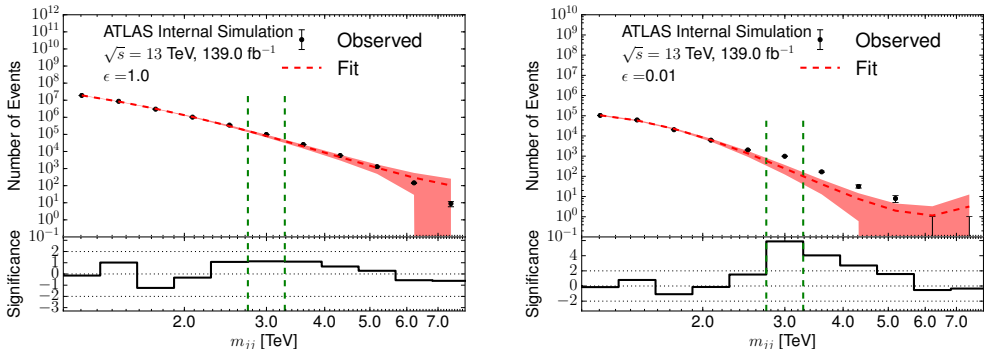


Figure 7.16: Neural network output for signal region 5 with a signal at $(m_A, m_B, m_C = 3000, 200, 400$ GeV) with (a) $\mu = 0$; (b) $\mu = 750$; (c) $\mu = 1000$; (d) $\mu = 1500$.

7.5.3 Fitting Results

Example fits in the coarse bins used for training are shown in Figure 7.17.



(a) (b)

Figure 7.17: Example fits to the m_{JJ} spectrum using the functional form given in Equation 7.3. There is an injected signal sample at $m_A, m_B, m_C = 3000, 200, 200$ GeV, which lies mostly in signal region 5, and with $\frac{S}{\sqrt{B}} \sim 2$ in that region. (a) With no cut on the NN output; (b) Training a NN with signal region $s = 5$ (2.74–3.28 TeV), and making a cut on the NN output at efficiency $\epsilon = 0.01$. The red band indicates the uncertainty on the fit, which is due to the covariance matrix from the fit parameters. The green dashed lines indicate the signal region for training.

7.5.4 Limits

As described in Section 7.4.7.2, there is a coarse scan of injected μ values to train the NN, following which, for each injected μ, limits are set following the CL_s procedure by scanning the overall signal strength as the POI while keeping the NN fixed. The values of μ injected for this analysis are given in Table 7.6.

Table 7.6: Injected μ values.

Bin	m_a [GeV]	Values
5	3000	[750,1000,1500,2000]

The 95% confidence exclusion limit μ_{95} on a variety of signal models is shown in Figure 7.18, expressed as the limit on the cross section times branching ratio,

$\sigma_{95} \times B = \frac{\mu_{95}}{\mathcal{L}}$, where $\mathcal{L} = 139$ fb^{-1} is the total integrated luminosity. The result is shown for a few different values of ϵ, where the NN is trained with signal region $s = 5$, which is the signal region in which these signal models mostly lie. For comparison, we also show limits for the ATLAS inclusive dijet search [19] and for the ATLAS all-hadronic diboson resonance search [20] The inclusive dijet limits are calculated using the W' signals from this analysis and the full analysis pipeline of that search; in particular, small-radius jets were used, so that the limits from that search get worse at higher m_B, m_C as the small-radius jets are not sufficient to contain all of the decay products of the daughter resonances. The diboson search limits are computing using the Heavy Vector Triplet [406] W' signal used in that search.

The $\epsilon = 1$ regime of the search has no machine learning tagging and is therefore similar to the inclusive dijet search. The limits from the $\epsilon = 1$ search are about the same as the inclusive dijet search for lower m_B, m_C; but they are much better at higher m_B, m_C, because small-radius jets are often not sufficient to contain all of the decay products of the daughter resonances. As expected, the targeted diboson search is more sensitive than the relatively generic inclusive dijet search and the difference in limits is about a factor of 10. Away from the SM dibosons, CWoLa hunting with $\epsilon < 1$ outperforms the other searches by as much as a factor of 5 for the heaviest masses, and almost a factor of 2 for the lighter masses where the inclusive dijet search is most sensitive.

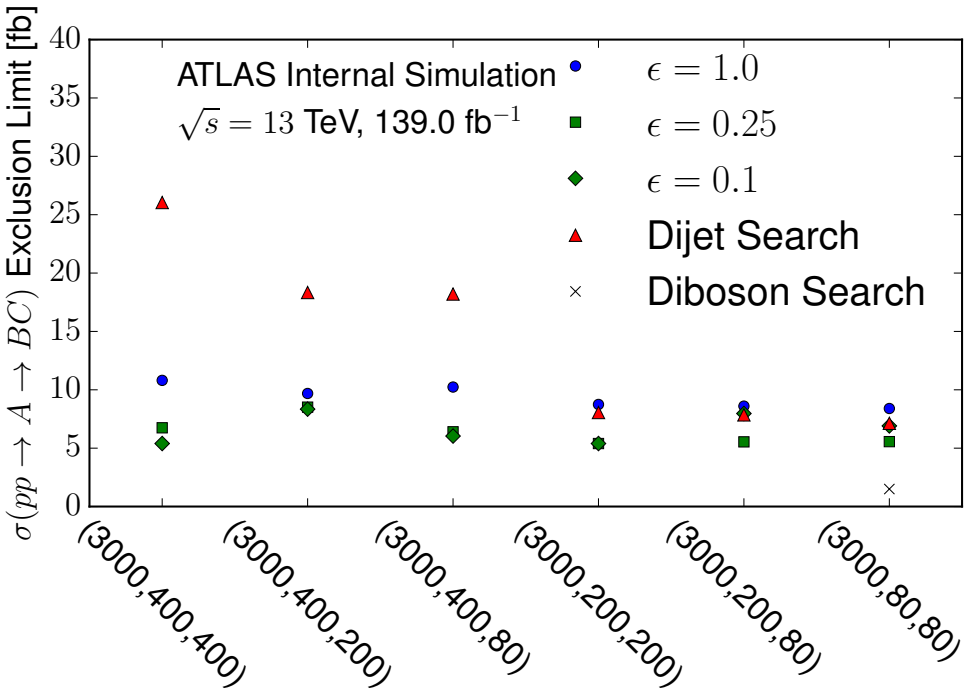


Figure 7.18: 95% confidence exclusion limits on the cross section of a variety of signal models, labeled by (m_A, m_B, m_C), in GeV. The limits are shown for NN cuts at $\epsilon = [1.0, 0.25, 0.1]$. Also shown are the limits from the ATLAS dijet search [19] and the ATLAS all-hadronic diboson search [20].

7.6 Validation Analysis

In this section, the analysis is validated on a data sample with an inverted rapidity cut, as detailed in section 7.6.1.

7.6.1 Event Selection

The event selection is similar to the standard selection from Section 7.4.1, only that the rapidity cut is inverted: $|y_1 - y_2| > 1.2$. Note that the signal samples have the usual non-inverted cut applied ($|y_1 - y_2| \leq 1.2$).

The distributions of $|y_1 - y_2| > 1.2$, η_1, and η_2 after this selection are shown in Figure 7.19. The jet mass distributions in a few representative signal regions are

shown in Figure 7.20.

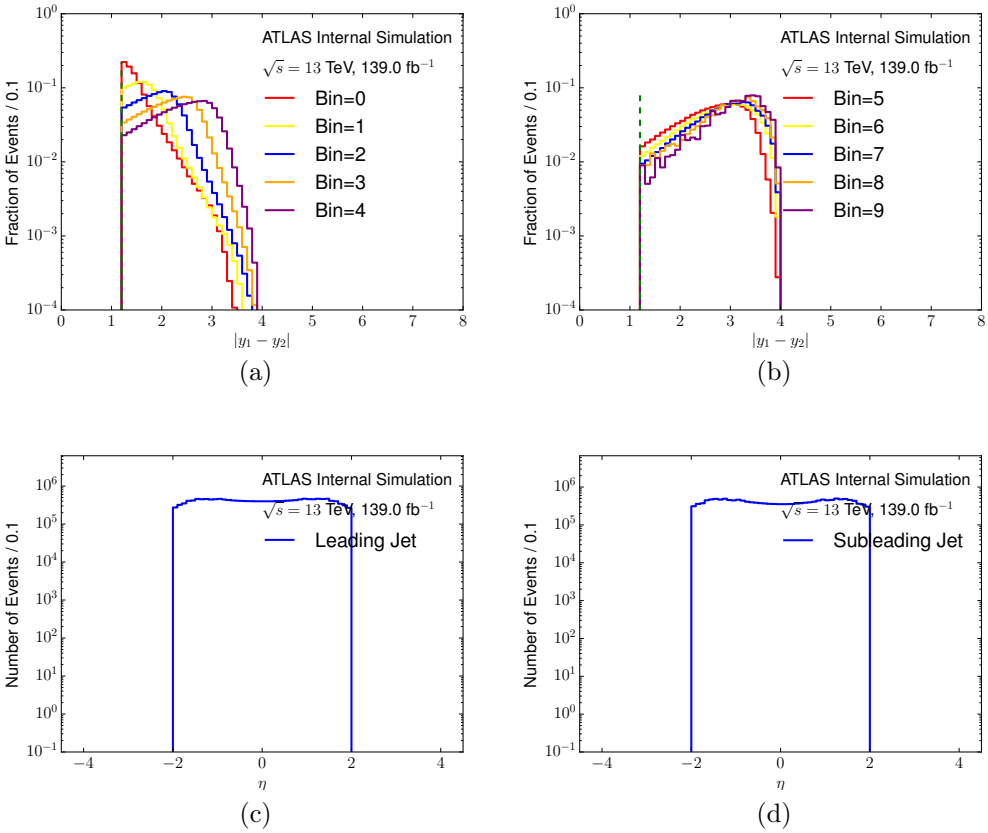
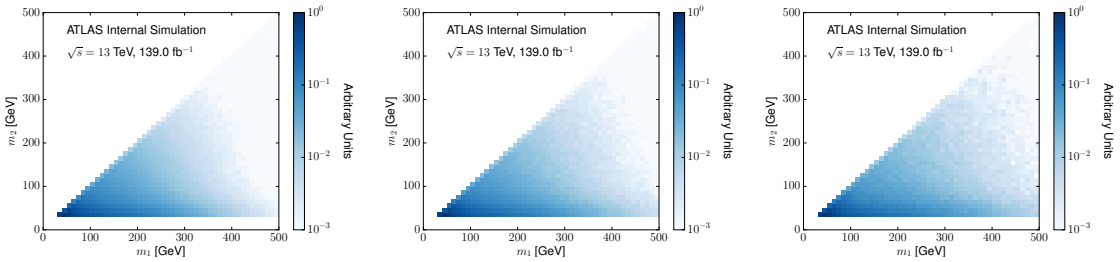


Figure 7.19: The difference in rapidity for the two selected jets. The η distributions of both jets are shown below. Note that these are data plots and not simulation, using the inverted rapidity cut data selection.



(a) (b) (c)

Figure 7.20: The mass distribution of the two selected jets in (a) signal region 4; (b) signal region 5; (c) signal region 6. Note that these are data plots and not simulation, using the inverted rapidity cut data selection.

7.6.2 Neural Network Output

The output of the neural network training with no injected signal is presented in Figure 7.21. There are some features in this neural network output, corresponding to likely statistical fluctuations, especially at high (m_1, m_2) where the data are poorly populated by events.

The output of the neural network training with an injected signal is presented in Figure 7.22. In all cases, the location of the injected signal aligns well with an identified region in the neural network map. The $(80, 80)$ GeV point is the hardest to find as it would sit on top of the largest background from all of the benchmarks.

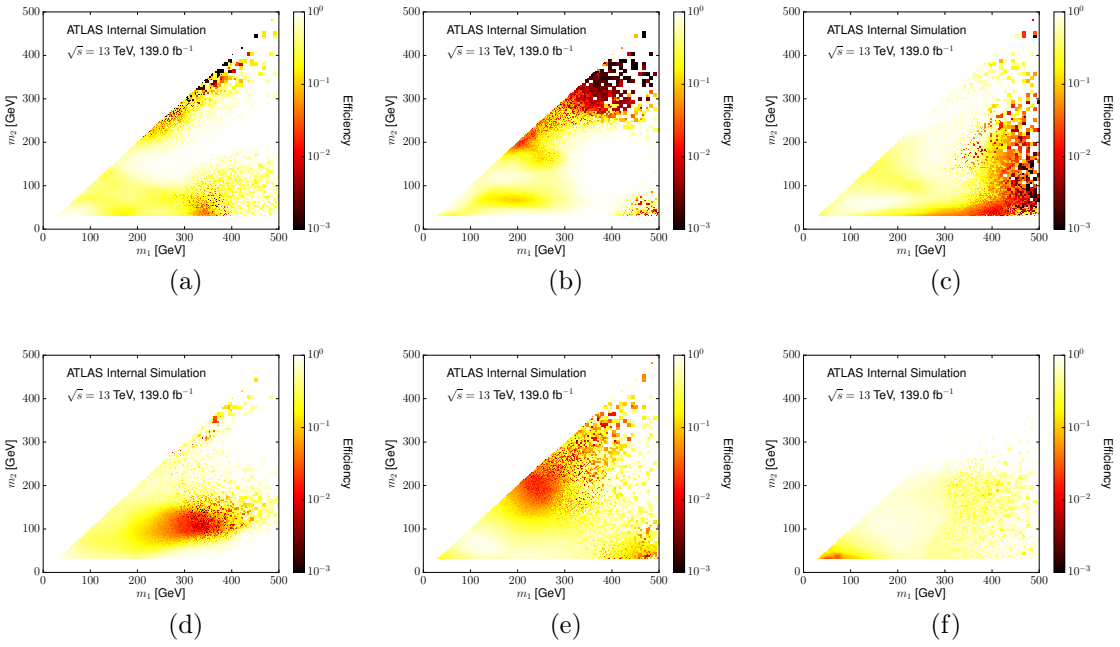


Figure 7.21: The output of the neural network when there is no injected signal, in (a) signal region 4; (b) signal region 5; (c) signal region 6; (d) signal region 7; (e) signal region 8; (f) signal region 9. Note that these are data and not simulation, using the inverted rapidity cut data selection.

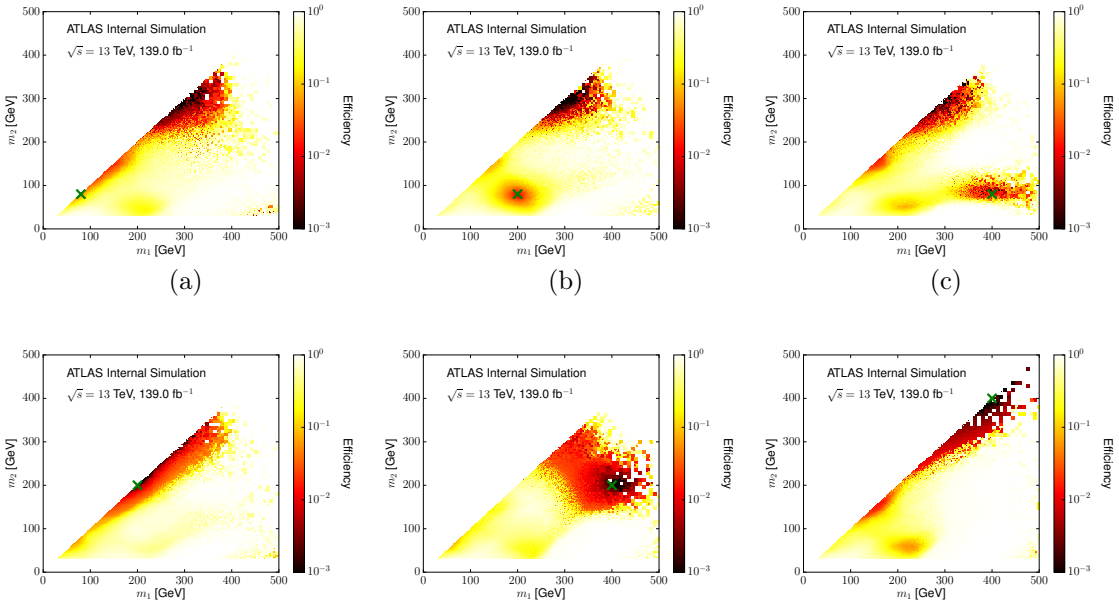


Figure 7.22: Neural network output for signal region 5 with $\mu = 1000$ for a signal with $m_A = 3000$ GeV and (a) $(m_B, m_C = 80, 80$ GeV); (b) $(m_B, m_C = 200, 80$ GeV); (c) $(m_B, m_C = 400, 80$ GeV); (d) $(m_B, m_C = 200, 200$ GeV); (e) $(m_B, m_C = 400, 200$ GeV); (f) $(m_B, m_C = 400, 400$ GeV). Note that these are data and not simulation, using the inverted rapidity cut data selection.

7.6.3 Fitting Results

The fit results for signal region 5, with no injected signal, are shown in Figure 7.23. In this region, there is a slight excess in the masked region for the $\epsilon = 1.0$ dataset, which goes away for $\epsilon < 1$. This indicates that the NN is not sculpting the background distribution in favor of the signal region, in agreement with Assumption 2.

The fit results with no injected signal for the other signal regions are included in Appendix C.7. These fit results with no injected signal indicate that the fit functions are insufficient to describe the background shape at $\epsilon = 1$ and $\epsilon = 0.25$, motivating the choice to limit the analysis to only $\epsilon = 0.1$ and $\epsilon = 0.01$ for the full unblinded analysis. In all signal regions, at these lower values of ϵ, there is no evidence of sculpting of the background, adding support to the claim that Assumption 2 is holding true.

Even at these lower values of ϵ, there is still a slight bias in the fit results; this bias is measured with these fits and applied as a correction, as a function of m_{JJ}, on the fit results in the full unblinded fit results. The correction is validated in the sidebands of the fits in the full unblinded dataset, and the uncertainty on this correction is added as an additional uncertainty on the fit; this correction is explained in detail in Appendix C.10.

The fit results for signal region 5, with an injected signal at $m_A = 3000, m_B = 200, m_C = 200$ GeV, for various efficiency points ϵ, are shown in Figure 7.24. The injected signal is a $\sim 1.5\sigma$ excess in the $\epsilon = 1.0$ case (note that there is a slight excess in the data for $\epsilon = 1.0$ already, so the observed p-value in the signal region there is artificially low), and would not be excluded at the 95% confidence level. After cutting on the output of the NN, the background is reduced while the signal remains, until at $\epsilon = 0.01$ the excess can be seen by eye and this level of signal would be clearly excluded.

The fit results for signal region 5, with a variety of injected signals with $m_A = 3000$ GeV, at either efficiency $\epsilon = 0.1$ or $\epsilon = 0.01$ depending on the signal, are shown in Figure 7.25. The injected signal is a $\sim 1.7\sigma$ excess in the $\epsilon = 1.0$ case, and would not be excluded at the 95% confidence level. After cutting on the output of the NN, the background is reduced while the signal remains. For the mass

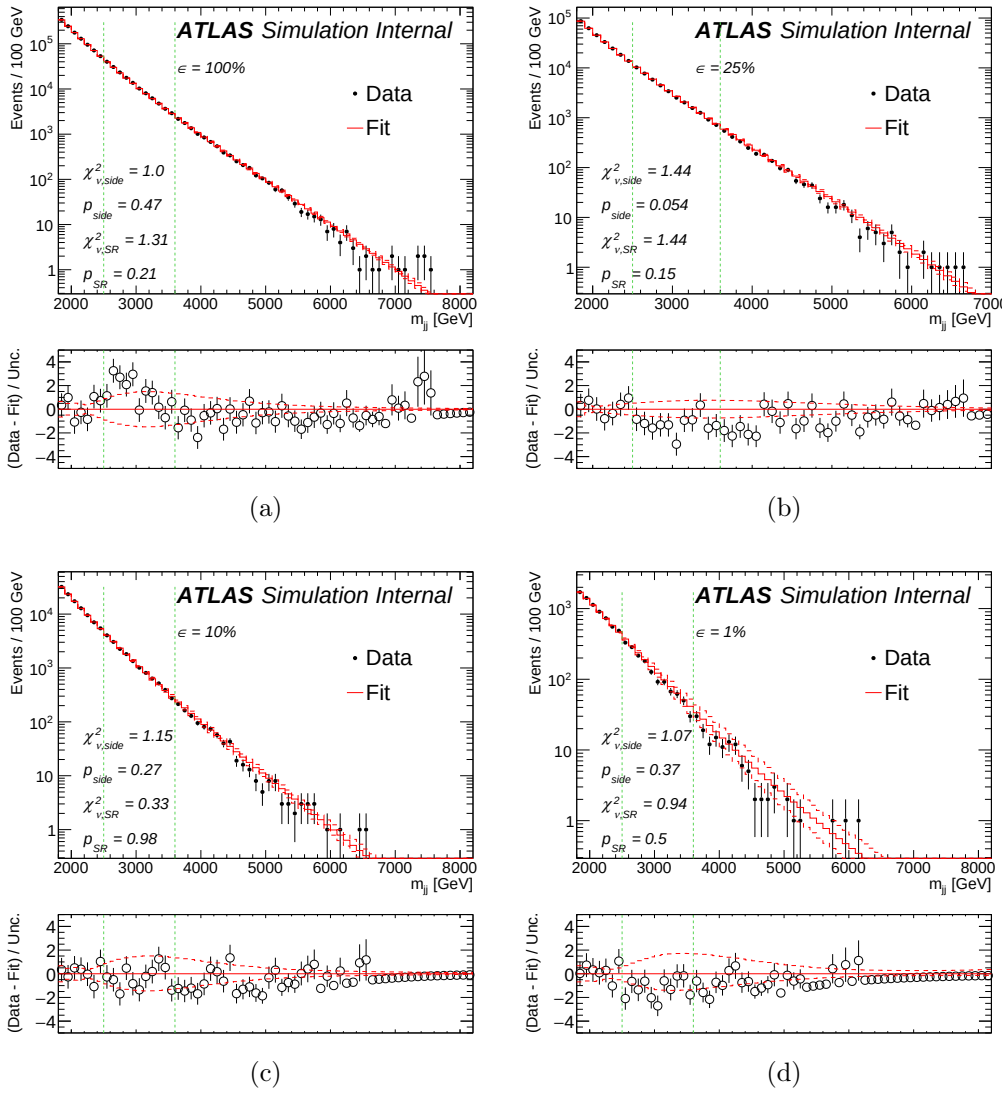


Figure 7.23: The background fit when there is no injected signal, in signal region 5, for various efficiency points ϵ. Note that these are data and not simulation, using the inverted rapidity cut data selection. The red dashed lines indicate the fit uncertainty. The green dashed lines indicate the signal region for training.

points at $m_B, m_C = (80, 80), (80, 200), (80, 400)$ GeV, the efficiency point $\epsilon = 0.1$ is shown. At this cut level, for the higher mass points, there is a maximum excess of around 3σ relative to the background expectation. For the mass points at $m_B, m_C = (200, 200), (200, 400), (400, 400)$ GeV, the efficiency point $\epsilon = 0.01$ is

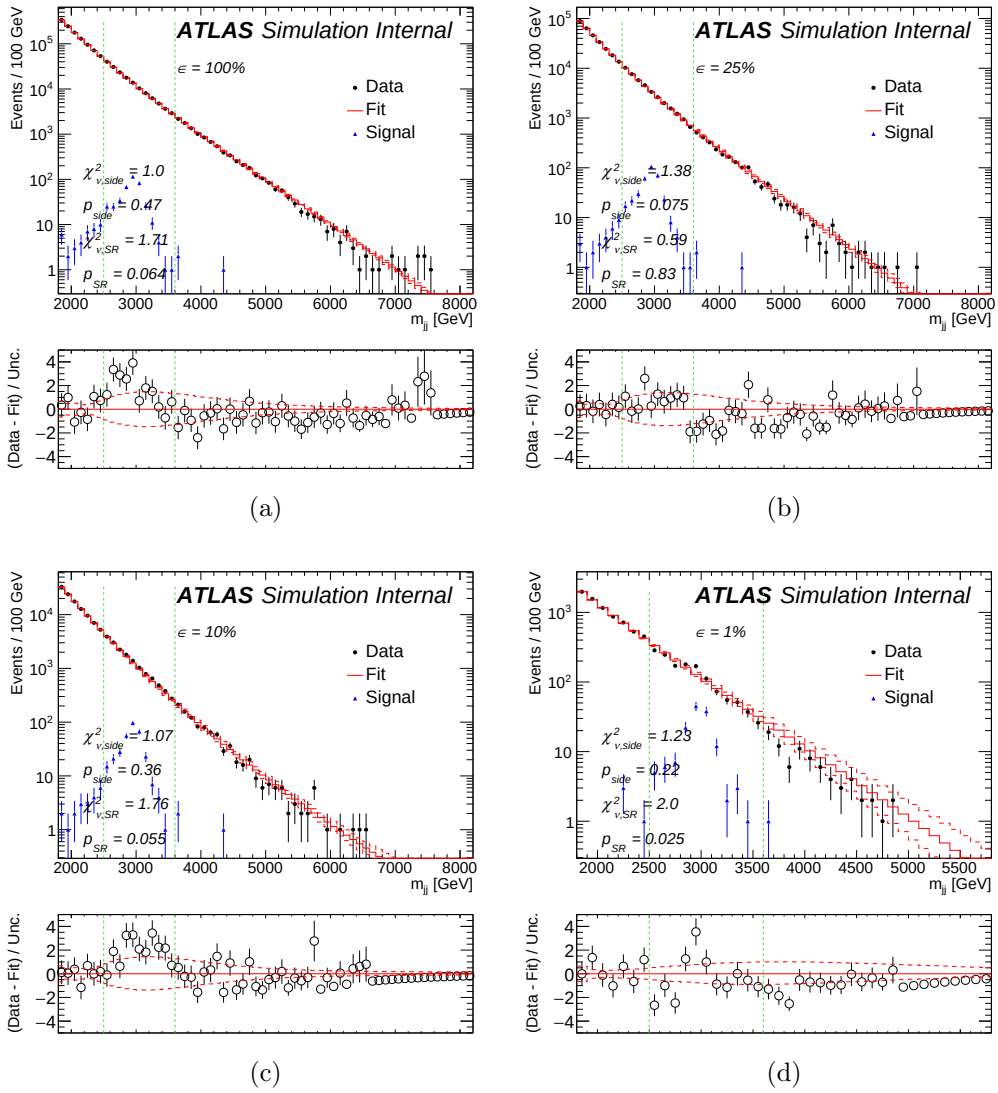


Figure 7.24: The fit with an injected signal at $m_A = 3000, m_B = 200, m_C = 200$ GeV, in signal region 5, for various efficiency points ϵ. The strength of the signal is $\mu = 750$, corresponding to a $\sim 1.5\sigma$ excess in this bin. **Note that the data points include both the observed data and the injected signal**. Note that other than the injected signal, these are data and not simulation, using the inverted rapidity cut data selection. The red dashed lines indicate the fit uncertainty. The green dashed lines indicate the signal region for training.

shown. At this cut level, for these mass points, there is a maximum excess of $\geq 5\sigma$

relative to the background expectation.

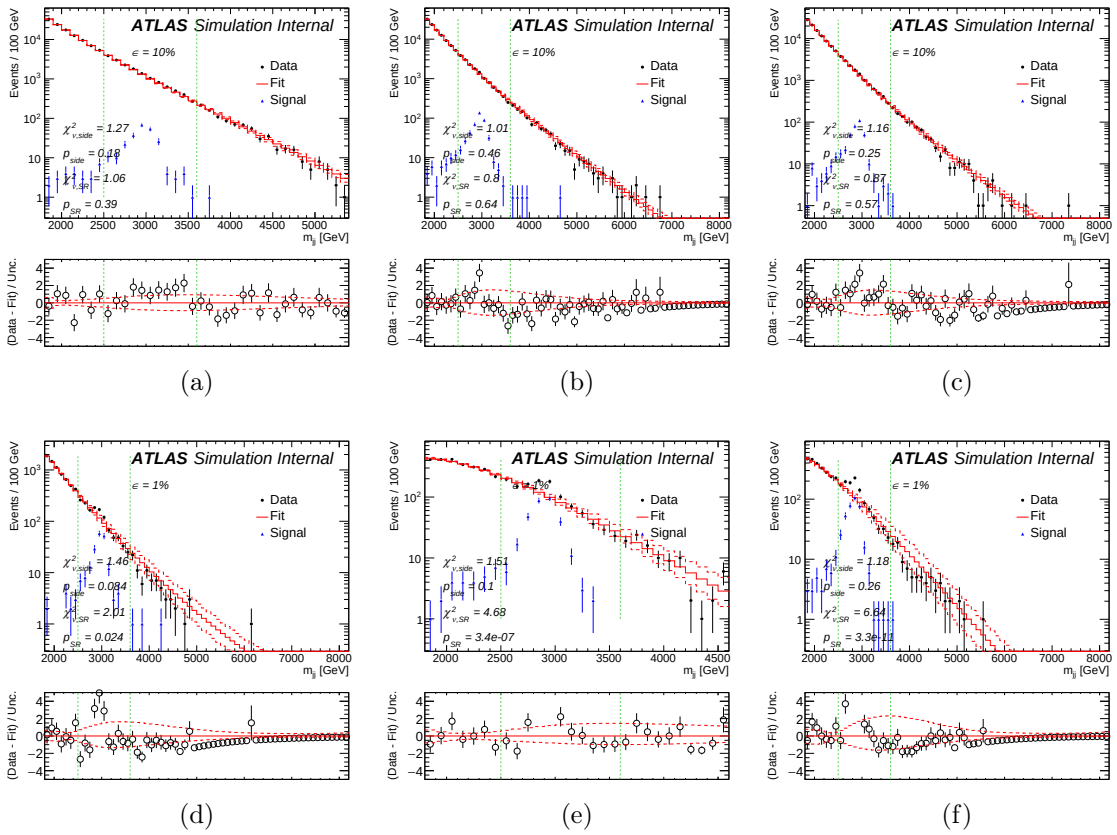


Figure 7.25: The fit with an injected signal at $m_A = 3000$ GeVand (a) $m_B = 80, m_C = 80$; (b) $m_B = 80, m_C = 200$; (c) $m_B = 80, m_C = 400$; (d) $m_B = 200, m_C = 200$; (e) $m_B = 200, m_C = 400$; and (f) $m_B = 400, m_C = 400$ GeV, in signal region 5, for efficiency points $\epsilon = 0.1$ (a,b,c) and $\epsilon = 0.01$ (d,e,f). The strength of the signal is $\mu = 1000$, corresponding to a $\sim 1.7\sigma$ excess in this bin. **Note that the data points include both the observed data and the injected signal**. Note that other than the injected signal, these are data and not simulation, using the inverted rapidity cut data selection. The red dashed lines indicate the fit uncertainty. The green dashed lines indicate the signal region for training.

7.6.4 Limits

The values of μ injected for this analysis are given in Table 7.7.

Table 7.7: Injected μ values.

Bin	m_a [GeV]	Values
5	3000	[500,750,1000,1500]
8	5000	[50,75,100,150]

The 95% confidence exclusion limit μ_{95} on a variety of signal models is shown in Figure 7.26, expressed as the limit on the cross section times branching ratio, $\sigma_{95} \times B = \frac{\mu_{95}}{\mathcal{L}}$, where $\mathcal{L} = 139$ fb^{-1} is the total integrated luminosity. Note that these are not real limits, as the validation data has an inverted rapidity cut applied while the signal samples have the usual non-inverted cut. However, since the validation data has been rescaled to match the expected statistics in the non-inverted region, as explained in Section 7.3.3, these results can be considered as estimates for the exclusion power of the method in the final non-inverted signal region. The result is shown for a few different values of ϵ, where the NN is trained with signal region $s = 5$, which is the signal region in which these signal models mostly lie. For comparison, we also show limits for the ATLAS inclusive dijet search [19] and for the ATLAS all-hadronic diboson resonance search [20] The inclusive dijet limits are calculated using the W' signals from this analysis and the full analysis pipeline of that search; in particular, small-radius jets were used, so that the limits from that search get worse at higher m_B, m_C as the small-radius jets are not sufficient to contain all of the decay products of the daughter resonances. The diboson search limits are computing using the Heavy Vector Triplet [406] W' signal used in that search.

The $\epsilon = 1$ regime of the search has no machine learning tagging and is therefore similar to the inclusive dijet search. The limits from the $\epsilon = 1$ search are about the same as the inclusive dijet search for lower m_B, m_C; but they are much better at higher m_B, m_C, because small-radius jets are often not sufficient to contain all of the decay products of the daughter resonances. As expected, the targeted diboson search is more sensitive and the difference in limits is about a factor of 10. Away from the SM dibosons, CWoLa hunting with $\epsilon < 1$ outperforms the other searches by as much as a factor of 5 for the heaviest masses, and almost a factor of 2 for the lighter masses where the inclusive dijet search is most sensitive.

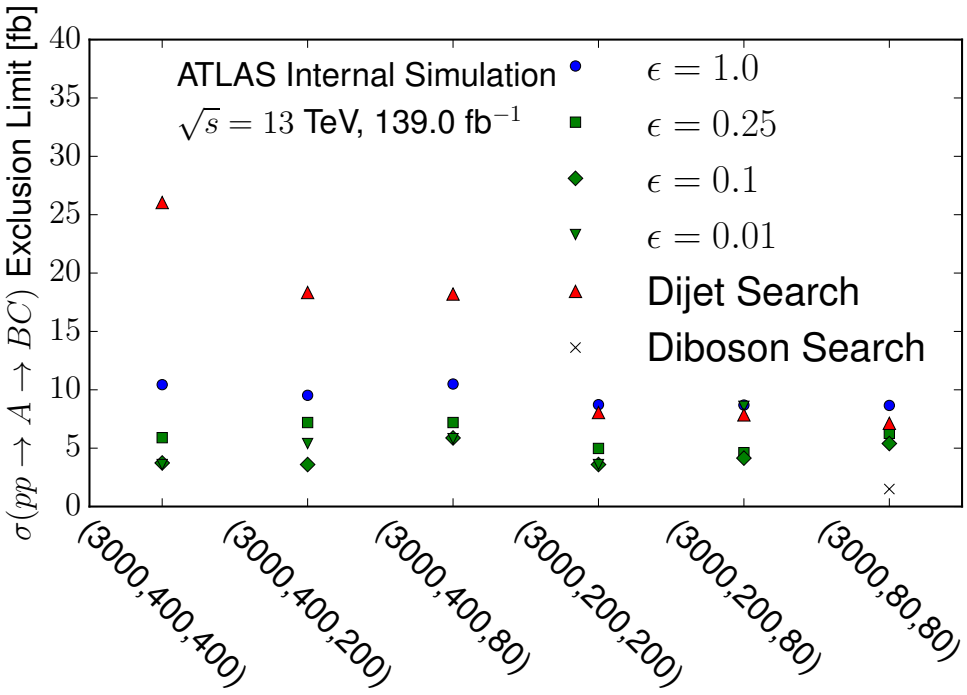


Figure 7.26: 95% confidence exclusion limits on the cross section of a variety of signal models, labeled by (m_A, m_B, m_C), in GeV. The limits are shown for NN cuts at $\epsilon = [1.0, 0.25, 0.1]$. Also shown are the limits from the ATLAS dijet search [19] and the ATLAS all-hadronic diboson search [20]. Note that for the CWoLa limits these are data and not simulation, using the inverted rapidity cut data selection.

7.7 Unblinded Analysis

In this section, the final analysis is performed on the unblinded dataset. The key difference to the analysis in Section 7.6 is that the rapidity cut is not inverted, as detailed in section 7.3.3.

7.7.1 Event Selection

The event selection is the same as the one outlined in Section 7.4.1, with the uninverted rapidity cut: $|y_1 - y_2| \leq 1.2$. The distributions of $|y_1 - y_2|$, η_1, and η_2 after this selection are shown in Figure 7.27. The jet mass distributions in a few representative signal regions are shown in Figure 7.28.

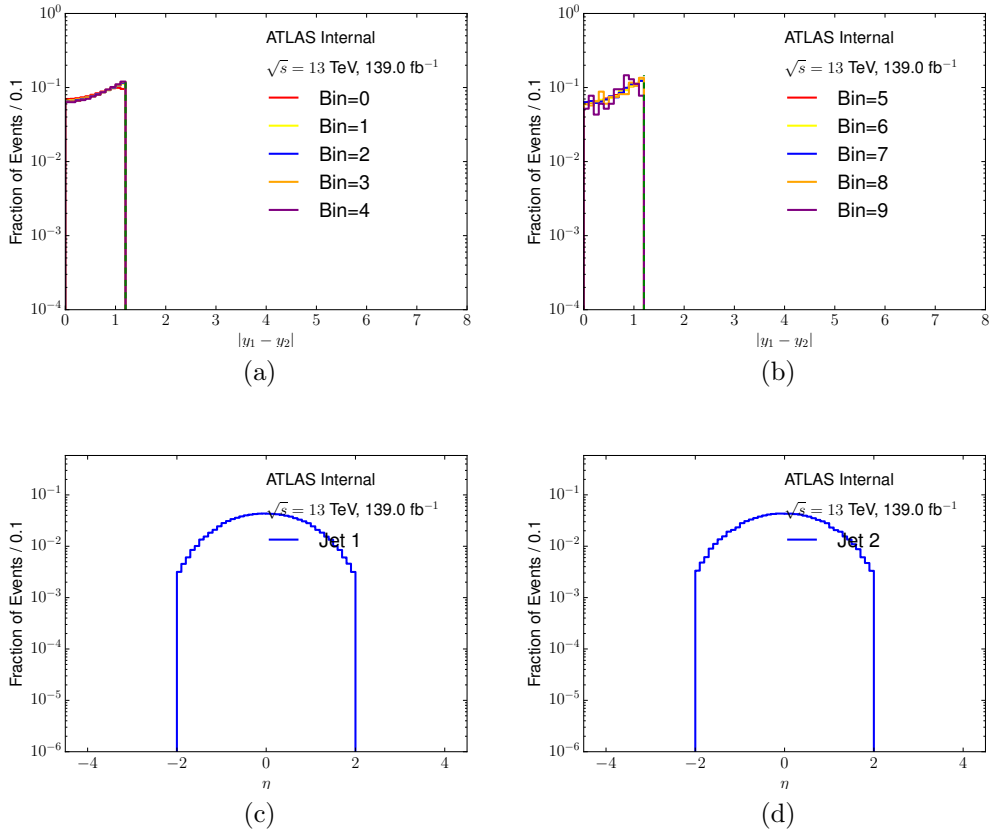


Figure 7.27: The difference in rapidity for the two selected jets. The binning is described in Section 7.7.2. The η distributions of both jets are shown below.

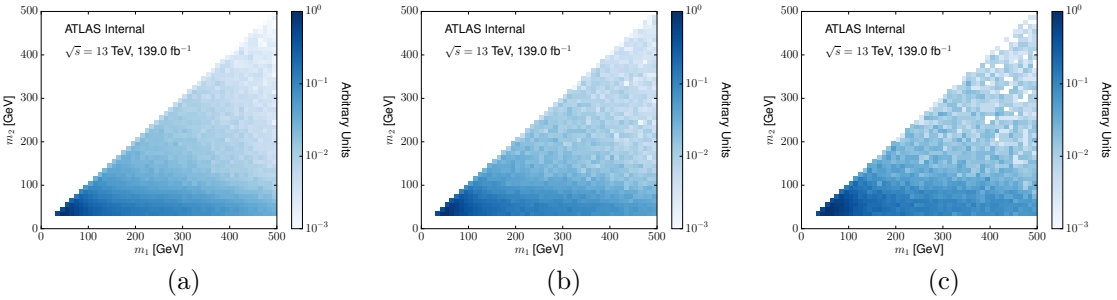


Figure 7.28: The mass distribution of the two selected jets in (a) signal region 4; (b) signal region 5; (c) signal region 6. The binning is described in Section 7.7.2.

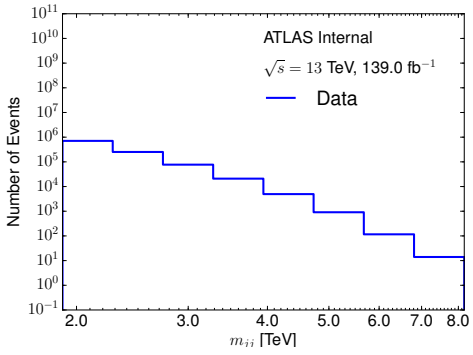
As mentioned in Section 7.4.6, the efficiency values used in this analysis are restricted to $\epsilon = \{0.1, 0.01\}$. These values are listed in Table 7.8.

Table 7.8: Chosen values of NN cuts with efficiency ϵ for analysis.

	Values
ϵ	[0.1,0.01]

7.7.2 Binning

The binning for the learning is the same as in Section 7.4.2, with signal regions starting from bin 4 and going up to bin 9, corresponding to the signal regions that fit into the ultimate fit range ($1.8 < m_{JJ} < 8.2$ TeV) with room for sidebands. The number of events in each of the bins 3-10 (bins 3 and 10 are used as sidebands, though not as signal regions) is shown in Figure 7.29.



(a)

Figure 7.29: The number of events in each of the bins 3-10.

7.7.3 Neural Network Output

The output of the neural network training with no injected signal is presented in Figure 7.30. There are some features in this neural network output, corresponding to likely statistical fluctuations, especially at high (m_1, m_2) where the data are poorly

populated by events. Note that by construction, the NN has to have a low efficiency somewhere and due to their smoothness, it is likely that the regions of low efficiency are also relatively localized.

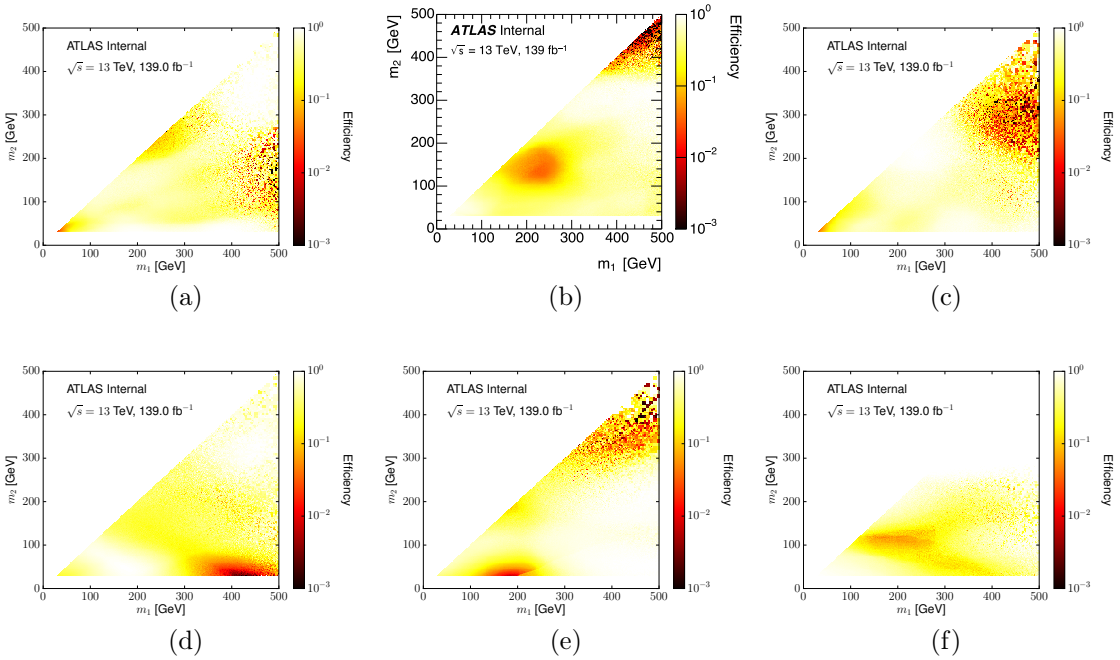


Figure 7.30: The output of the neural network when there is no injected signal, in signal region (a) 4; (b) 5; (c) 6; (d) 7; (e) 8; and (f) 9.

The output of the neural network training is presented with an injected signal at $m_A = 3000$ GeV in Figure 7.31 and with an injected signal at $m_A = 5000$ GeV in Figure 7.32. In each case, the NN with the median efficiency on the signal from the 5 random samplings of the signal is shown. These plots show the areas of the (m_1, m_2) plane that the network has identified as signal-like. In all cases, the location of the injected signal aligns well with an identified region in the neural network map. The $(80, 80)$ GeV point is the hardest to find as it sits on top of the largest background from all of the benchmarks. The NN output for all 5 random samplings as a function of μ is given in Appendix C.9.

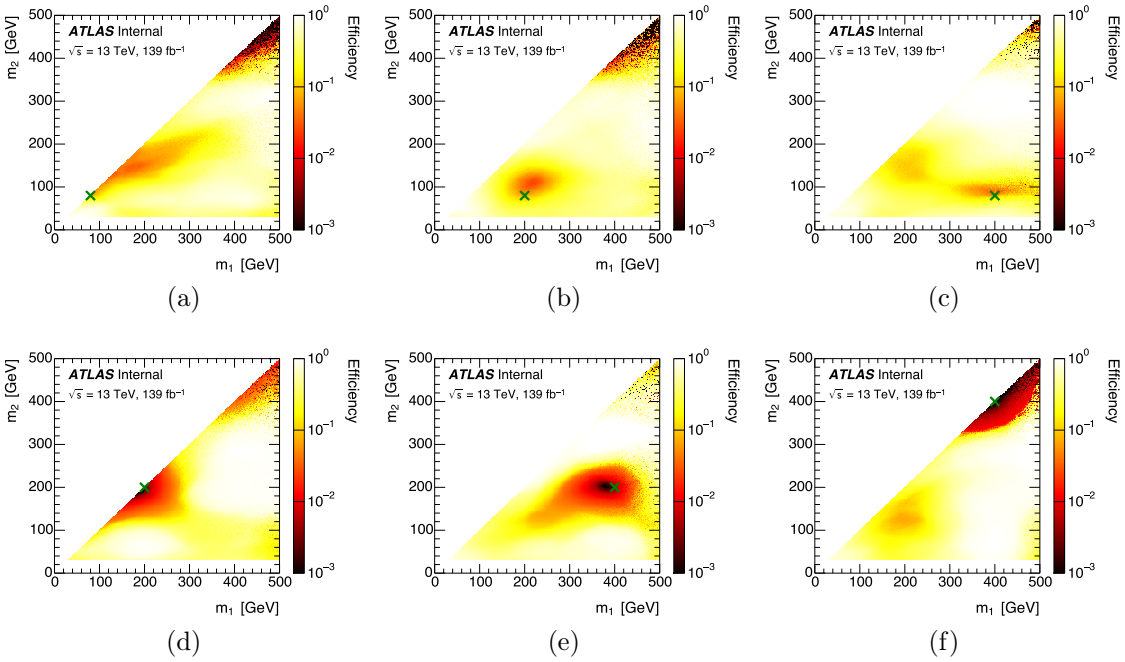


Figure 7.31: Neural network output for signal region 5 with $\mu = 1000$ for a signal with $m_A = 3000$ GeV and (a) $(m_B, m_C = 80, 80$ GeV); (b) $(m_B, m_C = 200, 80$ GeV); (c) $(m_B, m_C = 400, 80$ GeV); (d) $(m_B, m_C = 200, 200$ GeV); (e) $(m_B, m_C = 400, 200$ GeV); (f) $(m_B, m_C = 400, 400$ GeV).

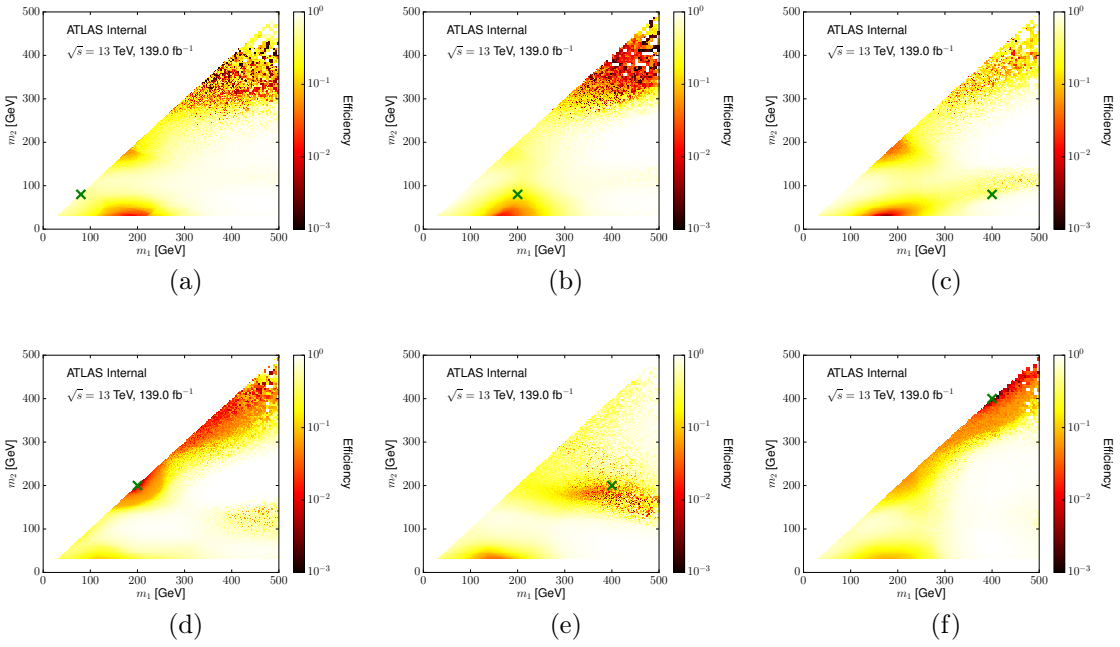


Figure 7.32: Neural network output for signal region 8 with $\mu = 350$ for a signal with $m_A = 5000$ GeV and (a) $(m_B, m_C = 80, 80$ GeV); (b) $(m_B, m_C = 200, 80$ GeV); (c) $(m_B, m_C = 400, 80$ GeV); (d) $(m_B, m_C = 200, 200$ GeV); (e) $(m_B, m_C = 400, 200$ GeV); (f) $(m_B, m_C = 400, 400$ GeV).

7.7.4 Fitting Results

The fit results for all the signal regions, with no injected signal, are shown in Figure 7.33 for $\epsilon = 0.1$ and Figure 7.34 for $\epsilon = 0.01$. The largest local excess in the masked regions occurs in signal region 4 around 2500 GeV, with a maximum local significance of 2.93 at $\epsilon = 0.1$ and 1.85 at $\epsilon = 0.01$. If this excess corresponds to a signal, the NN output for this signal region indicates an excess around $m_1, m_2 \sim (> 450, 150)$ GeV. The distribution of significances, in particular the rate of large excesses, is consistent with the background-only expectation (Appendix C.11) at around the 1.3σ level (i.e. the rate of large excesses is this large or larger $\sim 10\%$ of the time).

Some fit results with an injected signal are shown in Figure 7.35. For signal region 5 there is an injected signal at $m_A = 3000, m_B = 200, m_C = 200$ GeV, with $\mu = 500$; the injected signal is a $\sim 1.0\sigma$ excess in the $\epsilon = 1.0$ case, and would not be excluded at the 95% confidence level. For signal region 8 there is an injected signal at $m_A = 5000, m_B = 200, m_C = 200$ GeV, with $\mu = 225$; the injected signal is a $\sim 1.5\sigma$ excess in the $\epsilon = 1.0$ case, and would not be excluded at the 95% confidence level. In both cases, after cutting on the output of the NN, the background is reduced while the signal remains; the excess could possibly be seen by eye and this level of signal would be clearly excluded.

The fit results for all the signals at the injected μ value that gives rise to the limits given in Section 7.7.5 can be found in Appendix C.12.

These fits are summarized in Figure 7.36. For each signal region training, only the region of the fit that corresponds to the signal region is shown, and these regions are stitched together across the signal regions 4-9. Because the signal regions from independent fits (with independently trained NNs, and therefore possibly entirely different selections) are stitched together, it is possible for the background fit to be discontinuous when crossing from one signal region to another, even if the background fit is smooth within a single signal region. The figure also shows simultaneously a signal injected at $m_A = 3000$ GeV trained in signal region 5 and a signal injected at $m_A = 5000$ GeV trained in signal region 8. In each case, the signal strength is just less than that already excluded by existing searches (Section 7.7.5). After tagging with the NN, the signal at $m_A = 3000$ GeV would be easily discovered, and the signal

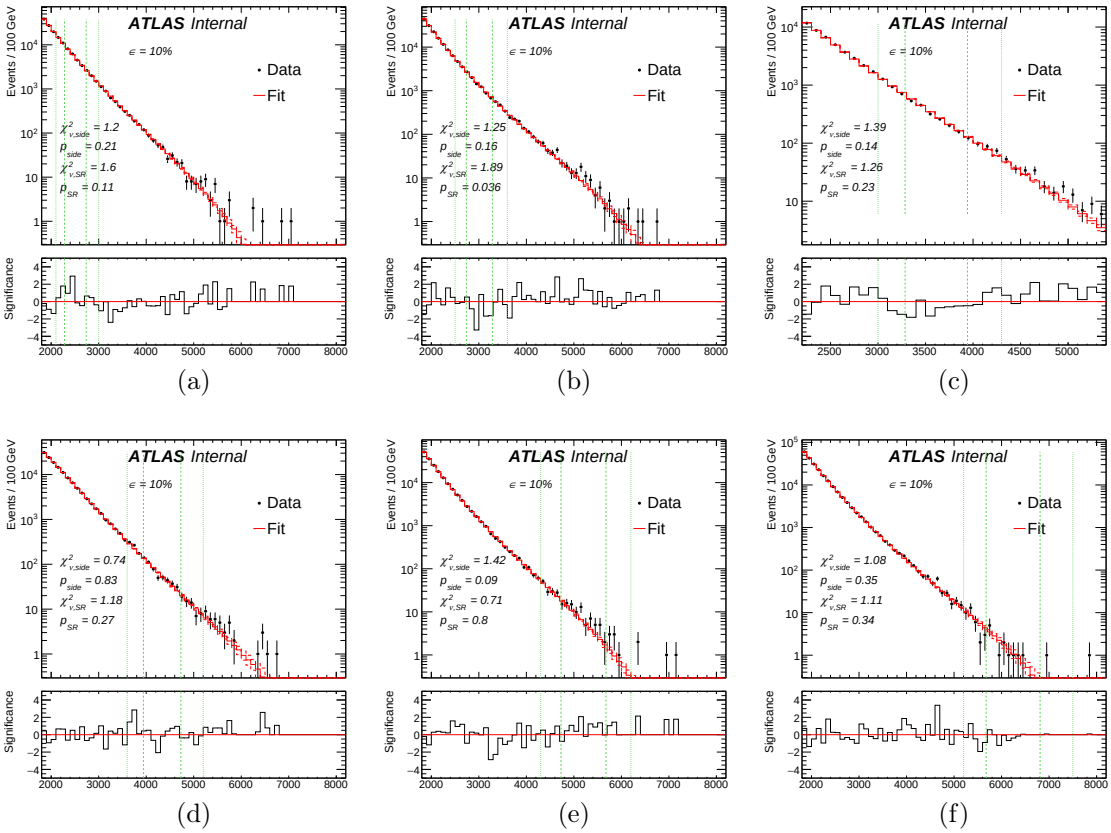


Figure 7.33: The background fit when there is no injected signal, for efficiency point $\epsilon = 0.1$, in signal region (a) 4; (b) 5; (c) 6; (d) 7; (e) 8; and (f) 9. The red dashed lines indicate the fit uncertainty. The green dashed lines indicate the signal region for training, and the green dotted lines indicate the masked region for evaluating the fit quality.

at $m_A = 5000$ GeV would form a noticeable excess. Note that in signal region 8, after tagging the signal constitutes a noticeable fraction of the total number of events; since the NN has a fixed efficiency in the signal region, the amount of background goes down proportionally.

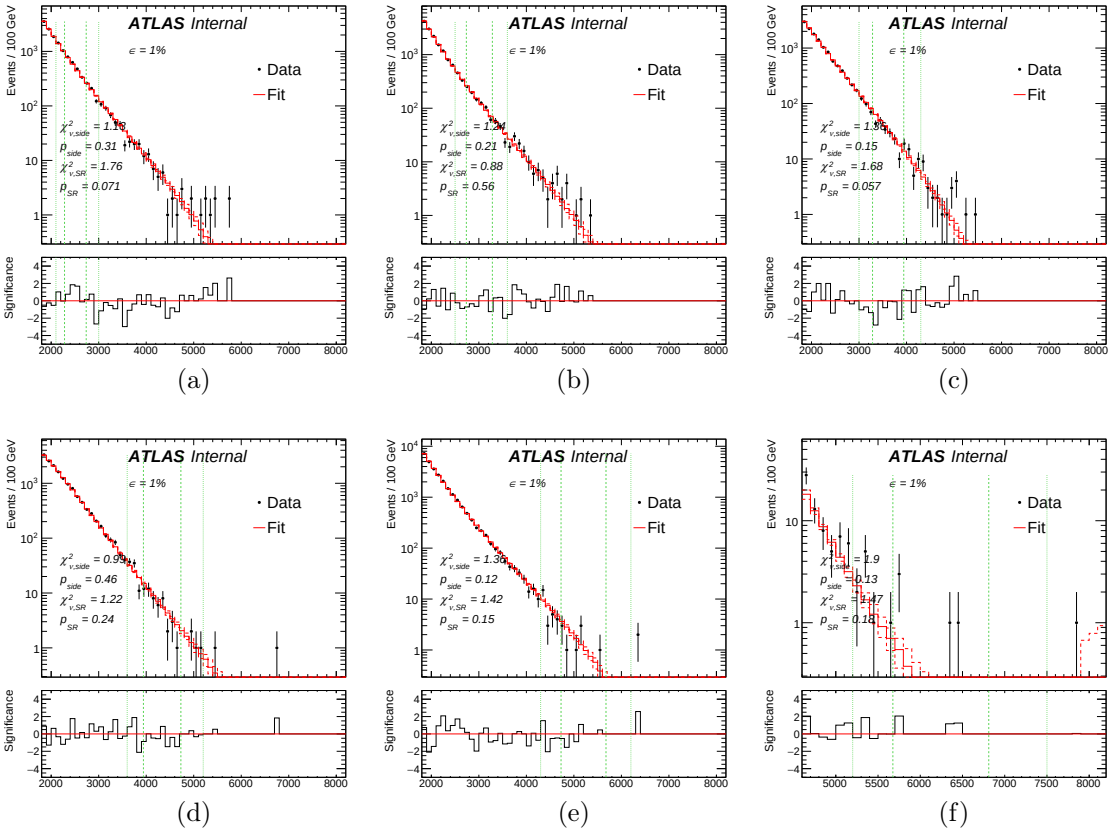


Figure 7.34: The background fit when there is no injected signal, for efficiency point $\epsilon = 0.01$, in signal region (a) 4; (b) 5; (c) 6; (d) 7; (e) 8; and (f) 9. The red dashed lines indicate the fit uncertainty. The green dashed lines indicate the signal region for training, and the green dotted lines indicate the masked region for evaluating the fit quality.

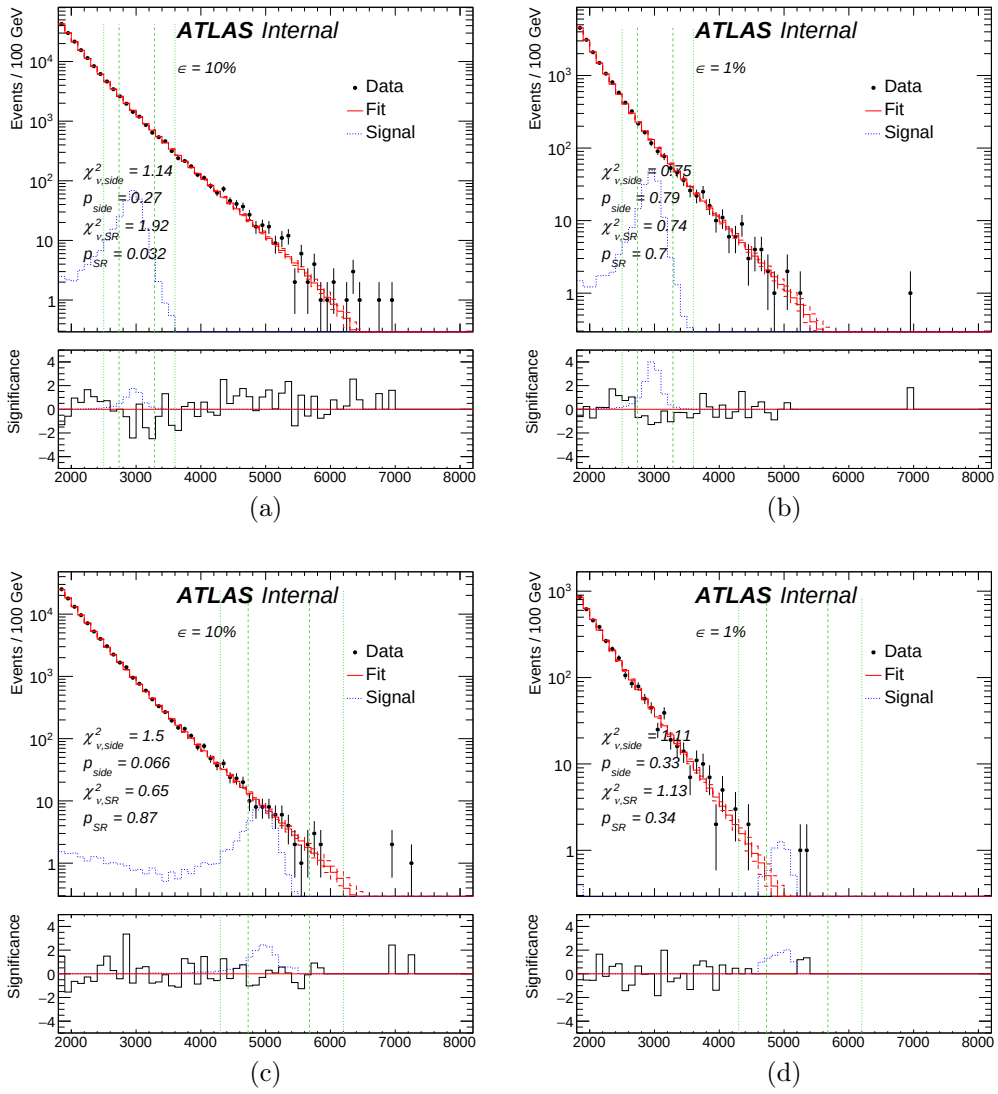


Figure 7.35: (a,b) The fit with an injected signal at $m_A = 3000, m_B = 200, m_C = 200$ GeV, in signal region 5, for the two efficiency points ϵ. The strength of the signal is $\mu = 500$, corresponding to a $\sim 1.0\sigma$ excess in this bin with no cuts. (c,d) The fit with an injected signal at $m_A = 5000, m_B = 200, m_C = 200$ GeV, in signal region 8, for the two efficiency points ϵ. The strength of the signal is $\mu = 225$, corresponding to a $\sim 1.5\sigma$ excess in this bin with no cuts. The red dashed lines indicate the fit uncertainty. The green dashed lines indicate the signal region for training, and the green dotted lines indicate the masked region for evaluating the fit quality.

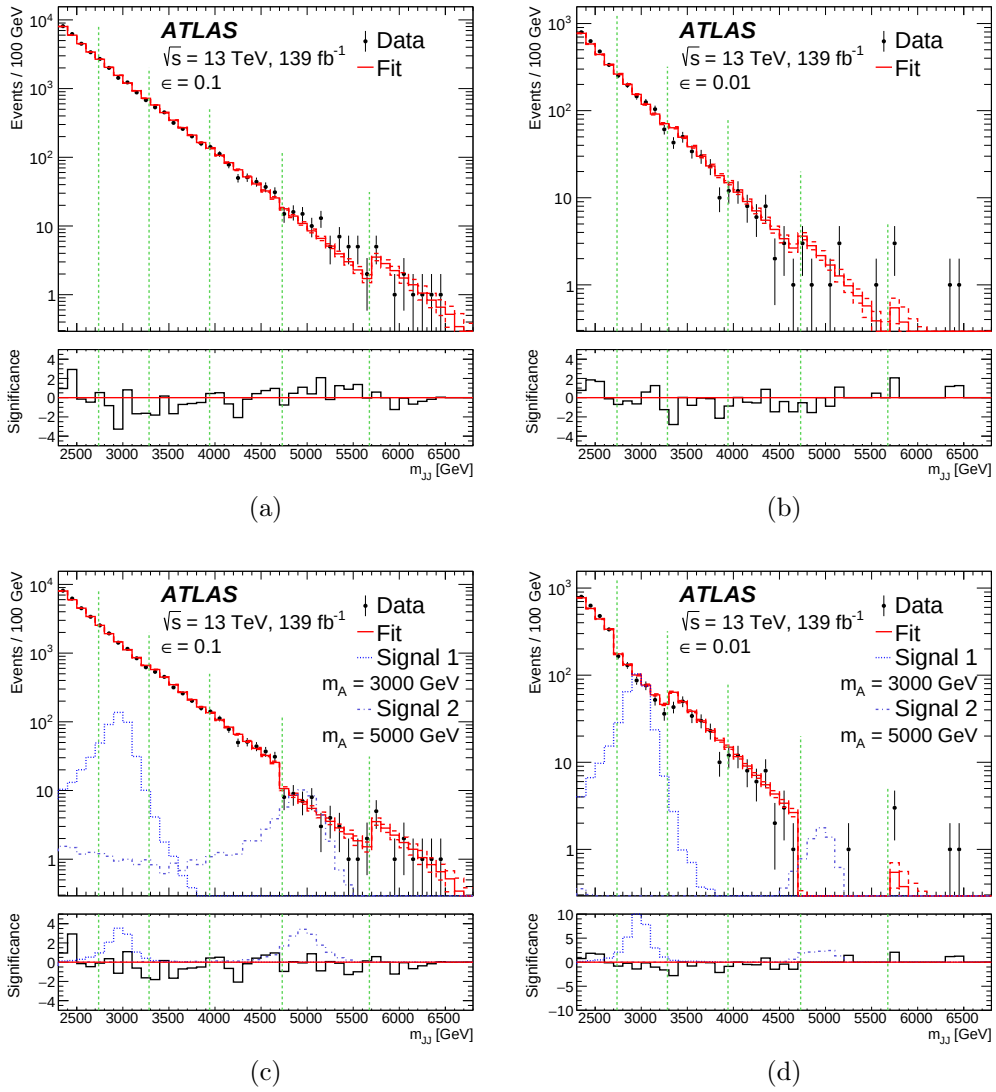


Figure 7.36: (a,b) The fit in each of the signal regions with no injected signal, stitched together across the signal region trainings, for $\epsilon = 0.1$ and $\epsilon = 0.01$, respectively. (c,d) The fit in each of the signal regions stitched together across the signal region trainings, with an injected signal at $m_A = 3000, m_B = 200, m_C = 200$ GeV with signal strength $\mu = 1000$ in signal region 5, and with an injected signal at $m_A = 5000, m_B = 200, m_C = 200$ GeV with signal strength $\mu = 280$ in signal region 8, for $\epsilon = 0.1$ and $\epsilon = 0.01$, respectively. In each case, the injected signal strength corresponds to just less than a 2.0σ excess, and are at just less than the level already excluded by existing searches. The green dashed lines indicate the boundaries of the signal regions. The red dashed lines indicate the fit uncertainty.

7.7.5 Limits

The values of μ injected for this analysis are given in Table 7.9.

Table 7.9: Injected μ values.

Bin	m_a [GeV]	Values
5	3000	[150, 185, 225, 350, 500, 600, 750, 850, 1000, 1250, 1500]
8	5000	[50, 60, 75, 100, 150, 225, 280, 350, 500, 750, 1000]

The 95% confidence exclusion limits μ_{95} are derived as discussed in Section 7.4.7.2. The dependence of the limits on μ for each of the 5 random samplings of the signal is given in Appendix C.13. As discussed in Section 7.4.7.2, the limit that is used for the given μ value is the limit derived from the sampling which gives the median expected limit over the 5 random samplings. The derived limits for each of the 5 random samplings are given in Appendix C.13 - for most signals where non-trivial limits are set, the additional variance from the random samplings is negligible relative to the original bands. The final expected limits that are used are the limits derived from the μ value that gives the minimum median expected limit, after setting every limit to the $\max(\mu, \mu_{95}(\mu))$. The 95% confidence exclusion limits μ_{95} on a variety of signal models is shown in Figure 7.37, expressed as the limit on the cross section times branching ratio, $\sigma_{95} \times B = \frac{\mu_{95}}{\mathcal{L}}$, where $\mathcal{L} = 139$ fb^{-1} is the total integrated luminosity. The result is shown for the two different values of $\epsilon = 0.1$ and $\epsilon = 0.01$, and for two different signal models: a signal model with $m_A = 3000$ GeV and NN trained on signal region 5; and a signal model with $m_A = 5000$ GeV and NN trained on signal region 8.

For comparison, limits are also shown for the ATLAS inclusive dijet search [19] and for the ATLAS all-hadronic diboson resonance search [20] The inclusive dijet limits are calculated using the W' signals from this analysis and the full analysis pipeline of that search; in particular, small-radius jets were used, so that the limits from that search get worse at higher m_B, m_C as the small-radius jets are not sufficient to contain all of the decay products of the daughter resonances. The limits from the inclusive dijet search are visually separated between signal models where $\langle \frac{2m}{p_T} \rangle < 0.4$ and the

dijet search is expected to be fully sensitive, and signal models where $\langle \frac{2m}{p_T} \rangle > 0.4$ and the dijet search loses sensitivity. The diboson search limits are computing using the Heavy Vector Triplet [406] W' signal used in that search. The acceptance for the W' using the selection in this search, compared to the W' acceptance in [20], is 86% and 54% at $m_{W'} = 3$ and 5 TeV, respectively. That search uses dedicated W jet taggers that make use of more features than just the mass of the jet, and so is expected to achieve sensitivity to those signals that would not be possible with the setup of this search (but would be possible in a future iteration of this analysis - Section 7.8). As expected, the targeted diboson search is more sensitive and the difference in limits is about a factor of 10.

For the SM dibosons, CWoLa hunting has the worst performance, as the NN is often unable to find the signal in the densest region of the background when the signal injected is less than that already excluded by other searches. For all other signals at $m_A = 3000$ GeV, CWoLa hunting sets new limits, with particularly large improvements of up to a factor of 10 when the masses of the daughter particles m_B, m_C are large, and up to a factor of 4 when m_B, m_C are smaller and the inclusive dijet search is more sensitive. For $m_A = 5000$ GeV, CWoLa hunting suffers from low statistics in the training, but is able to set new limits for large m_B, m_C, including signals models for which the inclusive dijet search is fully sensitive.

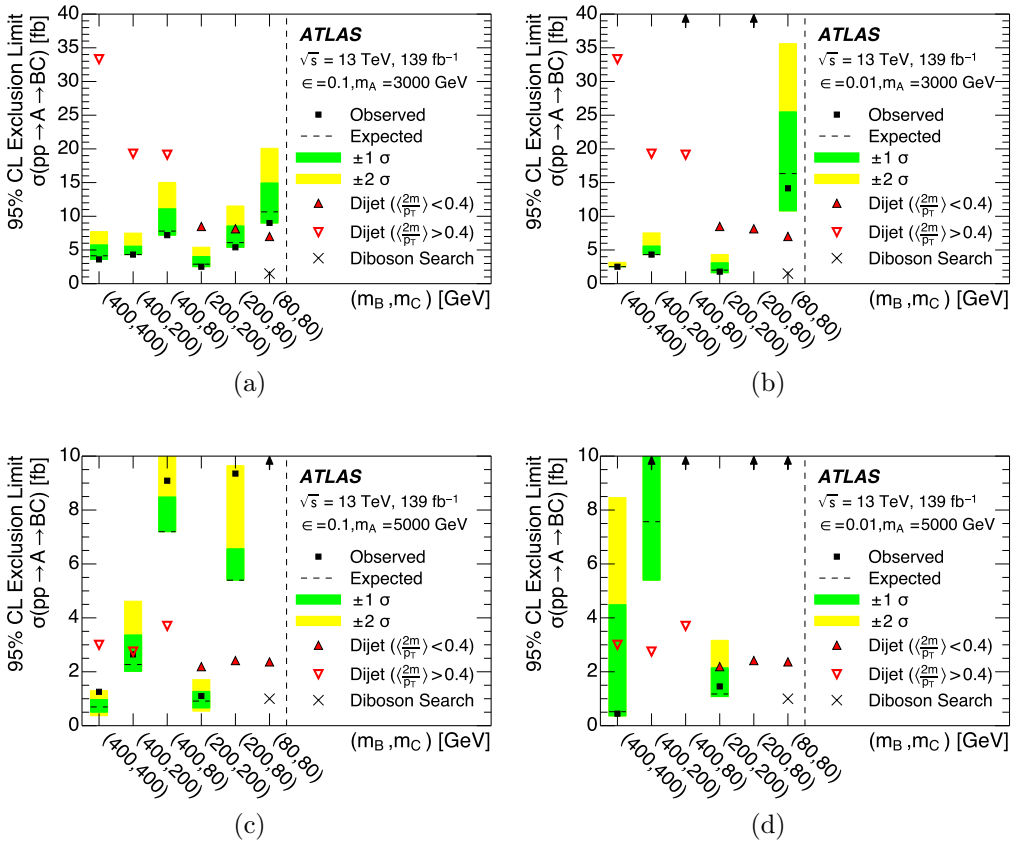


Figure 7.37: 95% confidence exclusion limits on the cross section of a variety of signal models, labeled by (m_A, m_B, m_C), in GeV. The limits are shown for signal models with (a,b) $m_A = 3000$ GeV and NN trained on signal region 5; and (c,d) $m_A = 5000$ GeV and NN trained on signal region 8. The limits are broken down between the analyses with (a,c) $\epsilon = 0.1$ and (b,d) $\epsilon = 0.01$. The limits from the ATLAS dijet search [19] are shown with red triangles. For signal models where $\langle \frac{2m}{p_T} \rangle < 0.4$ and the dijet search is expected to be fully sensitive, the markers are full and pointing up; for signals where $\langle \frac{2m}{p_T} \rangle > 0.4$ and the dijet search loses sensitivity, the markers are hollow and inverted. The limits from the ATLAS all-hadronic diboson search [20] are shown as well (black cross). Missing markers are higher than the plotted range.

7.8 Discussion

The search presented in this Chapter demonstrates for the first time the use of a generic signal and background model-independent method to search for unanticipated

signals directly in data. As such, considerable effort was put into the neural network training, especially the intricate cross-validation procedure (Section 7.4.4), which is necessary for being able to tag true signals without a high rate of false signals (Appendix C.1). However, some of the most challenging aspects of this analysis were the fitting procedure (Section 7.4.6) and the subsequent limit setting (Section 7.4.7.2). While the task of fitting an m_{JJ} spectrum with a parametric fit function over large numbers of events is one aspect that this analysis shares with almost every other dijet resonance search, the number of such fits required in this analysis is far greater than any existing analysis. Most dijet searches only do a few fits ultimately, especially if the background fit is decoupled from the POI scan (so that the POI scan fits have only a couple free parameters and converge quickly) like in this analysis. However, this analysis uniquely requires a huge number of background fits for the limit setting - one for each signal hypothesis, for each injected signal strength μ, for each NN selection ϵ - amounting to hundreds of background fits. The limit setting procedure used in this analysis is also novel, due to the unique situation where the event selection depends on the presence of the signal and therefore the likelihood function $\mathcal{L}(\mu, \theta)$ is extremely complicated. Therefore, the likelihood function can only be approximated by examining it at a few discrete points μ, rather than having an analytic dependence as is the usual case. Each of these challenges could be helped by simply increasing the computing time spent on this analysis. That being said, the computing resources used in this analysis are already considerable, taking days-weeks just to train all the neural networks. More details about the computing resources used in this analysis can be found in Appendix C.14.

In this iteration of the analysis, the features used were simply the masses of the two jets $\{m_1, m_2\}$. It is anticipated that this kind of 3-dimensional search in $\{m_A, m_B, m_C\}$ will turn into a standard result ATLAS publishes like the inclusive dijet search. For this search, the only handles on tagging the signal are the masses of the jets, and in particular rely heavily on an accurate calibration and good resolution of the large-R jet mass. The dijet resolution is also important for narrowing the m_{JJ} signal regions, which depends on an accurate calibration and good resolution of the large-R jet energy. The techniques presented in Chapter 8 and Chapter 9 focus mostly on calibrations

of small-R jets, but these techniques are intended as general calibration techniques which should apply readily to the calibration of large-R jets as well. In particular, the Author has contributed to a project intended to better calibrate the mass and energy of large-R jets [292] based directly on the novel technique presented in Chapter 9.

A natural extension of this analysis is to include more features of the two jets, especially the jet substructure which is known to tag jets resulting from top quarks and W/Z hadronic decays [94]. Indeed, the original proposal for this analysis [17] uses jet substructure information like N-subjettiness [202] to tag the jets and be much more sensitive to new signals that have non-trivial distributions in those features. Other ideas for using additional information about the jets are to include features which are typically used for heavy flavor tagging [182–184], di-b-tagging for Higgs identification [95], photon and electron identification [181, 281], hadronic τ-tagging [186, 302–304], or quark-gluon tagging [87]. Each of these sets of features would allow this analysis to be more sensitive to resonant decays to the given standard model particles, but also other more exotic jets produced from BSM particles. For example, using the entire jet image [90,407,408] to construct features with convolutional neural networks [315–317] could enable the low-mass extension of the search presented in Chapter 6, in which the a particles are each reconstructed as single small-R jets with exotic substructures due to the boosted gg and γγ decays.

A word of caution should be raised here about these exciting possibilities, though. The first thing to mention is that one advantage of using a reduced feature space is that calibrations and systematic uncertainties are available on all the included features (Section 7.4.5). This allows limits to be set on specific signal models while including all the known uncertainties between the simulation and the data. The use of any other jet substructure features may not come with these calibrations and uncertainties, meaning that the limit projections may be biased or that additional performance work will need to be done to enable the analysis[21]. Of course, one of

[21]Note that other more typical searches do use jet substructure features which are not themselves calibrated. In those searches the signal selection is set in advance, often by training in simulation, so that the overall classifier can be calibrated between data and simulation and the uncertainties understood. Since in this search the selection is not set in advance, this kind of approach would not work.

the main advantages of CWoLa hunting in the first place is that it does not require a specific signal model hypothesis and simply learns to tag whatever signal happens to be in data, if such a signal exists; i.e., the network itself does not need to be calibrated, and the only reason simulations are required are to understand what would have happened if a given signal were present. The second consideration that should go into adding additional features has to do with accounting for correlations between the features and m_{JJ}. This analysis spent considerable effort on imposing a decorrelation between the features and m_{JJ} (Section 7.4.3.1), and it is not clear exactly how this could be extended to arbitrary numbers and types of features. That being said, it is also true that of all the jet substructure features other than the jet p_T itself, the jet mass is probably the most correlated with the jet p_T and therefore with m_{JJ}, so that some other set of jet substructure features may not need to be decorrelated in this way.

There are even more speculative extensions using the principle of CWoLa beyond the dijet search involving more complicated topologies or even whole event information. The formulation of CWoLa and CWoLa hunting (Section 7.3.1) is quite general and any set of features satisfying the underlying principles should lend itself to the CWoLa hunting technique.

Finally, as discussed in Section 7.3.1, there are many competing alternative proposals to CWoLa hunting which can also be broadly sensitive to new unanticipated signal models in data.

7.9 Conclusion

This Chapter has presented a model agnostic search in the dijet final state using the full LHC Run 2 dataset recorded by ATLAS. Weakly supervised neural networks are trained entirely in data, without a specific signal model hypothesis, under the principle of classification without labels (CWoLa) in order to identify potential signals. The search is sensitive to generic final states of the form $A \to BC$, where all of A, B, C are massive and may be BSM, with $m_A \sim$ O(TeV) and $m_B, m_C \sim$ O(100 GeV). This is the first search to that completely covers the space of $\{m_A, m_B, m_C\}$, and in doing so

avoids the large trials factor associated with searching in this 3-dimensional feature space. No significant evidence for a localized excess is found in the dijet spectrum between 1.8 and 8.2 TeV. The search significantly improves the limits for certain specific signal models relative to the incusive dijet search, especially at high m_B, m_C. As the first analysis to use this new technique, a reduced feature set is used to tag the jets corresponding to the B and C particles, but there is great potential to extend this analysis to include additional features and therefore be more sensitive to specific final states. This analysis and these possible extensions enhance the coverage of the ATLAS search program and provide an essential means for ensuring ATLAS is sensitive to unanticipated scenarios.

Chapter 8

Numerical Inversion for Jet Calibrations

Numerical inversion is a general method used for calibrating the properties of objects observed in detectors and in particular is used for the calibration of the jets observed in the ATLAS detector. The key property of numerical inversion as a calibration method is that it is independent of the underlying spectrum. In this chapter this method is put into a formal framework and its statistical properties are explored rigorously. The results in this chapter are published in [349]. There are a few key novel results from this study. First, the method itself is established to be inherently biased, and the size of this bias is estimated analytically for the first time. Second, common approximations for the calibrated jet energy resolution are shown to be inaccurate, and more accurate estimations are presented in their place. Finally, extensions and corrections to numerical inversion are presented which can reduce the inherent bias. These approximations and corrections are shown to be increasingly important to consider as the LHC moves to higher instantaneous luminosities and pile-up conditions.

8.1 Introduction

Both of the main searches in this Thesis (Chapter 6 and Chapter 7) use jets extensively in their final states, and in general the theme of this Thesis is the use of jets in

ATLAS to search for new physics. As mentioned in their respective chapters, these searches would both benefit from improved kinematic resolution of the jets involved - in Chapter 6, the small-R jet energy resolution, and in Chapter 7, the large-R jet energy and mass resolution. The detector is not perfect, and so the observed quantities observed from the detector do not match exactly to the pre-detector, or particle-level, quantities; this can be confirmed either by using simulations of particles passing through the detector or by using calibration or test beam data where the pre-detector quantities are well-known. Because of this, all of these jet observables are calibrated from their detector-level quantities in order to both set the scale at the same level as at particle-level and to reduce the spread of the difference between the calibrated and the particle-level quantities [409] (Section 5.5.4.1). These calibration procedures at ATLAS [80, 81] and also at CMS [410, 411] involve several steps to correct for multiple nearly simultaneous pp collisions (pile-up), the non-linear detector response, the η-dependence of the jet response, flavor-dependence of the jet response, and residual data/simulation differences in the jet response. For all but the last step, simulations are used to determine the particle-level quantity. The final step, which is applied only to jets in data (*in-situ* correction), corrects for differences between the detector-level quantities in data and simulation by comparing the relative response to a well-measured reference object in the same event[1]. All of these calibrations, including the in-situ correction, proceed via a process known as *numerical inversion*.

While numerical inversion is essential for the jet calibration at ATLAS and CMS, the documentation is surprisingly sparse. The term was first introduced in a non-public ATLAS document [412] which amusingly is accompanied by the note "I asked that this note be refereed over 1 month ago and nobody has gotten back to me", dated August 2009. The motivation behind introducing this method is to ensure that the calibration is independent of the underlying spectrum used to derive the calibration. This is important at a general-purpose detector like ATLAS, because the energy spectra corresponding to the searches for different particles can be wildly

[1]This step occurs after the jets in simulation have already been corrected to be on the same scale as at particle-level. Since it is much more expensive to know particle-level quantities at data level (with calibration or test beam data), this relative ratio method is used to correct detector-level jets in data to the detector-level simulation jets, and therefore in turn to the particle-level.

different; the calibration should perform well regardless of what are the needs at the analysis level. This is further explained in Section 8.2.1.

The purpose of this chapter is to formally document numerical inversion and describe some of its properties. This chapter, and the paper in which these results are published, are intended to shed light on the properties of numerical inversion. In addition, the insight provided by these studies is used to propose corrections and extensions to numerical inversion. In particular, the deep understanding of the method provided by this work led directly to the techniques described in Chapter 9, which incorporate machine learning via neural networks into the numerical inversion process, and in doing so improve the jet energy calibration.

8.2 Calibrations

Numerical inversion is a method that can be used to calibrate any quantity that is observed in the detector, e.g. the jet transverse momentum p_T or the jet mass m. We focus here on the case of calibrating the E_T for sake of concreteness.

In what follows, X will be a random variable representing the particle-jet E_T and Y will be a random variable representing the reconstructed jet E_T [2] [3] . As mentioned in Section 8.1, there is a spread in Y given X, and furthermore Y is not equal to X even on average, due to various effects including the non-compensating and non-linear response of the ATLAS calorimeter, pile-up effects, and out-of-cone effects.

[2] Capital letters represent random variables and lower case letters represent realizations of those random variables, i.e. $X = x$ means the random variable X takes on the (non-random) value x. Functions that take random variables as arguments use square brackets $[\cdot]$ while functions that take real variables as arguments use curved brackets $(\cdot)$.

[3] As mentioned in Section 8.1, X and Y can be known simultaneously in simulation or using calibration or test beam data. For the in-situ correction, X^{data} and X^{MC} represent the p_T of the reference object in data and in simulation, while Y^{data} and Y^{MC} represent the p_T of the reconstructed jet, respectively. Then $c(x) \equiv R^{\text{data}}(x)/R^{\text{MC}}(x)$ is defined and Y^{data} is corrected using $c(x)$ as the analogue of the response function $R(x)$ in numerical inversion as described in Section 8.2.2. In order to prevent confusion and remain concrete, the simpler case where Y is being corrected to X directly is focused on.

Define

$$f_{me}(x) \equiv \mathbb{E}[Y|X = x] \tag{8.1}$$

$$R_{me}(x) \equiv \mathbb{E}\left[\frac{Y}{x}\middle|X = x\right] = \frac{f_{me}(x)}{x}. \tag{8.2}$$

Where $\mathbb{E}[\cdot]$ is defined as taking the expected value (mean) of its argument, and the subscript indicates that the mean is being used to define f and R. In practice, sometimes the core of the distribution of $Y|X = x$ is fit with a Gaussian and so the effective measure of central tendency is the mode of the distribution. Therefore in analogy to Equations 8.1 and 8.2, we define

$$f_{mo}(x) \equiv \text{mode}[Y|X = x] \tag{8.3}$$

$$R_{mo}(x) \equiv \text{mode}\left[\frac{Y}{x}\middle|X = x\right] = \frac{f_{mo}(x)}{x}. \tag{8.4}$$

Similarly, mode$[\cdot]$ is defined as taking the mode of its argument, and the subscript indicates the definition of f and R. We will often drop the subscript of f and R for brevity in the text, when it is clear which definition we are referring to. If not specified, f and R will refer to a definition using a generic definition of central tendency. For all sensible notions of central tendency, we still have that $R(x) = \frac{f(x)}{x}$. The function $R(x)$ is called the *response function*.

The *resolution* $\sigma(x)$ measures the spread of Y given X and is defined to be

$$\sigma(x)^2 \equiv \sigma[Y|X = x]^2 = \mathbb{E}[Y^2|X = x] - \mathbb{E}[Y|X = x]^2 \tag{8.5}$$

where $\sigma[\cdot]$ is defined as taking the standard deviation of its argument.

The resolution is sometimes broken up into terms based on the dependence on x [342, 413][4]:

$$\sigma(x) \sim N \oplus S \cdot \sqrt{x} \oplus C \cdot x \tag{8.6}$$

[4] Actually, this breakdown is usually applied to the resolution after calibration, but the intuition is the same for the uncalibrated resolution.

where N is the *noise term*, S is the *stochastic term*, and C is the *constant term*. The noise term corresponds to sources of variance independent of the jet energy; for example, pile-up is a source of noise independent of the true jet energy[5]. The stochastic term corresponds to Poisson sources of noise; in particular, the number of particles produced by a jet passing the noise thresholds is roughly Poisson distributed with mean scaling with x. The constant term corresponds to sources of noise that are a constant fraction of x, e.g. energy lost in passive material in the detector. As the scalings suggest, the N term dominates for lower energy jets, the C term dominates for higher energy jets, and the S term dominates in the intermediate regions; however, in practice it is difficult to unequivocally disentangle their contributions. The results presented in this Chapter are general and do not depend on this breakdown. However, the toy model used to present results (Appendix D.9) focuses on lower energy jets and high pile-up conditions, as might be expected at the high-luminosity LHC (Section 4.4), where the N term dominates, with $\sigma'(x) = 0$.

It is often useful to think of $Y|X = x \sim \mathcal{N}(f(x), \sigma(x))$; however, the discussion in this Chapter remains general unless stated otherwise. Just to be clear, in this case we have that $\mathbb{E}[Y|X = x] = \text{mode}[Y|X = x] = f(x)$, and that $\sigma[Y|X = x] = \sigma(x)$, so the definitions are consistent.

The functions f and R should be thought of as parameterizing the average response of the detector to jets with known energies X, and σ as parameterizing the spread of the response. In particular, f, R, and σ are independent of the underlying spectrum of X - if two samples are examined, one with more jets produced at a lower energy, for example, the values of f, R, and σ derived from these two samples should be the same (to within statistical fluctuations).

In the calibration we are given a list of ordered pairs (X, Y), from which we can derive f and R, and we need to correct Y to be on the same scale as X when averaged over the jets in the sample. The goal of the calibration therefore is to devise a calibration function $c : Y \mapsto Z = c(Y)$ such that after the calibration, the scale of Z is the same as the scale of X. There are two ways one might judge that.

[5]Since the detector response is non-linear, this is only true to first order.

One way is to define the *closure*

$$\mathsf{C}_{\mathrm{me}}(\mathsf{x}) \equiv \mathbb{E}\left[\frac{\mathsf{Z}}{\mathsf{x}}\middle|\mathsf{X}=\mathsf{x}\right] = \frac{\mathbb{E}\left[\mathsf{Z}|\mathsf{X}=\mathsf{x}\right]}{\mathsf{x}}, \tag{8.7}$$

and another way is to define the *prior-dependent closure*

$$\mathsf{C}^{\mathsf{p}}_{\mathrm{me}}(z) \equiv \mathbb{E}\left[\frac{\mathsf{X}}{z}\middle|\mathsf{Z}=z\right] = \frac{\mathbb{E}\left[\mathsf{X}|\mathsf{Z}=z\right]}{z}, \tag{8.8}$$

with C_{mo} and $\mathsf{C}^{\mathsf{p}}_{\mathrm{mo}}(z)$ defined analogously. The symbols C and C^{p} will denote the closure for a generic notion of central tendency.

As the name suggests, the prior-dependent closure depends on the prior distribution of X in the sample. A sample with more jets produced at lower energies, for example, will tend to be biased towards a lower prior-dependent closure than another sample with more jets produced at higher energies. As mentioned in the motivation, at ATLAS there are samples targeting different physics regimes with wildly different prior distributions of true energies, so the prior-dependent closure cannot be relied on to be useful in physics analysis.

On the other hand, the closure C is independent of the prior distribution of X, given c. Therefore, this is the agreed-upon figure of merit for the calibration. We say that the calibration has *achieved closure* or simply *closes* if, for all x,

$$\mathsf{C}(\mathsf{x}) = 1. \tag{8.9}$$

8.2.1 Learning the Prior

One way to define c is to learn a function $\mathsf{M}(\mathsf{Y})$ that predicts X given Y directly. I.e., define

$$\mathsf{M}_{\mathrm{me}}(\mathsf{y}) \equiv \mathbb{E}[\mathsf{X}|\mathsf{Y}=\mathsf{y}] \tag{8.10}$$

and then let $\mathsf{c} = \mathsf{M}$ so that $\mathsf{Z} = \mathsf{M}(\mathsf{Y})$.

This method is *not* the way calibrations are done in ATLAS, and the reasoning

for this decision is given below.

It's clear that this definition of calibration function depends on the prior distribution of X in the sample, and in particular the prior-dependent closure (Equation 8.8) is achieved for the sample used to learn M.

That being said, c is just some function, even if its definition is prior-dependent. One might think that one could define c on a representative sample, e.g. with a uniform prior, and then apply the calibration function across various samples, achieving closure (Equation 8.7).

However, even if the underlying distribution of X is uniform across some range, this method does not necessarily lead to closure. In particular, if $f(x)$ is nonlinear or if the resolution $\sigma(x)$ is nonconstant, then even with a uniform prior distribution of X there can be large non-closures. Since both of these effects are present in the ATLAS jet calibration [81], it is expected that these nonclosures will be present if this method were attempted in the ATLAS jet calibration. This effect can be seen in Fig. 8.1, using a toy model that approximates the ATLAS nonlinear $f(x)$ (Appendix D.9) with a constant resolution and another toy model with a linear f but with a nonconstant resolution $\sigma(x)$ that approximates the resolution in ATLAS.

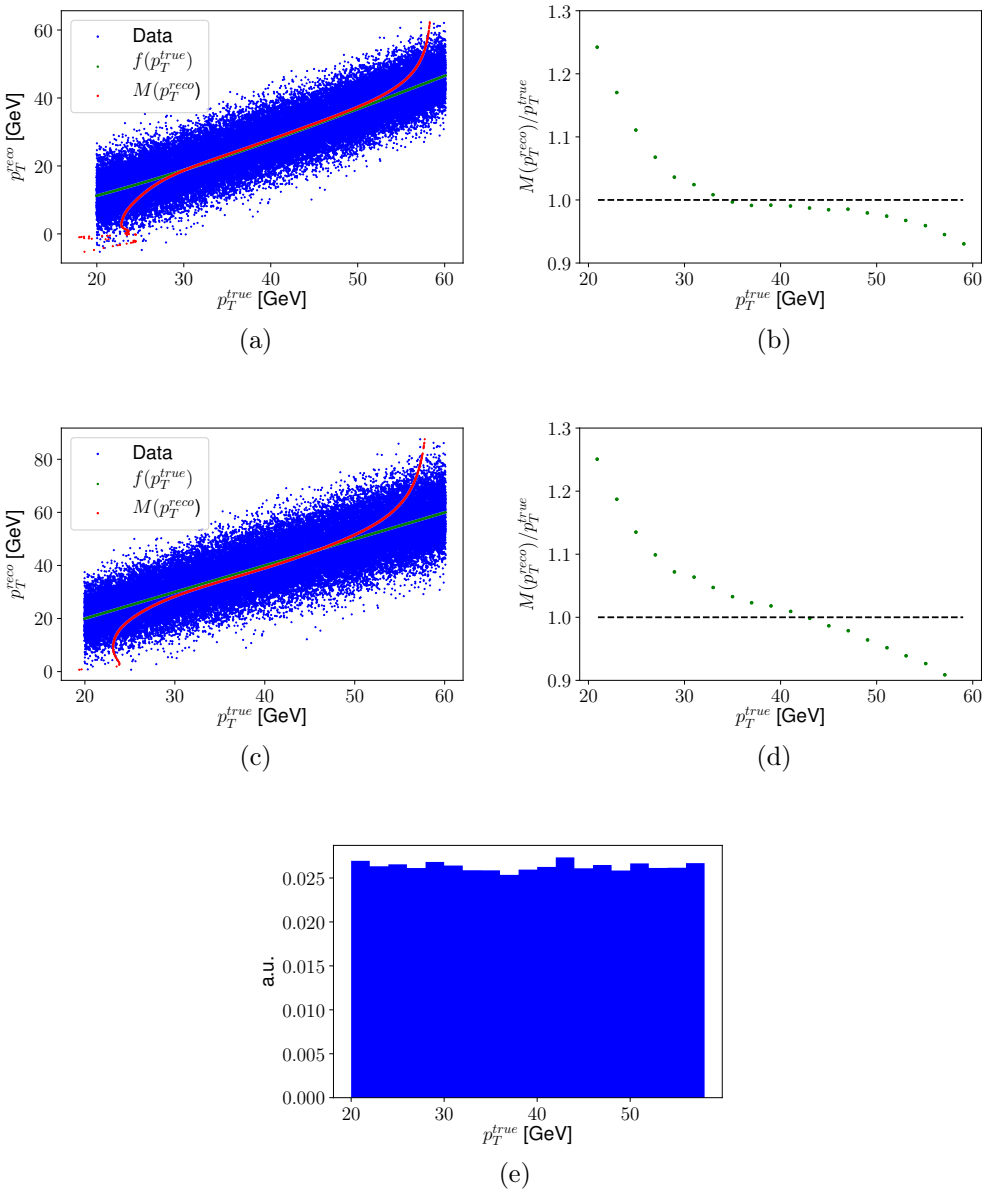


Figure 8.1: The effects of learning a function $\mathsf{M}(\mathsf{Y})$ that predicts X directly given Y, with a uniform underlying distribution of X (e). (a,b) A nonlinear $\mathsf{f}(\mathsf{x})$ intended to model the ATLAS nonlinear response function, with a constant resolution. (c,d) A nonconstant $\sigma(\mathsf{x})$ intended to model the ATLAS nonconstant resolution, with a linear $\mathsf{f}(\mathsf{x})$. In this figure only X and Y correspond to the truth and reconstructed jet p_T, respectively, rather than the E_T, as will be the case for the remainder of this chapter; however, the principle is exactly the same.

8.2.2 Numerical Inversion

Instead, the ATLAS calibration uses numerical inversion. Formally, numerical inversion is the following procedure:

1. Compute $f(x)$, $R(x)$.
2. Let $\tilde{R}(y) = R(f^{-1}(y))$.
3. Apply a jet-by-jet correction: $Y \mapsto Y/\tilde{R}(Y)$.

The intuition for step 2 is that for a given value y drawn from the distribution $Y|X = x$, $f^{-1}(y)$ is an estimate for x and then $R(f^{-1}(y))$ is an estimate for the response at the value of x that gives rise to Y. Let $\rho_X(x)$ be the prior probability density function of $X = E_T$. Since f is defined conditioning on a given value of X, f is by construction independent of $\rho_X(x)$. Therefore, this calibration is independent of $\rho_X(x)$, since f and thus f^{-1} do not depend on $\rho_X(x)$. This is what is meant when it is said that numerical inversion is independent of the underlying spectrum.

The question of whether numerical inversion achieves closure will be discussed in detail in the following sections; in particular, under certain conditions it is found that numerical inversion does *not* achieve closure. This raises the question of whether using f^{-1} instead of M is indeed a better method of calibration, under the criterion of achieving closure in a prior-independent way. The first thing to note is that at least the definition of the calibration function is independent of the prior distribution of X, unlike the definition of M. Second, it will be found that the non-closures can be approximated and understood in terms of properties of the distributions of $Y|X$, whereas the non-closures present when using M to calibrate would depend necessarily on the prior distribution $\rho_X(x)$ used to learn M. Third, this is the method used in ATLAS and CMS for jet calibrations, and therefore the method deserves to be studied and its shortcomings understood. Finally, later in this chapter (Section 8.4.4) a calibration function will be proposed different from both f^{-1} and M which attempts to achieve better closure than either in a prior-independent way.

We can see now the first result, which will be useful for the rest of this chapter:

The correction derived from numerical inversion is $Y \mapsto Z = f^{-1}(Y)$.

Proof.

$$\begin{aligned} \tilde{R}(Y) &= R(f^{-1}(Y)) \\ &= \frac{f(f^{-1}(Y))}{f^{-1}(Y)} \\ &= \frac{Y}{f^{-1}(Y)} \\ \rightarrow Z &= \frac{Y}{\tilde{R}(Y)} \\ &= f^{-1}(Y) \qquad \square \end{aligned} \tag{8.11}$$

I.e., the calibration function is simply $c = f^{-1}$.

8.2.3 Assumptions and Definitions

The general results presented in the following sections are based on three assumptions listed below. These requirements should be satisfied by real detectors using calorimeters and trackers to reconstruct jets, given that the detector-level reconstruction is of sufficiently high quality.

1. $f^{-1}(y)$ exists for all y in the support of Y, and f^{-1} is single-valued. These may seem like obvious statements, but are not vacuous, even for a real detector. For example, pile-up corrections can result in non-zero probability that $Y < 0$, so the function f must be computed for all possible values of Y, even if the transverse energy is negative. At the high-luminosity LHC (Section 4.4), the level of pile-up will be so high that the jet energy resolution may be effectively infinite at low transverse energies (no correlation between particle-level and detector-level jet energy). In that case, f^{-1} may not be single valued and numerical inversion cannot be strictly applied as described in Section 8.2.2.

2. $f(x)$ is monotonically increasing: $f'(x) > 0$ for all x. This condition should trivially hold: detector-level jets resulting from particle-level jets with a higher

E_T should on average have a higher E_T than those originating from a lower E_T particle-level jet[6]. We note also that Assumption 1 implies that $f'(x) \geq 0$ or $f'(x) \leq 0$ for all x; so Assumption 2 is equivalent to the additional assumptions that $f'(x) \neq 0$ for any x, and that $f'(x) > 0$ (as opposed to $f'(x) < 0$).

3. f is twice-differentiable. The first derivative of f has already been assumed to exist in Assumption 2, and the second derivative will also be required to exist for some of the later results. In practice it is expected that f is differentiable out to any desired order.

We note that as long as the above three assumptions hold, the theorems stated in the remainder of this chapter are valid. In particular, this implies that x could be any calibrated quantity that satisfies the above constraints. Assumption 1 is necessary for numerical inversion to be performed, while Assumptions 2 and 3 are necessary for deriving the analytical results given in this Chapter.

The results in this chapter have been divided into "Proofs" and "Derivations". The "Proofs" require only the three assumptions stated above, and in particular do not assume anything about the shape of the underlying distributions, e.g. that the distributions $Y|X = x$ are Gaussian or approximately Gaussian. The "Derivations" are useful approximations that apply in the toy model described in Appendix D.9; we expect them to apply in a wide variety of cases relevant to LHC jet physics. In particular, it is expected that these approximations hold in cases with properties similar to the toy model presented here - e.g., good approximation of f by its truncated Taylor series about each point and approximately Gaussian underlying distributions of $Y|X = x$. [7] Some of the derivations are illustrative and are included in the main

[6]Note that this is only true for fixed η. Detector technologies can depend significantly on η (Section 5.2) and the η-dependence of f (for a fixed x) can be non-monotonic. When calibrating η itself, though, it should still be the case that as the truth-level η increases the detector-level η also increases, satisfying this Assumption.

[7]Note that we do not require that $Y|X = x$ is exactly Gaussian, only that it is approximately Gaussian, which is true for a wide range of energies and jet reconstruction algorithms at ATLAS and CMS. In particular, there are non-negligible (but still often small) asymmetries at low and high E_T at ATLAS and CMS [81,410,411]. In any case, even if $Y|X = x$ *is* Gaussian, $Z|X = x$ is in general *not* Gaussian, for non-linear response functions (Appendix D.1).

body of the text; however, some derivations are moved to the Appendix (D) in order to make the text more readable and highlight the main results.

Finally, in the rest of this chapter, we write $\rho_P(p)$ to represent the probability distribution of a random variable P, so that $\rho_{Y|X}(y|x)$ represents the probability distribution of Y given $X = x$, and $\rho_{Z|X}(z|x)$ is the probability distribution of Z given $X = x$. A standard fact about the probability distribution from changing variables is that

$$\rho_{Z|X}(z|x) = f'(z)\rho_{Y|X}(f(z)|x). \tag{8.12}$$

To ease the notation, we will often use $\rho_Y(y)$ and $\rho_Z(z)$ interchangeably with $\rho_{Y|X}(y|x)$ and $\rho_{Z|X}(z|x)$, respectively, when it is clear (as is usually the case) that we are conditioning on some true value x.[8]

8.3 Results

In the subsequent sections, we will derive properties about the closure C for three different definitions of the central tendency: mean (Section 8.3.1), mode (Section 8.3.2), and median (Section 8.3.3).

8.3.1 Mean

In the following section only, for brevity, we will let f be f_{me} and C be C_{me}.

8.3.1.1 Closure

We can write the closure (Equation 8.7) as

[8] In practice with a real sample it is necessary to condition on a small range of X, e.g. $X \in [x, (1+\epsilon)x]$. If ϵ is large then there can be complications from the changing of $f(x)$ over the specified range and from the shape of the prior distribution of X over the specified range. These challenges can be solved by generating large enough simulation or real datasets. We therefore assume that $\epsilon \ll 1$ and consider complications from finite ϵ beyond the scope of this Thesis.

$$C(x) = \mathbb{E}\left[\frac{Z}{x}\middle| X = x\right] = \frac{1}{x}\int dy \rho_{Y|X}(y|x) f^{-1}(y). \tag{8.13}$$

We find that for many functions f, numerical inversion does not close. This is summarized in the following result:

Let the notion of central tendency be the mean. If f is linear, then numerical inversion closes. If f is not linear, then numerical inversion does not necessarily close.

Proof. Let f be linear, $f(x) = a(x + b)$. Then[9] $f^{-1}(y) = \frac{y}{a} - b$. We can see that we necessarily have closure as Equation 8.13 can be written

$$\begin{aligned} C(x) &= \frac{1}{x}\int dy \rho_{Y|X}(y|x)\left(\frac{y}{a} - b\right) \\ &= \frac{1}{x}\left(\frac{1}{a}\mathbb{E}\left[Y|X = x\right] - b\right) \\ &= \frac{1}{x}\left(\frac{1}{a}f(x) - b\right) \\ &= 1. \end{aligned} \tag{8.14}$$

Now let f be nonlinear, and so therefore f^{-1} is also nonlinear. We note that the statement being proved is that f does not necessarily close in this case; not that f necessarily does not close. Thus, it is sufficient to find a counterexample that does not close in order to demonstrate this statement. Let $f(x) = \left(\frac{x}{c}\right)^{\frac{1}{3}}$ with $c \neq 0$, so that $f^{-1}(y) = cy^3$, which is a simple non-linear monotonic function. We will also need to specify some higher moments of the distribution $\rho_{Y|X}$. With the standard definitions of the variance and skew, respectively:

$$\sigma(x)^2 \equiv \mathbb{E}\left[(Y - \mathbb{E}\left[Y\right])^2\middle| X = x\right] \tag{8.15}$$

$$\sigma(x)^3\gamma_1(x) \equiv \mathbb{E}\left[(Y - \mathbb{E}\left[Y\right])^3\middle| X = x\right]. \tag{8.16}$$

[9] We have $a > 0$ from the assumption that $f'(x) > 0$.

We specify the weak conditions that $\sigma(x) > 0$ (which is always true as long as $\rho_{Y|X}$ is not a delta function), and that $\gamma_1(x) = 0$ (which is true if $\rho_{Y|X}$ is symmetric). Then, the closure (Equation 8.13) can be written

$$\begin{aligned} C(x) &= \frac{1}{x}\int dy \rho_{Y|X}(y|x)\left(cy^3\right) \\ &= \frac{c}{x}\left(\mathbb{E}\left[Y^3|X=x\right]\right). \end{aligned} \tag{8.17}$$

With $\gamma_1(x) = 0$, we have that

$$\begin{aligned} \mathbb{E}\left[Y^3|X=x\right] &= 3\sigma(x)^2\mathbb{E}\left[Y|X=x\right] + \mathbb{E}\left[Y|X=x\right]^3 \\ &= 3\sigma(x)^2 f(x) + f(x)^3 \\ &= 3\sigma(x)^2\left(\frac{x}{c}\right)^{\frac{1}{3}} + \frac{x}{c}. \end{aligned} \tag{8.18}$$

Then we see we do not have closure, as

$$\begin{aligned} C(x) &= \frac{c}{x}\left(\mathbb{E}\left[Y^3|X=x\right]\right) \\ &= \frac{c}{x}\left(3\sigma(x)^2\left(\frac{x}{c}\right)^{\frac{1}{3}} + \frac{x}{c}\right) \\ &= 1 + 3\sigma(x)^2\left(\frac{x}{c}\right)^{-\frac{2}{3}} \\ &> 1. \qquad \square \end{aligned} \tag{8.19}$$

Although the counterexample provided here only applies to a specific choice of $f(x)$ and $\rho_{Y|X}(y|x)$, we have reason to believe that closure is not achieved for non-linear f in the vast majority of cases, as can be seen in more detail in Appendix D.2. In addition, we can Taylor expand the closure C to derive an equation for the first non-closure term:

$$C(x) \approx 1 - \frac{1}{2}\frac{f''(x)}{f'(x)^3}\frac{\sigma(x)^2}{x}, \tag{8.20}$$

the derivation of which can be found in Appendix D.2.

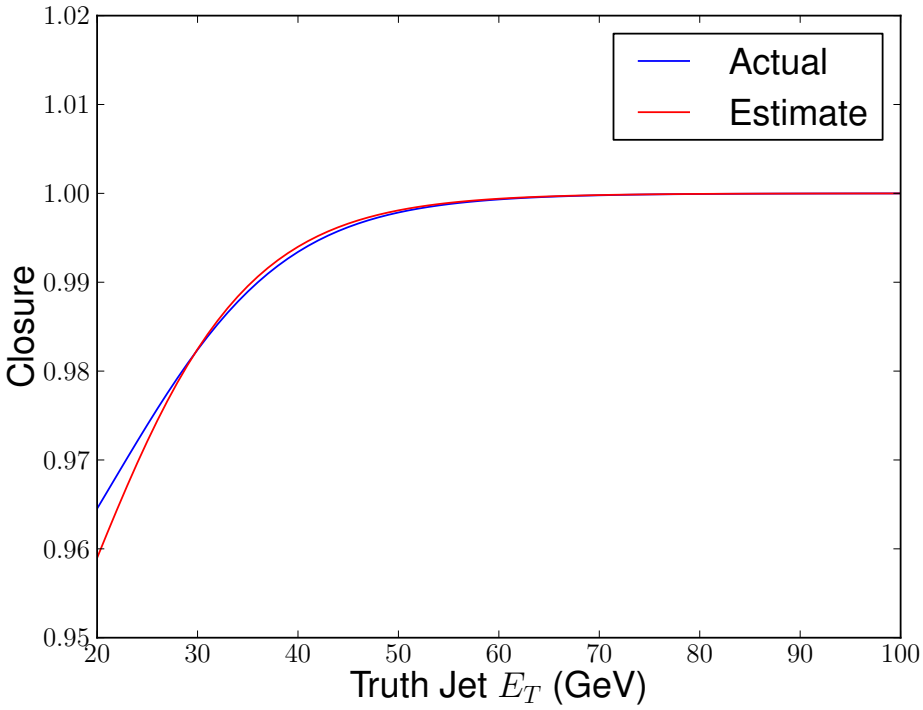


Figure 8.2: The closure of numerical inversion when using the mean to calibrate, using a toy model similar to conditions in ATLAS or CMS. In blue, the exact calculated closure. In red, the estimate of the closure using the first term of the higher-order correction given in Equation 8.20. For details of the model, see Appendix D.9.

Figure 8.2 shows the inherent non-closure in numerical inversion for a toy calculation using a response function $R(x)$ that is typical for ATLAS or CMS, and the first term of the higher-order correction (Equation 8.20).

8.3.1.2 Calibrated Resolution

We often care about how precise the reconstructed transverse energy of the jets is relative to the detector-level quantity, which is quantified by examining the standard deviation of the calibrated resolution Z.

The calibrated resolution of the reconstructed jets is defined to be the standard

deviation of the Z distribution, with $X = x$, which is given by

$$\hat{\sigma}(x)^2 \equiv \sigma\left[Z|X=x\right]^2 = \mathbb{E}\left[Z^2\middle|X=x\right] - \mathbb{E}\left[Z|X=x\right]^2, \tag{8.21}$$

and the fractional resolution is just given by $\sigma\left[\frac{Z}{x}|X=x\right]$. The fractional resolution, to first order in the Taylor series, is given by

$$\sigma\left[\frac{Z}{x}|X=x\right] = \frac{1}{x}\hat{\sigma}(x) \approx \frac{1}{x}\frac{\sigma(x)}{f'(x)}, \tag{8.22}$$

the derivation of which can be found in Appendix D.3.

A standard diagnostic technique when a full calibration is not applied is to examine the quantity $\frac{\sigma(x)}{R(x)}$. This diagnostic is also often used [174, 187, 413] when there are some remaining non-closures after the calibration (which is expected, as shown above) by examining $\frac{\hat{\sigma}(x)}{C(X)}$ as a proxy for the case where there are no non-closures. In fact technically there is no distinction between these two cases, as after the calibration one could imagine defining a new response $R_{new}(x) = C(x)$ and new resolution $\sigma_{new}(x) = \hat{\sigma}(x)$ and performing numerical inversion a second time in order to remove the remaining non-closures (Section 8.4.3).

It should be noted that $f'(x)$ is *not* the response $R(x) = \frac{f(x)}{x}$. In particular, $f'(x) = R(x) + R'(x)x$, so $f'(x) \neq R(x)$ unless $R'(x) = 0$, or equivalently $f(x) = kx$ for some constant k (which is not the case at ATLAS nor at CMS). Figure 8.3 verifies Equation 8.22 and compares it to the method of dividing the width of the distribution by R [10] .

[10] Taking $\frac{\sigma(x)}{R(x)}$ to estimate the calibrated resolution, and taking $\frac{\hat{\sigma}(x)}{C(x)}$ to estimate the resolution with no non-closures, are technically equivalent. $C(x)$ is close to 1, while $R(x)$ can be very different from 1, so the difference between $f'(x)$ and $R(x)$ is larger than the difference between $(C(x)x)'$ and $C(x)$. Figure 8.3 emphasizes the difference between the two estimates in the former case.

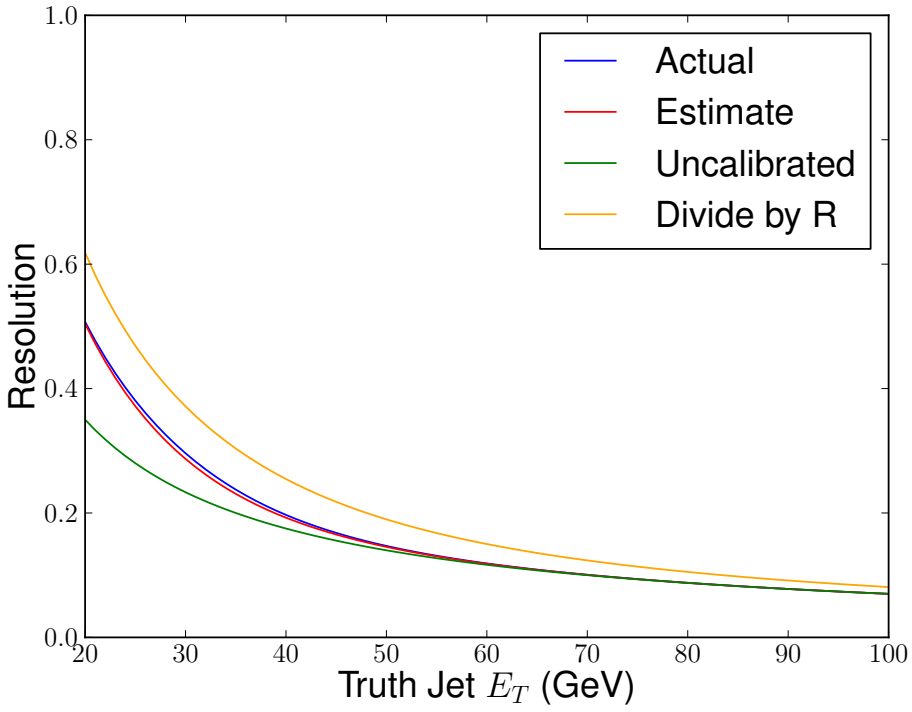


Figure 8.3: The resolution of the E_T distribution following numerical inversion when using the mean to calibrate, using a toy model similar to conditions in ATLAS or CMS. In blue, the exact calculated resolution. In red, the estimate of the resolution using the first term of the higher-order correction in Equation 8.22. In green, the uncalibrated resolution. In orange, the resolution when dividing by the response $R(x)$. For details of the model, see Appendix D.9.

8.3.2 Mode

In the following section only, for brevity, we will let f be f_{mo} and C be C_{mo}. The distribution $\rho_{Y|X}(y|x)$ is usually unimodal and Gaussian fits to the "core" of this distribution are essentially picking out the mode of the distribution. Therefore, the results of this section are a good approximation to what is often used in practice. We note that in the case that the underlying distribution is multimodal, it is not clear how to unambiguously define the mode of the distribution, and so the results of this section cannot be applied naively.

8.3.2.1 Closure

Assuming that the probability distribution function of a random variable P is unimodal, the mode is the point at which the first derivative of the probability distribution function is 0:

$$\text{mode}[P] \equiv p^* \text{ s.t. } \rho_p'(p^*) = 0. \tag{8.23}$$

Then we can write the closure condition (Equation 8.7) as

$$\begin{aligned} \text{mode}\left[\frac{Z}{x}\middle| X = x\right] &= 1 \\ \rightarrow \text{mode}\left[Z|X = x\right] &= x \\ \rightarrow \rho_Z'(x) &= 0. \end{aligned} \tag{8.24}$$

Using this definition, we can prove a result similar (but stronger) to the closure condition for the mean in the previous section:

Let the notion of central tendency be the mode. Numerical inversion closes if and only if f is linear.

Proof. We have from Equation 8.12 that

$$\rho_Z(z) = f'(z)\rho_Y(f(z)). \tag{8.25}$$

Therefore,

$$\rho'_Z(z) = f''(z)\rho_Y(f(z)) + f'(z)^2\rho'_Y(f(z)), \tag{8.26}$$

and

$$\begin{aligned}\rho'_Z(x) &= f''(x)\rho_Y(f(x)) + f'(x)^2\rho'_Y(f(x)) \\ &= f''(x)\rho_Y(y^*) + f'(x)^2\rho'_Y(y^*) \\ &= f''(x)\rho_Y(y^*),\end{aligned} \tag{8.27}$$

where $\rho_Y(y^*) > 0$ since y^* is the mode of the distribution ρ_Y. Then we see that if $f''(x) = 0$, then $\rho'_Z(x) = 0$ and closure is achieved. In contrast, if $f''(x) \neq 0$, then $\rho'_Z(x) \neq 0$ and closure is not achieved. □

The closure when using the mode to calibrate, to first order in the Taylor series, is given by

$$C(x) \approx 1 + \frac{f''(x)}{f'(x)^3}\frac{\tilde{\sigma}(x)^2}{x}, \tag{8.28}$$

where $\tilde{\sigma}(x)$ is the width of a Gaussian fitted to just the area near the peak of the function $\rho_{Y|X}(y|x)$ (defined precisely in the next section). The derivation of Equation 8.28 can be found in Appendix D.4.

Figure 8.4 shows the inherent non-closure in numerical inversion for a toy calculation using a response function $R(x)$ that is typical for ATLAS or CMS, and the first term of the higher-order correction given in Equation 8.20, when using the mode for calibration. This non-closure agrees with ATLAS published results, e.g. [413] (Figure 5).

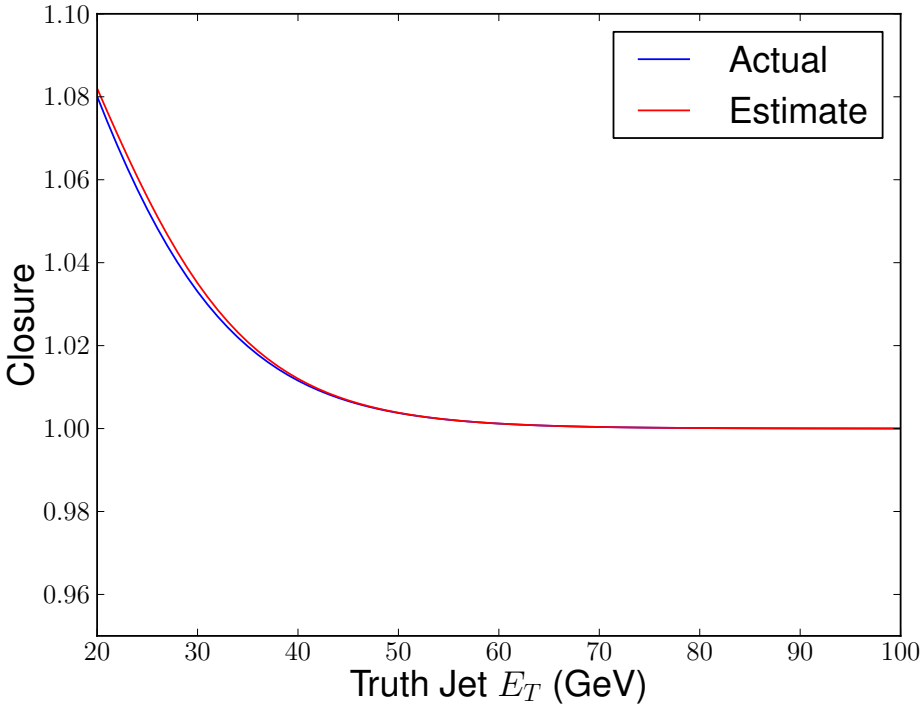


Figure 8.4: The closure of numerical inversion when using the mode to calibrate, using a toy model similar to conditions in ATLAS or CMS. In blue, the exact calculated closure. In red, the estimate of the closure using the first term of the higher-order correction given in Equation 8.28. For details of the model, see Appendix D.9.

8.3.2.2 Resolution

Let $z^*(x)$ be the mode of the distribution $Z|X = x$, which is not necessarily equal to x given the above result. It is often the case at ATLAS and CMS that a Gaussian is fit to the distributions $\rho_{Y|X}(y|x)$ and $\rho_{Z|X}(z|x)$ only in the vicinity of the modes $f(x)$ and $z^*(x)$, respectively, since it is assumed that the distributions have a Gaussian core but non-Gaussian tails. The width of the Gaussian core found in this fit is then used as a measure of the resolution of the distribution. We thus define a "trimmed resolution" for a distribution P about its mode m, which is valid if $P \sim \mathcal{N}(m, \tilde{\sigma})$ for p near m:

$$\tilde{\sigma}[P]^2 \equiv -\frac{\rho_P(m)}{\rho_P''(m)}. \tag{8.29}$$

The definition in Equation 8.29 is chosen because it reduces to the usual variance for a Gaussian distribution. For the distributions $\rho_{Y|X}(y|x)$ and $\rho_{Z|X}(z|x)$, we thus have the trimmed resolutions

$$\tilde{\sigma}(x)^2 \equiv \tilde{\sigma}\left[Y|X = x\right]^2 = -\frac{\rho_Y(f(x))}{\rho_Y''(f(x))} \tag{8.30}$$

$$\hat{\tilde{\sigma}}(x)^2 \equiv \tilde{\sigma}\left[Z|X = x\right]^2 = -\frac{\rho_Z(z^*(x))}{\rho_Z''(z^*(x))}. \tag{8.31}$$

The calibrated fractional trimmed resolution $\tilde{\sigma}\left[\frac{Z}{x}|X = x\right]$, to first order in the Taylor series, is given by

$$\tilde{\sigma}\left[\frac{Z}{x}|X = x\right] = \frac{1}{x}\hat{\tilde{\sigma}}(x) \approx \frac{1}{x}\frac{\tilde{\sigma}(x)}{f'(x)}, \tag{8.32}$$

the derivation of which can be found in Appendix D.5.

8.3.3 Median

In the previous sections we have examined using the mean or the mode to define f and C, and found that both results do not lead to closure in general. We propose a

new definition, using the median of the reconstructed jet E_T distributions:

$$f_{med}(x) = \text{median}[Y|X = x] \tag{8.33}$$

$$R_{med}(x) = \text{median}\left[\frac{Y}{x}\middle|X = x\right] = \frac{f_{med}(x)}{x}. \tag{8.34}$$

And define C_{med} analogously. In the following section only, for brevity, we will let f be f_{med} and C be C_{med}.

8.3.3.1 Closure

The median of the distribution of a random variable P is the point at which 50% of the distribution is above and 50% is below:

$$\text{median}[P] \equiv p^* \text{ s.t. } \int_{-\infty}^{p^*} \rho_P(p)dp = 0.5. \tag{8.35}$$

Then the closure condition (Equation 8.7) can be written

$$\begin{aligned} &\text{median}\left[\frac{Z}{x}\middle|X = x\right] = 1 \\ \rightarrow\ &\text{median}\left[Z|X = x\right] = x \\ \rightarrow\ &\int_{-\infty}^{x} \rho_Z(z)dz = 0.5. \end{aligned} \tag{8.36}$$

We can see then the following result under this definition of central tendency:

Let the notion of central tendency be the median. Then numerical inversion always closes.

Proof. We have from Equation 8.12 that

$$\rho_Z(z) = f'(z)\rho_Y(f(z)). \tag{8.37}$$

So the closure condition in Equation 8.36 becomes

$$\begin{aligned} 0.5 &= \int_{-\infty}^{x} \rho_Z(z)dz \\ &= \int_{-\infty}^{x} f'(z)\rho_Y(f(z))dz. \end{aligned} \tag{8.38}$$

Then with $u = f(z), du = f'(z)dz$ we have

$$\begin{aligned} 0.5 &= \int_{-\infty}^{f(x)} \rho_Y(u)du \\ &= \int_{-\infty}^{y^*} \rho_Y(u)du \\ &= 0.5. \qquad \square \end{aligned} \tag{8.39}$$

8.3.3.2 Resolution

A natural definition of resolution when using the median to calibrate jets is the 68% interquantile range, defined as follows for a random variable P:
With I_P^- and I_P^+ defined by

$$\int_{-\infty}^{I_P^-} \rho_P(p)dp \equiv \Phi(-1), \tag{8.40}$$

$$\int_{-\infty}^{I_P^+} \rho_P(p)dp \equiv \Phi(+1); \tag{8.41}$$

the 68% interquantile range is defined as

$$\sigma_{IQR}[P] \equiv \frac{1}{2}\left(I_P^+ - I_P^-\right). \tag{8.42}$$

Where $\Phi(x) = \frac{1}{2}\text{erfc}\left(\frac{-x}{\sqrt{2}}\right)$ is the cumulative distribution function of the normal distribution. The definition is designed so that if $P \sim \mathcal{N}(\mu, \sigma)$ then $\sigma_{IQR}[P] = \sigma$. The quantity σ_{IQR} is called the "68% interquantile range" because $\Phi(+1) - \Phi(-1) \approx 0.68$. For the distributions $Y|X = x$ and $Z|X = x$, define:

$$\sigma_{\mathrm{IQR}}(x) \equiv \sigma_{\mathrm{IQR}}[Y|X = x] \tag{8.43}$$

$$\hat{\sigma}_{\mathrm{IQR}}(x) \equiv \sigma_{\mathrm{IQR}}[Z|X = x]. \tag{8.44}$$

Then we can see the following result for the calibrated resolution $\sigma_{\mathrm{IQR}}[\frac{Z}{x}|X = x]$:

The 68% IQR of the calibrated response distribution is given by $\sigma_{IQR}[\frac{Z}{x}|X = x] = \frac{1}{2x}\left(f^{-1}(I_Y^+) - f^{-1}(I_Y^-)\right)$.

Proof. We have

$$\int_{-\infty}^{I_Z^-} \rho_Z(z)dz = \Phi(-1) \tag{8.45}$$

$$\int_{-\infty}^{I_Z^+} \rho_Z(z)dz = \Phi(+1). \tag{8.46}$$

From Equation 8.12,

$$\rho_Z(z) = f'(z)\rho_Y(f(z)), \tag{8.47}$$

so that

$$\begin{aligned}
\Phi(-1) &= \int_{-\infty}^{I_Z^-} f'(z)\rho_Y(f(z))dz \\
&= \int_{-\infty}^{f(I_Z^-)} \rho_Y(u)du \\
\rightarrow f(I_Z^-) &= I_Y^- \qquad (8.48) \\
\Phi(+1) &= \int_{-\infty}^{I_Z^+} f'(z)\rho_Y(f(z))dz \\
&= \int_{-\infty}^{f(I_Z^+)} \rho_Y(u)du \\
\rightarrow f(I_Z^+) &= I_Y^+. \qquad (8.49)
\end{aligned}$$

Therefore,

$$\begin{aligned}\sigma_{\mathrm{IQR}}[Z|X=x] &= \frac{1}{2}\left(I_Z^+ - I_Z^-\right) \\ &= \frac{1}{2}\left(f^{-1}(I_Y^+) - f^{-1}(I_Y^-)\right), \end{aligned} \tag{8.50}$$

and

$$\sigma_{\mathrm{IQR}}\left[\frac{Z}{x}|X=x\right] = \frac{1}{2x}\left(f^{-1}(I_Y^+) - f^{-1}(I_Y^-)\right). \qquad \square \tag{8.51}$$

8.4 Discussion

After a quick summary in Section 8.4.1 of the results presented so far, Section 8.4.2 discusses the benefits and drawbacks of various methods of calibration, and Sections 8.4.3 and 8.4.4 describe extensions of numerical inversion that may help to improve closure.

8.4.1 Summary of Results

In Section 8.2 the concept of closure was defined in the process of calibrating the E_T of jets. It was found in Sections 8.3.1.1 and 8.3.2.1 that when using the mean or mode, respectively, of the distribution $Y|X=x$ to calibrate, closure is not necessarily achieved; with the response functions found at ATLAS or CMS, it is expected that numerical inversion will not close. Estimates were also provided for the non-closure for the mean (Equation 8.20) and for the mode (Equation 8.28). In those estimates it is found that as the underlying resolution $\sigma(x)$ or $\tilde{\sigma}(x)$ of the uncalibrated jet distribution $Y|X=x$ increases, the non-closure gets worse. This indicates that the non-closure issues raised in this chapter will become more important as the LHC moves to conditions with higher pile-up in the future.

A new calibration scheme based on the median of $Y|X=x$ is proposed in Section 8.3.3.1. With this method of calibration, closure is always achieved.

Each section also explored various definitions of the resolution of the fractional

calibrated jet distribution $\frac{Z}{x}|X = x$, where the most natural definition depends on the manner in which calibration has been performed (i.e., whether using the mean, mode, or median to calibrate). Useful estimates were provided for the standard deviation (Equation 8.22), the trimmed Gaussian width (Equation 8.32), and an exact formula for the 68% IQR (Equation 8.51). These expressions can be used to quickly estimate the final resolution of a jet algorithm without having to actually apply the calibration jet-by-jet.

8.4.2 Recommendation for Method of Calibration

As mentioned in the summary above, we have that for a non-linear response function closure is not necessarily achieved when using the mode or mean to calibrate, and closure is necessarily achieved when using the median. While this indicates that the median is a useful metric to use if closure is the main objective, we accept that there might be reasons to use the mode instead (for example, if the tails of $\rho_{Y|X}(y|X = x)$ are cut off, then the mode should stay constant while the median and mean will change). Thus we leave it to the reader to decide which method of calibration is most appropriate to use for their specific purposes. To that end, we also have discussion below about methods to improve the closure when the mode is being used to calibrate.

8.4.3 Iterated Numerical Inversion

A natural question is whether it is useful for the purposes of achieving closure to implement numerical inversion again on the calibrated jet collection, if closure has not been achieved the first time. We define the *iterated numerical inversion* process as follows:

With $C(x)$ defined as in Equation 8.7, let

$$R_{new}(x) \equiv C(x) \tag{8.52}$$

$$f_{new}(x) \equiv C(x)x. \tag{8.53}$$

Then, apply numerical inversion on the calibrated distribution Z:

$$Z \mapsto Z_{\text{new}} = f_{\text{new}}^{-1}(Z). \tag{8.54}$$

We then ask if the closure of this new distribution, $C_{\text{new}}(x)$ (defined analogously as in Equation 8.7), is closer to 1 than $C(x)$. In general, this is a difficult question to answer, but we have derived analytic approximations when the mode is used to derive the calibration (Appendix D.6). Iterating numerical inversion does *not* always help:

$$\frac{|C_{\text{new}}(x) - 1|}{|C(x) - 1|} \approx \frac{12 f''(x)^2 \tilde{\sigma}(x)^2}{f'(x)^4}. \tag{8.55}$$

If the ratio in Equation 8.55 is greater than 1, then the closure gets worse after a second iteration of numerical inversion. In particular, as $\tilde{\sigma}(x)$ gets larger, the iterated closure gets worse relative to the original closure. So we expect at higher levels of pile-up that iterating numerical inversion will not be useful. In Figure 8.5 we can see that iterating numerical inversion does make the closure worse than the original closure, in a model simulating higher pile-up conditions. The next section provides another scheme to correct for the residual non-closure that does not require iterating the process of numerical inversion.

8.4.4 Corrected Numerical Inversion

As noted above, when using the mean or mode of the distribution $Y|X = x$ to calibrate, closure is not achieved in general. With the closed-form estimates of the non-closure provided in the text, one might think to simply "subtract off" the non-closure. However the non-closure estimates provided are in terms of the truth E_T value x. Since x is not available in data, a sensible proxy is to use numerical inversion as an estimate for x. This is actually equivalent to iterated numerical inversion, which as shown in the previous section does not always help.

Another possibility is to use a different original response function to perform the calibration. Suppose that instead of using $f(x) = R(x)x$, there was a new function

$g(x) \neq f(x)$ such that if the calibration is performed with this new function, $Y \mapsto Z_{\text{corr}} = g^{-1}(Y)$, the new calibrated distribution $Z_{\text{corr}}|X = x$ does achieve closure or gets closer to achieving closure than when calibrating using f.

We define the *corrected numerical inversion* process as follows:

1. Calculate $f(x) = f_{\text{mo}}(x) = \text{mode}[Y|X = x]$.
2. Let $g(x) = g(x; f(x))$ be a calibration function depending on the fitted function $f(x)$.
3. Apply the calibration $Y \mapsto Z_{\text{corr}} = g^{-1}(Y)$ jet-by-jet.

We then can examine the closure

$$C_{\text{corr}}(x) = \text{mode}\left[\frac{Z_{\text{corr}}}{x}\middle| X = x\right]. \tag{8.56}$$

And say we have achieved closure if

$$C_{\text{corr}}(x) \equiv 1. \tag{8.57}$$

We examine the case of using the mode to measure closure, again because in practice that is what is often used when there are significant non-Gaussian tails.

One way to specify g is by explicitly requiring closure. In Appendix D.7 it is shown that in the case that closure is achieved exactly, g necessarily satisfies the differential equation[11]

$$0 = g''(x) - g'(x)^2 \frac{g(x) - f(x)}{\tilde{\sigma}(x)^2}. \tag{8.58}$$

In principle Equation 8.58 can be solved numerically given numerical fitted values $f(x)$ and $\tilde{\sigma}(x)$, though in practice such a method may prove intractable.

[11] As noted in the derivation, this equation also assumes the following: that the underlying distribution $Y|X = x$ is approximately Gaussian in the vicinity of its mode $f(x)$; and that the correction is small, with $|g(x) - f(x)| \ll \tilde{\sigma}(x)$.

Another way to specify g is to use external parameters

$$g(x) = g(x; f(x); a_1, ..., a_n). \tag{8.59}$$

Then the parameters $a_1, ..., a_n$ can be chosen such that the closure is as close to 1 as possible. This is the method used to find the corrected calibration curve in Figure 8.5, and explained in more detail in Appendix D.8. The absolute non-closure $|C - 1|$ is significantly smaller than the original non-closure, even in a model simulating very high pile-up conditions.

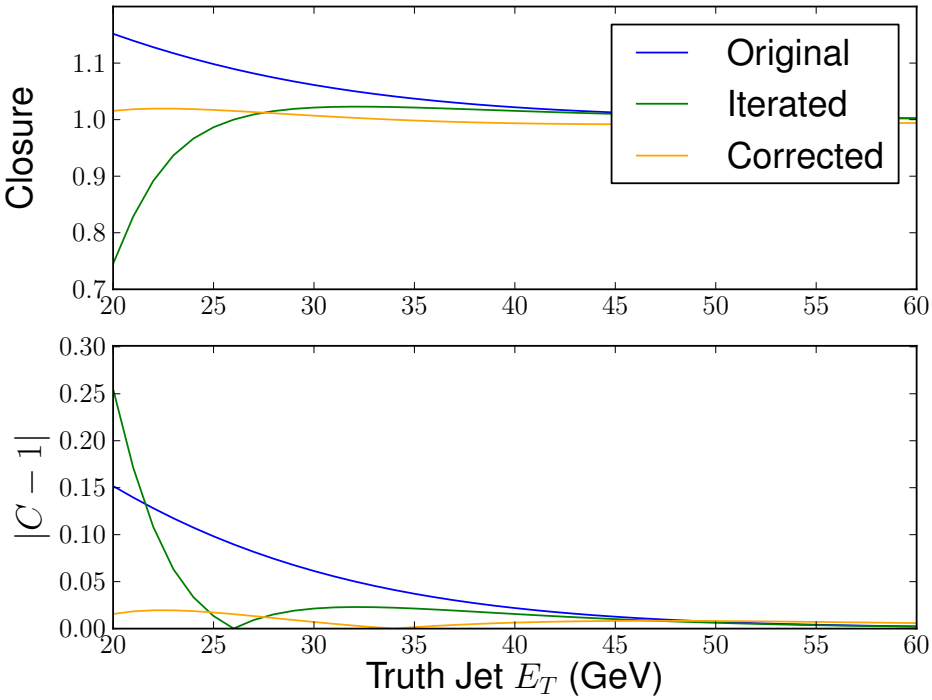


Figure 8.5: The top plot shows the closure of numerical inversion when using the mode to calibrate, using a toy model similar to conditions in ATLAS or CMS but increasing $\sigma(x)$ by a factor of 1.4 in order to simulate higher pile-up conditions. In blue, the original closure as defined in Equation 8.7. In green, the closure after iterating numerical inversion once as in Section 8.4.3. In orange, the closure after using the parameterized corrected numerical inversion technique as in Section 8.4.4. For details of the model, see Appendix D.9. The bottom plot shows the absolute non-closure $|C-1|$. In particular, at low E_T, iterating numerical inversion does worse, while corrected numerical inversion does better than the original calibration.

8.5 Conclusion

Jets are ubiquitous at the LHC and are used extensively in the two main searches presented in this Thesis (Chapter 6 and Chapter 7). The calibration of the energy and mass of jets is necessary for their practical use in these analyses or any other, and understanding and improving this calibration can improve the sensitivity to new physics. The standard technique used for jet calibration at ATLAS and at CMS is numerical inversion. This chapter serves to formally define numerical inversion and derive many of its properties.

This study has resulted in a few novel discoveries and clarifications about this standard technique.

It is firmly established that numerical inversion does achieve its key property, which is that it is independent of the underlying distribution used to derive the calibration.

However, numerical inversion is established to be inherently biased: when using the mean or mode of the reconstructed kinematic distributions to calibrate, the calibrated response function does not close, if the uncalibrated response function is nonlinear. The size of this nonclosure is approximated for the first time and is shown to get worse as the resolution of jets gets worse, in particular as is expected as the LHC moves to higher instantaneous luminosity and pile-up conditions in the future. When using the median as the notion of central tendency, the method is guaranteed to close. However, there are good reasons to use the mode instead of the median, and one method is not recommended over the other because of these respective drawbacks.

An approximation for the calibrated jet resolution is provided which can be used to estimate the resolution before actually doing the calibration. Previous approaches to do this estimation were shown to be inaccurate; the approximation derived here uses the derivative of the uncalibrated response, and is shown to be more accurate.

Given the inherent nonclosure present in numerical inversion, the question of whether numerical inversion can be improved is answered in a couple different ways. First, simply iterating numerical inversion is shown for the first time to not always help. Second, an extension is proposed in which a modified calibration function can

be constructed to achieve a better closure than the standard technique.

In addition to the corrections and extensions proposed, the deep understanding generally of numerical inversion provided for the creation of the technique described in Chapter 9. The realization that numerical inversion was defined by a single function (Section 8.2.2), and that this function could be modified to achieve better closure, led directly to the idea that this function could be approximated with a neural network, which could simultaneously and automatically take into account multiple auxiliary variables.

While this Thesis focuses on the use of jets at the LHC, numerical inversion is in fact a generic detector calibration method that can be applied as long as there is a reliable method for matching the value of an observable before and after passing through the detector (either via simulation or high-fidelity calibration data). The results presented here therefore can be applied to objects other than jets at the LHC, or indeed even beyond the realm of high energy physics.

Chapter 9

Improving Jet Calibrations with Machine Learning

Jets reconstructed in the ATLAS detector have their energies corrected in order to be an unbiased estimate of the particle-level jet. This correction proceeds via multiple steps, the most significant of which include an absolute MC-based jet energy scale correction and following a correction based on auxiliary features of the jet observed in the detector. These two steps both use numerical inversion (Chapter 8) in order to remain independent of the underlying spectrum. The current method of accounting for these features relies on a sequential application of numerical inversion to one feature at a time. This Chapter introduces a new technique which can perform this correction accounting for multiple feature simultaneously, accomplished through the use of neural networks. The multivariate nature of the new technique allows to account for correlations between the energy dependence on multiple features, which is not accounted for in the current method. The new technique, called *generalized numerical inversion*, generalizes numerical inversion by retaining the key property of being independent of the underlying prior distribution, while accounting for multiple features simultaneously. This new technique further improves the correction on a variety of metrics relative to the sequential method. As a method for improving the jet energy scale and resolution, the new technique is widely applicable to analyses in ATLAS, and in particular to the two searches presented in this Thesis (Chapter 6

and Chapter 7). The results presented in this Chapter are published in [350].

9.1 Introduction

The two searches presented in this Thesis (Chapter 6 and Chapter 7) use jets extensively, and besides that nearly every analysis at the Large Hadron Collider (LHC) uses jets in some capacity, both Standard Model (SM) measurements and searches for physics beyond the SM. All of these analyses rely on an accurate and precise measurement of the jet energy. As mentioned in Section 5.5.4.1, due to various distortions arising from detector effects, the reconstructed jet energy deviates from the particle-level jet energy in the absence of a detector. These distortions are corrected for using a series of calibration steps [80, 81, 409]. All of these steps proceed ultimately via numerical inversion, the properties of which are expounded upon in detail in Chapter 8. This Chapter is concerned with improving one particular and important part of the current calibration procedure, the Global Sequential Calibration (GSC). This step incorporates auxiliary measurements about the jet, such as its reconstructed charged particle multiplicity, to improve the overall quality of the energy reconstruction [414].

The correction based on auxiliary features intends to further improve the jet calibration via a few metrics. First, using features correlated with the provenance of the jet, the difference in energy scale between, e.g., quark and gluon jets can be reduced. Second, by reducing the dependence of the scale on auxiliary features, the effects of mismodeling in the underlying spectrum of these features can be mitigated. Third, the correction can remove the additional component of the jet energy resolution relative to the particle-level quantity due to the spread of these underlying features.

The current method for the GSC treats the impact of each auxiliary feature sequentially and independently, which requires the set of features used in the correction to be chosen in such a way to have independent effects on the jet energy. However, there is additional information available which could improve the calibration and yet is not used in the current GSC because of this requirement of independence. This Chapter presents a new technique using neural networks to simultaneously take into account multiple auxiliary variables in this correction which may or may not have

correlated effects on the jet energy. The technique relies both on the general capacity of neural networks to approximate well complicated functions in high dimensions [415] and on recent advances in jet substructure [86] which indicate useful features for applying this correction. As mentioned in Chapter 8, it is important that the calibration is independent of the underlying p_T spectrum, and therefore that ultimately it relies on numerical inversion. The new technique presented here, called *generalized numerical inversion*, incorporates neural networks into the numerical inversion framework in order to accomplish both goals of simultaneously taking into account the effect of multiple auxiliary variables and of remaining independent of the underlying spectrum. Ultimately this technique is shown to improve the overall quality of the jet calibration on the three metrics outlined above.

This Chapter is organized as follows. Section 9.2 discusses the simulated samples used in this study, and Section 9.3 reviews the GSC and describes it in the notation used in this Thesis. Generalized numerical inversion is described in Section 9.4 and first results using this new method are presented in Section 9.5. There is a discussion of the results in Section 9.6 and the Chapter concludes with Section 9.7.

9.2 Event Simulations

Studies documented within this chapter are performed using a variety of Monte Carlo (MC) simulated samples (Section 5.3). Dijet events are generated at Leading Order (LO) in Pythia8.1 [244] with the $2 \to 2$ matrix element convolved with the NNPDF2.3LO Parton Distribution Function (PDF) set [379] and using the A14 set of tuned parameters [251]. An additional sample is simulated using Herwig 7.0 [248] using the NNPDF3.0 NLO PDF and the default set of tuned parameters for the underlying event. Both Pythia and Herwig are interfaced with EvtGen for heavy flavor decays [253, 416].

All simulated events have been reconstructed using a full simulation of the ATLAS detector [340] implemented in Geant4 [255], which describes the interactions of particles with the detector and the subsequent digitization of analog signals. The effects of multiple simultaneous pp collisions (pile-up) are simulated with minimum

bias pp collisions using PYTHIA8.1 and overlaid on the nominal dijet interactions.

9.3 Global Sequential Calibration

As mentioned in Section 5.5.4.1, the ATLAS jet calibration proceeds via multiple steps. The largest correction is the absolute MC-based correction which brings the overall scale of the reconstructed jet E to the truth or particle-level E. The step immediately proceeding is the GSC, which corrects the dependence of the jet p_T on various jet quantities derived from information in the tracker, calorimeter, and muon system [80, 342, 413].

There may be a little confusion here in moving from the calibration of E_T to p_T. The first thing to note is that this is just simply how the ATLAS calibration does it, first an absolute correction to E_T and then following a residual correction to p_T. The second thing to note is that E_T and p_T have a well-defined relation: $E_T = \sqrt{p_T^2 + m^2}$, where m is the mass of the jet. The GSC is only applied to small-R jets, intended to apply to quark- and gluon-initiated jets in which the true mass is either 0 or much less than the p_T (the lowest p_T in the calibration is 20 GeV, while the highest quark mass reconstructed in a small-R jet is the bottom quark, with mass of ~ 4 GeV). This is related to the fact that the opening angle of a jet is generically roughly $\Delta R \sim \frac{2m}{p_T}$ [185]; for a small-R jet, with ΔR parameter of 0.4, $\frac{m}{p_T} < 0.2$. Because of the negligibility of this term, for small-R jets it is the case that $E_T \sim p_T$, and the term "jet energy" correction is used somewhat loosely in this Chapter (and in the previous one) to refer to the ATLAS jet E_T and p_T corrections in the calibration. Generally the distinction is not important, but where it is the text is clear about what is being corrected.

For large-R jets, the ΔR parameter is 1.0, allowing for jets with larger ratios of $\frac{m}{p_T}$. The GSC is not applied to these jets, but rather after the inclusive E_T correction an energy-dependent mass calibration is applied (Section 5.5.4.1). The Author has contributed to work which directly uses the techniques developed in this Chapter to employ neural networks to simultaneously calibrate the mass and energy of large-R jets [292].

As mentioned in Chapter 8, after the inclusive correction it is approximately true

that $f(x) \equiv \langle p_T^{reco} | p_T^{true} = x \rangle = x$[1] (Equation 8.28), i.e. that the calibration closes and that the average reconstructed or detector-level jet p_T (p_T^{reco}) is an unbiased estimator[2] of the truth or particle-level p_T (p_T^{true}) [3]. The reason to follow this with the GSC is that even though the overall response closes, $f(x)$ may have a residual dependence on auxiliary information available from the detector.

The GSC uses 5 auxiliary features which can be grouped into 3 categories based on the effect being corrected for. The first two features are the number of tracks in the jet (associated via ghost association [84]) and the width of tracks in the jet (the average p_T-weighted distance of tracks from the jet axis). Since the calorimeter-cell response is non-linear due to noise suppression, a jet with more particles (and therefore softer on average for fixed total p_T) will tend to have a lower measured energy than a jet with a few hard particles. The number of tracks in the jet is a good but not perfect measure of the number of particles, as tracks are insensitive to pile-up, but only correspond only to charged particles. This effect is correlated to the provenance of the jet, as gluon jets tend to have more particles and a wider radiation pattern than quark jets [170]. The track width is used to further discriminate between gluon and quark jets, and therefore gain more information about the number of particles produced in the jet. The correction for the first two features has the benefit of making the response more similar between quark- and gluon-initiated jets, which reduces the sample dependence of the jet p_T closure.

The second two features are the fraction of jet energy in the final layer of the electromagnetic calorimeter and in the first layer of the hadronic calorimeters; these features account for energy loss in the dead material between the electromagnetic and hadronic calorimeters.

The final feature counts the number of muon segments behind the jet that register hits; this feature accounts for energy not measured in the calorimeter due to showers

[1] In practice $f(x)$ is measured using the mode of the distribution rather than the mean.

[2] An estimate $\hat{\theta}$ of a fixed but unknown parameter θ is an unbiased estimator if, for all values of θ, $\mathbb{E}\left[\hat{\theta}|\theta\right] = \theta$.

[3] Of course, this statement is subject to the stipulations mentioned in that Chapter, and in fact a major result from that study was understanding exactly how and why the calibration *does not* close. However, ultimately the understanding from that study was that the major non-closures would be a bigger issue at higher pile-up conditions that will exist at the LHC in the future.

that punch through to the MS.

Let θ represent the available auxiliary information about a jet[4], e.g. the 5 features used in the GSC. Then, the function f can be generalized as $f_\theta(x) \equiv \langle p_T^{reco} | p_T^{true} = x, \theta \rangle$.

In general, if it is found that the jet energy $f_\theta(x)$ depends on some auxiliary variable θ, then it is desirable that the calibration removes this dependence, for a few reasons. First, if the distribution of θ is sample-dependent, e.g. correlated with the jet originating parton, then there could be closure in one sample but not another, which could cause suboptimal selections in the final analysis cuts. Second, if θ is not well-modeled by the simulation then even though there may be closure in the simulation there may not be closure in the data, and this could introduce systematic uncertainties in the final calibration. Third, there is a degradation of the resolution $\sigma(x)$ due to the spread of θ, which can cause the sensitivity of the final analysis selection to be worse than with an undegraded resolution. These are the three metrics by which we judge the performance of the GSC and of generalized numerical inversion in this Chapter.

However, this third point is more subtle than it naively appears. While it is true that there is a degradation of the resolution due to the spread of θ, it is *not* the case that the resolution always gets better when the dependence of the response on an auxiliary variable is corrected. In particular, there exist certain situations where correcting for the dependence of the response on an auxiliary variable can actually make the resolution worse. This can occur, for example, if θ is correlated with p_T^{true} in such a way that removing the correlation between p_T^{reco} and θ also removes some of the correlation between p_T^{reco} and p_T^{true}. These situations are explored in more detail in Appendix E.1.

It is important to note that this auxiliary information θ is available at the detector or reconstructed level. Therefore, while $f_\theta(x)$ is derived in simulation, for a given jet at detector level θ is known and $f_\theta(x)$ can be inverted as usual across its argument to give the calibration function. For $\theta \in \mathbb{R}$, the correction is therefore given by $p_T^{reco} \mapsto \hat{p}_T^{reco} = f_\theta^{-1}(p_T^{reco})$. In practice, the distribution of θ is binned and the numerical inversion is performed for different functions in each bin of θ.

[4]Note that θ can be a single value or a vector with multiple features.

When $\theta \in \mathbb{R}^n$, a sequential calibration like the GSC (for which $n = 5$) proceeds with a sequential (as the name suggests) application of numerical inversion:

$$p_T^{\text{reco}} \mapsto \hat{p}_T^{\text{reco}} = f_{\theta_n}^{-1}\left(\cdots f_{\theta_2}^{-1}\left(f_{\theta_1}^{-1}\left(p_T^{\text{reco}}\right)\right)\cdots\right). \tag{9.1}$$

The sequential method removes all residual dependencies when $f_{\theta_i}(x)$ is independent of $\theta_{j\neq i}$, i.e. when $f_\theta(x)$ is entirely determined by one feature at time. If there are such dependencies of $f_\theta(x)$ on more than one feature θ_i, then there could be residual dependencies on some combination of the θ_i after the full sequential correction.

9.4 Generalized Numerical Inversion

The idea of generalized numerical inversion is to simply replace the sequential approach in Equation 9.1 with a single inversion: $p_T^{\text{reco}} \mapsto \hat{p}_T^{\text{reco}} = f^{-1}_{\theta_1,\ldots,\theta_n}(p_T^{\text{reco}})$. Binning the response in n dimensions would require significant computing resources and so is practically infeasible. An unbinned approach that can simultaneously capture the dependence on many features is needed to exploit the potential correlations in the response on the θ_i. One powerful tool for this purpose is a neural network. In principle, neural networks can approximate any smooth function [415, 417, 418] and have been shown to provide excellent performance for a wide variety of classification and regression tasks with limited training data [391–396]. Inverting a neural network can be non-trivial, so the procedure for generalized numerical inversion is adapted as follows:

1. Learn a neural network approximation $L(x, \theta)$ to the function $f_\theta(x) = \langle p_T^{\text{reco}} | p_T^{\text{true}} = x, \theta\rangle$.
   Note that, if θ is a vector with n features, then $L(x, \theta) : \mathbb{R}^{n+1} \to \mathbb{R}$.

2. Learn a neural network $C(L(x, \theta), \theta)$ that tries to predict x given θ and $L(x, \theta)$. This is an approximation to the family of functions $f_\theta^{-1}(x)$. Note that learning the inverse this way is technically simple since L is single-valued.

3. Calibrate with $p_T^{reco} \mapsto \hat{p}_T^{reco} = C(p_T^{reco}, \theta)$. The calibration non-closure is defined as usual as the deviation of $\langle \hat{p}_T^{reco} / p_T^{true} | p_T^{true} = x, \theta \rangle$ from 1.

There are therefore two learning stages in this procedure. The first step learns a neural network that predicts the average behavior of p_T^{reco} given p_T^{true} and θ while the second network tries to approximate a single-valued function. In principle, the above procedure could be applied to the entire jet calibration procedure described in Section 5.5.4.1, but the focus here is on the component addressing residual dependencies, following the inclusive jet energy calibration.

A schematic of the generalized numerical inversion method is shown in Figure 9.1.

Figure 9.1: A schematic of the generalized numerical inversion method. First, a neural network L is trained to learn p_T^{reco} from p_T^{true} and θ, which is an approximation to the functions $f_\theta(x)$. Then, a second neural network C is trained to learn p_T^{true} from $L(p_T^{true}, \theta)$ and θ, which is an approximation to the functions $f_\theta^{-1}(x)$. Finally, C is used as the calibration function.

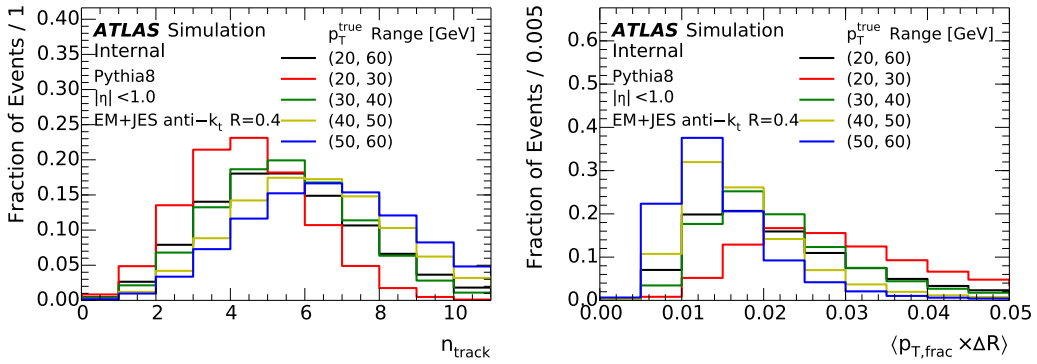
9.5 Results

To illustrate the potential of generalized numerical inversion, an example is provided using jets from the ATLAS full detector simulation. Neural network training is performed using scikit-learn [419] with a simple two-layer feed-forward neural network that has 100 hidden nodes in each layer and a rectified linear unit activation function. To lay the groundwork for application to the full GSC, two key jet features are studied: the number of tracks above 1 GeV associated to the jet (n_{track}) and the average

track radius[5]:

$$\Delta R_{\text{track,avg}} \equiv \begin{cases} \frac{1}{n_{\text{track}}+1} \sum_{\text{tracks}} (p_{T,\text{track}} / \sum_{\text{tracks'}} p_{T,\text{track'}}) \times \Delta R_{\text{track,jet}} & \text{if } n_{\text{track}} > 0 \\ -1 & \text{if } n_{\text{track}} = 0 \end{cases}$$

The average track radius is also sometimes denoted more explicitly as $\langle p_{T,\text{frac}} \times \Delta R \rangle$. As mentioned in Section 9.3, these observables account are correlated with the type of parton that produced the jet (gluon or quark), and account for effects on the p_T response due to the non-linear calorimeter-cell response. The underlying distributions of these two variables are given in Figure 9.2.



(a) (b)

Figure 9.2: The distributions of (a) n_{track} and (b) $\Delta R_{\text{track,avg}} = \langle p_{T,\text{frac}} \times \Delta R \rangle$ in bins of p_T^{true}.

Figure 9.3 shows the dependence of the response (following the inclusive energy calibration) on n_{track} and $\Delta R_{\text{track,avg}}$. The response depends strongly on these two quantities, varying by about 10-20% across the accessible range. The trends are also not uniform in jet p_T - the response of lower p_T jets shows a stronger dependence on both n_{track} and $\Delta R_{\text{track,avg}}$.

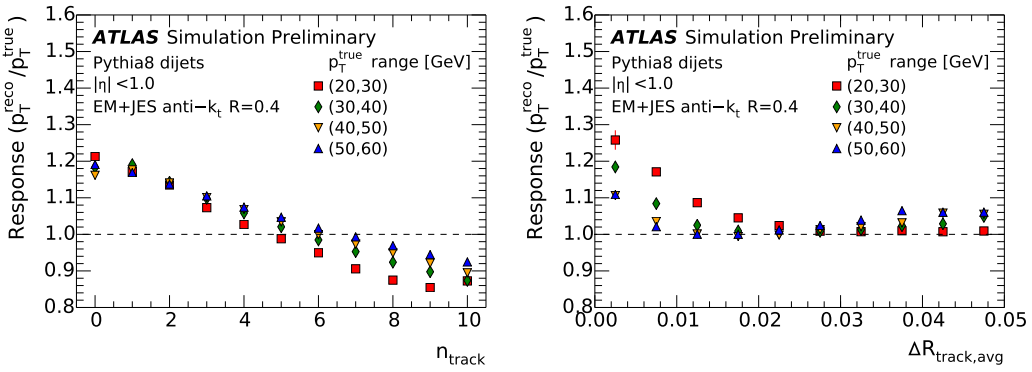
[5]The GSC currently used in ATLAS uses instead the track width, which is the average track radius multiplied by $(n_{\text{track}} + 1)$. The residual dependence of the n_{track}-calibrated response on the track width is negligible and is thus not useful for benchmarking generalized numerical inversion as the sequential calibration is nearly the same as the simultaneous calibration.

In order to understand the performance provided by a simultaneous instead of a sequential approach, approximations are derived for the calibration functions $f_{n_{track}}(x)$, $f_{\Delta R_{track,avg}}(x)$, and $f_{n_{track},\Delta R_{track,avg}}(x)$, and the sequential calibration $f^{-1}_{\Delta R_{track,avg}}(f^{-1}_{n_{track}}(p_T^{reco}))$ is compared with the simultaneous calibration $f^{-1}_{n_{track},\Delta R_{track,avg}}(p_T^{reco})$. The sequential calibration presented here differs in some details with the standard GSC used in ATLAS; a major difference is that the calibration presented here does an unbinned fit to p_T^{reco} and the feature θ while the GSC does a binned fit which is then smoothed. To control for this ability of the neural network to operate unbinned, the generalized numerical inversion approach is used for both the one- and two-feature cases in order to study only the differences between the sequential and the simultaneous calibrations. However, importantly the sequential approach demonstrated here does no worse than the GSC in correcting for the residual dependence of the response.

The learned functions L in the sequential calibration are presented in Figures 9.4a and 9.4d. As expected, Figure 9.4a looks similar to Figure 9.3a, as the neural network has learned to approximate the shape of the response with respect to n_{track} (Figure 9.4d is not expected to look similar to Figure 9.3b, since the previously applied n_{track} correction affects the dependence of the response on $\Delta R_{track,avg}$). The ratio of p_T^{reco} to $L(p_T^{true})$ (for the appropriate θ, at the appropriate step of the sequential calibration) is shown in Figures 9.4b and 9.4d, and the ratio is very close to 1, indicating the learning step is working properly. The closure of the calibrations at each step of the sequence is shown in Figures 9.4c and 9.4f. In both cases, the calibration closes, with an average calibrated response at unity, independent of the features.

Corresponding results for the network trained to simultaneously learn the dependence on n_{track} and $\Delta R_{track,avg}$ are shown in Figure 9.5. The network is able to learn the dependence on either of the two features in one dimension and Figures 9.5c and 9.5f also show that the calibration closes.

The advantage of the simultaneous method is highlighted in Figure 9.6. After the calibration, in each step of the sequential method, there is still a residual dependence of the calibrated response $\hat{R}$ on n_{track} in individual bins of $\Delta R_{track,avg}$. This residual dependence is monotonic, and so the single value $d\hat{R}/dn_{track}$ (obtained as the slope



(a) (b)

Figure 9.3: The dependence of the response on (a) n_{track} and (b) $\Delta R_{track,avg}$ in several bins of truth jet p_T.

when fitting $\hat{R}$ versus n_{track} to a line in bins of $\Delta R_{track,avg}$) parameterizes the remaining residual dependence. The residual dependence on n_{track} in bins of $\Delta R_{track,avg}$ is nearly the same and clearly not zero at all values of $\Delta R_{track,avg}$ if the n_{track} or sequential n_{track}, $\Delta R_{track,avg}$ calibrations are applied. However, the residual dependence is zero and independent of $\Delta R_{track,avg}$ when the simultaneous calibration is performed. Similarly, there is still a residual dependence of $d\hat{R}/d\Delta R_{track,avg}$ on n_{track} if the n_{track} or sequential n_{track}, $\Delta R_{track,avg}$ calibrations are applied; this residual dependence goes away with the simultaneous calibration.

For example, in the sequential calibration, $d\hat{R}/dn_{track} \approx -0.02$ for large $\Delta R_{track,avg}$. This means that two jets with values of n_{track} differing by 5 (and the same large values of $\Delta R_{track,avg}$) will have a response difference of ~ 0.1. Similarly, in the sequential calibration, $d\hat{R}/d\Delta R_{track,avg} \approx -4$ for large n_{track}. Two jets that differ in $\Delta R_{track,avg}$ by 0.02 (both with large n_{track}) will therefore have a response difference of ~ 0.08.

Because of this residual dependence, there can be nonclosures in specific regions of the parameter space, despite the calibration closing inclusively overall. Figure 9.6c shows a selection on n_{track} and $\Delta R_{track,avg}$ which is more efficient on gluons than on quarks, and Figure 9.6d shows a selection which is more efficient on quarks than on gluons. In the sequential calibration, large non-closures can be seen in some regions of the parameter space, while the simultaneous calibration almost completely removes

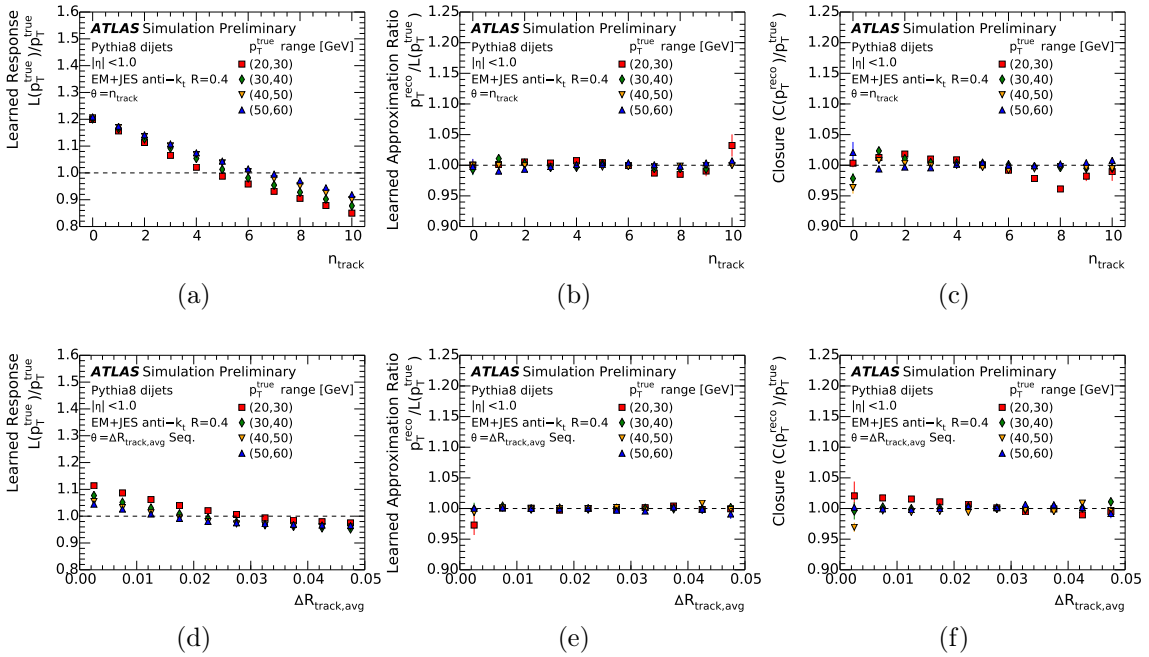


Figure 9.4: The dependence of the (a) learned response $L(p_T^{true}, \theta)/p_T^{true}$, (b) ratio of p_T^{reco} to learned approximation $L(p_T^{true}, \theta)$, and (c) calibrated response $C(p_T^{reco}, \theta)/p_T^{true}$ on n_{track} in several bins of truth jet p_T for $\theta = \{n_{track}\}$. Also, the dependence of the (d) learned response $L(p_T^{true}, \theta)/p_T^{true}$, (e) ratio of p_T^{reco} to learned approximation $L(p_T^{true}, \theta)$, and (f) calibrated response $C(p_T^{reco}, \theta)/p_T^{true}$ on $\Delta R_{track,avg}$ in several bins of truth jet p_T for $\theta = \{\Delta R_{track,avg}\}$ in sequence after the n_{track} correction.

these nonclosures.

As observables correlated with the quark or gluon provenance of a jet, the correction for the dependence of the energy on n_{track} and $\Delta R_{track,avg}$ can reduce the response difference between jet types. Figure 9.7 shows the difference between the response for quark and gluon jets as a function of the truth jet p_T for different calibration methods. Since gluon jets have a softer constituent p_T spectrum, their response is lower than for quark jets on average. The residual correction does reduce the response difference; the simultaneous approach is slightly better than the sequential approach at every value of truth jet p_T.

The properties of quark and gluon jets can depend on the specific model of jet fragmentation used; therefore, it is also important to check that the improvement in

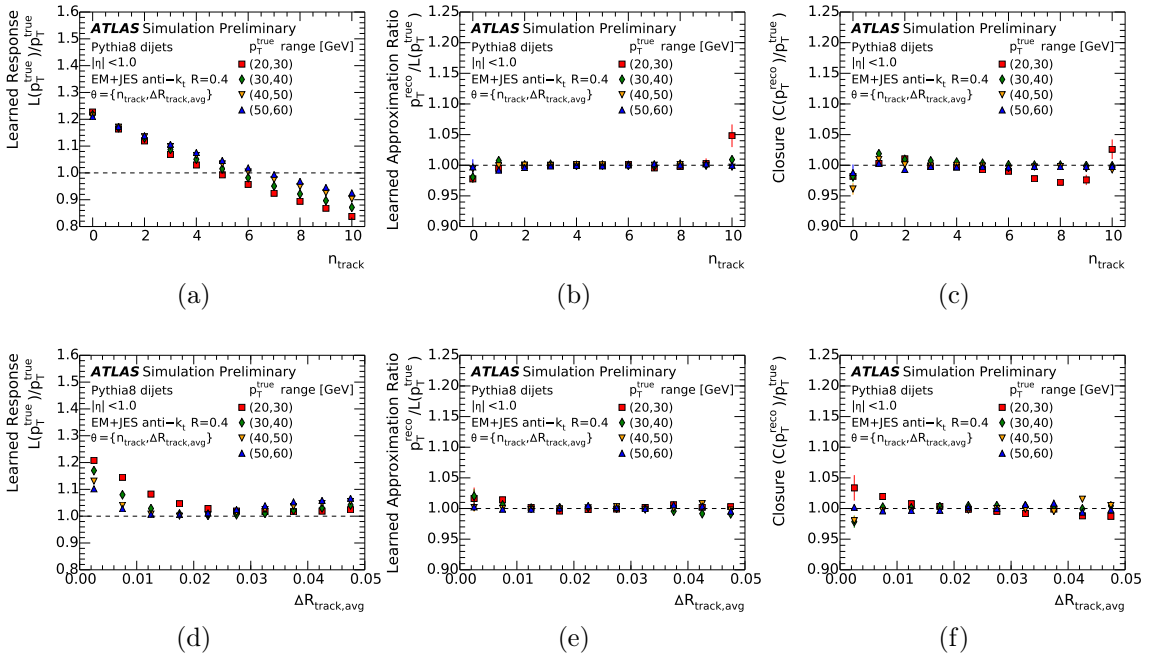


Figure 9.5: The dependence of the (a,d) learned response $L(p_T^{true}, \theta)/p_T^{true}$, (b,e) ratio of p_T^{reco} to learned approximation $L(p_T^{true}, \theta)$, and (c,f) calibrated response $C(p_T^{reco}, \theta)/p_T^{true}$ on (a,b,c) n_{track} and (d,e,f) $\Delta R_{track,avg}$, respectively, in several bins of truth jet p_T for a simultaneous calibration with $\theta = \{n_{track}, \Delta R_{track,avg}\}$.

the quark-gluon response difference is robust to these model differences. In addition to showing the results for PYTHIA 8, Figure 9.7b also demonstrates that (without retraining) there are also improvements for HERWIG 7.

Another goal of the correction based on auxiliary variables is to reduce the sensitivity of the jet energy to mismodeling of the underlying features. An important consequence of such a reduction could be a smaller systematic uncertainty associated with jet fragmentation modeling. Figure 9.8 quantifies this effect by showing the difference in the jet response between PYTHIA 8 and HERWIG 7 for quarks and gluons separately. It can be seen that the jet response difference between the generators is generally better in the simultaneous calibration method than in the sequential method. In particular, in some bins of p_T^{true}, the sequential method makes the difference between the generators worse than not including the second variable, while the

simultaneous method retains or slightly improves the difference between the generators. This could happen if the correlation between the two features changes between the generators in such a way that the sequential calibration effectively undoes the correction of the first feature in the sample the network was not trained on. Another possible effect is that the correlation between the two features stays the same between the generators, but the underlying distribution of one or both of the features in the jets changes. In either case, since the simultaneous method conditions on both features at once, this does not affect the final performance.

A final metric for studying the impact of generalized numerical inversion is the jet energy resolution. One component of the jet energy resolution is due to the spread in the jet energy response for various values of θ - in the extreme case that there is no spread in p_T^{reco} given p_T^{true} and θ, there will still be an effective resolution resulting from the spread in values over the range of θ. However, as mentioned in Section 9.3 and expounded upon in Appendix E.1, the calibrated resolution does not necessarily improve following the correction based on auxiliary variables. Figure 9.9 shows both the inclusive (not differential in θ) closure as well as the inclusive resolution as a function of the truth jet p_T, for different calibration methods. All methods achieve a similar inclusive closure and as desired, the residual calibration procedures reduce the resolution. The overall resolution reduction is similar for the sequential and simultaneous approaches, with the simultaneous slightly better at higher p_T and the sequential slightly better at low p_T. The difference in the resolutions from the various approaches is small compared to the gain over no residual calibration.

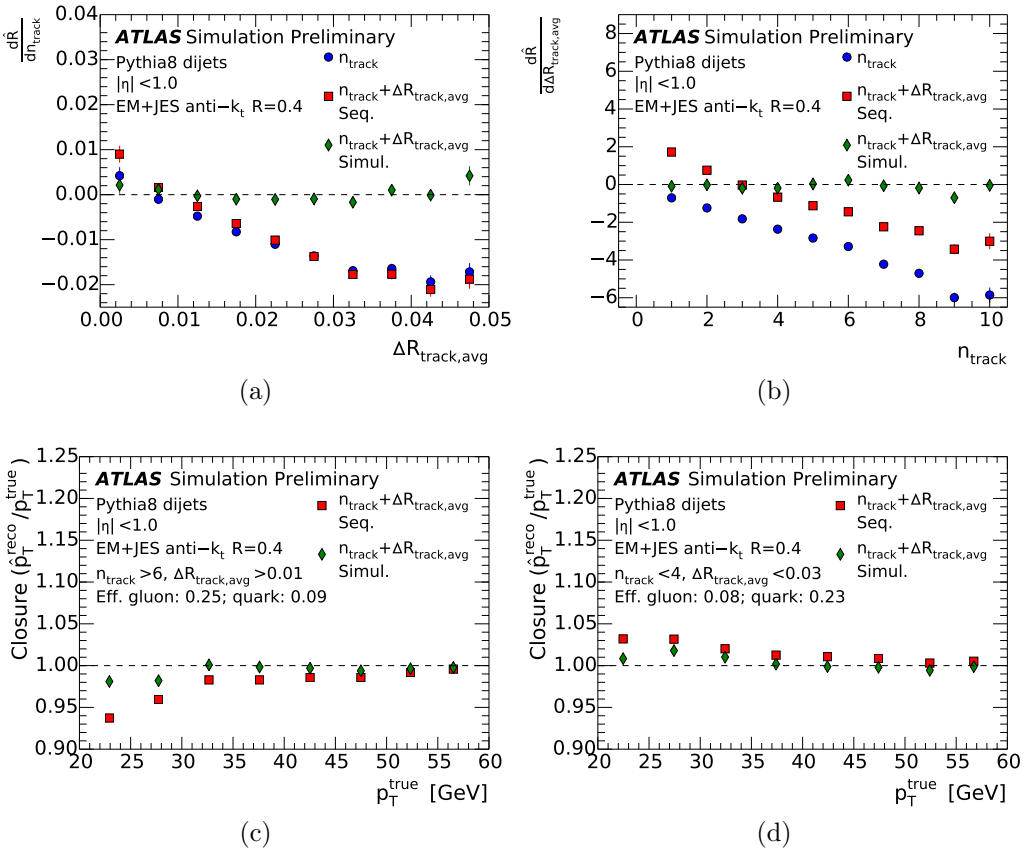
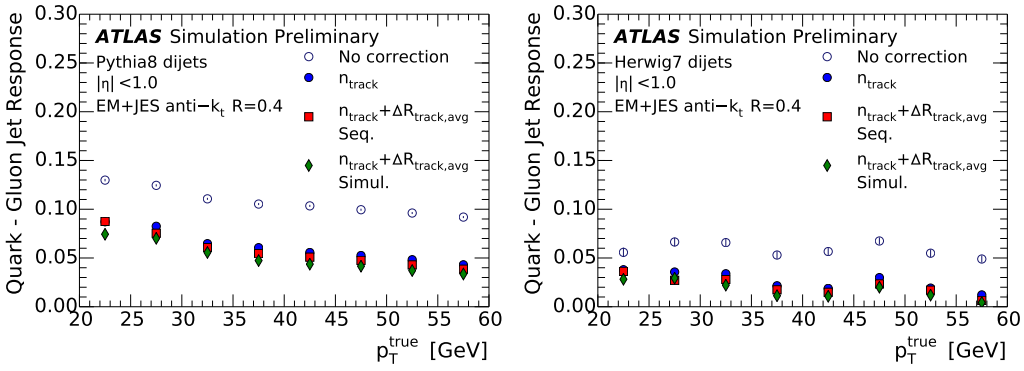
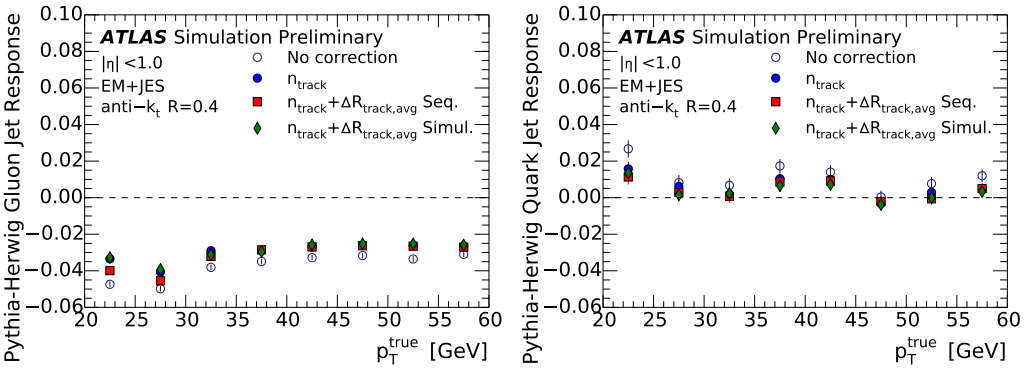


Figure 9.6: The dependence of (a) $\frac{d\hat{R}}{dn_{track}}$ on $\Delta R_{track,avg}$ and (b) $\frac{d\hat{R}}{d\Delta R_{track,avg}}$ on n_{track} for: a calibration using a network with $\theta = \{n_{track}\}$ (circles); a calibration using a network with $\theta = \{\Delta R_{track,avg}\}$ employed sequentially after correcting for n_{track} (squares); and a simultaneous calibration using a network with $\theta = \{n_{track}, \Delta R_{track,avg}\}$ (diamonds). Also, the closure as a function of p_T^{true}, in (c) a selection intended to target gluon jets; and (d) a selection intended to target quark jets.



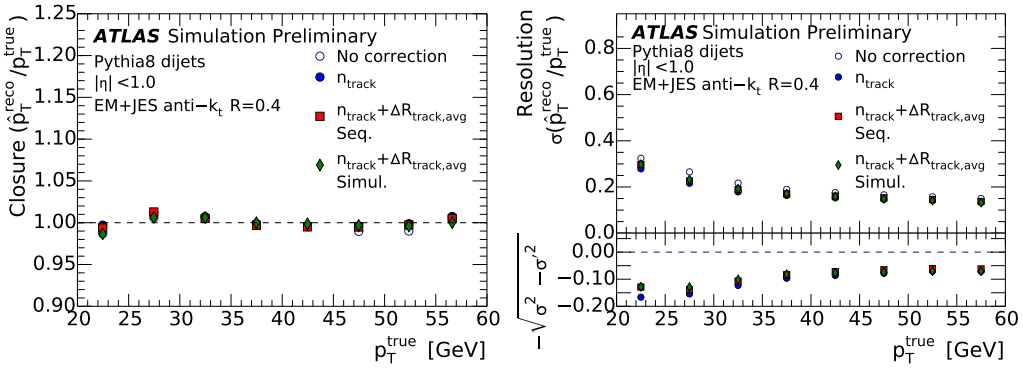
(a) Pythia 8 (b) Herwig 7

Figure 9.7: The difference between the response of quarks and gluons as a function of p_T^{true} for: before any n_{track} or $\Delta R_{track,avg}$ correction (open circles); a calibration using a network with $\theta = \{n_{track}\}$ (circles); a calibration using a network with $\theta = \{\Delta R_{track,avg}\}$ sequentially after correcting for n_{track} (squares); and a simultaneous calibration using a network with $\theta = \{n_{track}, \Delta R_{track,avg}\}$ (diamonds).



(a) (b)

Figure 9.8: The difference between the response of jets in Pythia8 and Herwig7 as a function of p_T^{true} in (a) gluon jets, and (b) quark jets, for: before any n_{track} or $\Delta R_{track,avg}$ correction (open circles); a calibration using a network with $\theta = \{n_{track}\}$ (circles); a calibration using a network with $\theta = \{\Delta R_{track,avg}\}$ sequentially after correcting for n_{track} (squares); and a simultaneous calibration using a network with $\theta = \{n_{track}, \Delta R_{track,avg}\}$ (diamonds).



(a) (b)

Figure 9.9: The (a) closure and (b) resolution as a function of p_T^{true} for: before any n_{track} or $\Delta R_{track,avg}$ correction (open circles); a calibration using a network with $\theta = \{n_{track}\}$ (circles); a calibration using a network with $\theta = \{\Delta R_{track,avg}\}$ sequentially after correcting for n_{track} (squares); and a simultaneous calibration using a network with $\theta = \{n_{track}, \Delta R_{track,avg}\}$ (diamonds). For the resolution, also shown is the (negative) improvement in quadrature of the resolution for a given calibration with resolution σ' to the resolution before any correction σ.

9.6 Discussion

The results presented in Section 9.5 demonstrate that the simultaneous neural network-based method introduced in this Chapter improves the performance of the calibration on a variety of metrics.

The search in Chapter 6 involves both quarks and gluons in the final state, and tagging the jets [87–90] may help both with background reduction and with the combinatorial problem of determining the jets decaying from the $\mathfrak{a}$ particle and the ones associated with the VBF production. Figures 9.6c and 9.6d show directly the reductions of non-closures when using the simultaneous method in selections targeting gluons and quarks, respectively. Any nonclosures can reduce the sensitivity of the analysis, in particular in the gluon jets, as the resolution of the $\mathfrak{m}_{jj}$ derived from the gluon jets is a major limitation in the analysis.

In addition, if there is a dependence of the energy response on the underlying features then any mismodeling in the distribution of those features can affect the final energy and introduce systematic uncertainties. The low-mass extension of this search discussed in Appendix B.1 proposes in particular the use of small-R jet substructure to tag the boosted low-mass decays, and so therefore reducing any uncertainties associated with that kind of selection can improve the optimality and sensitivity of the analysis.

The search in Chapter 7 uses only large-R jets, which as mentioned in Section 9.3 do not have any GSC applied. However, large-R jets actually have two relevant parameters with dimensions of energy associated with them - their energy and their mass. The large-R calibration calibrates the energy first and then the mass in sequence, taking into account the energy as a feature. Generalized numerical inversion can apply directly to this second step of the large-R jet calibration in order to improve the mass scale - as this search tags on the masses of the involved large-R jets, any improvement on the mass scale directly improves the tagging capabilities and performance of the search.

The technique outlined here does not address the question of how to simultaneously calibrate multiple features, e.g. the jet energy and mass at the same time.

The Author has contributed to a project [292] in which an extension of generalized numerical inversion is proposed that addresses exactly this question, and it is found that it is possible to calibrate multiple variables simultaneously while taking into account multiple simultaneous auxiliary variables. In particular, this project answers this question in the context of the large-R jet calibration, which could directly further improve the sensitivity of the search in Chapter 7.

One point that is not addressed in the formulation of generalized numerical inversion is the difference between the mean, median, and mode of the target distributions; this difference was expounded on particularly thoroughly in the discussion of numerical inversion itself in Chapter 8. The neural network L in generalized numerical inversion learns whatever the loss function indicates it to learn; e.g., if the loss is a mean squared loss then it learns the mean over the corresponding distribution. The Author has contributed to a project [420] which discusses exactly how to learn the mean, median, or mode of the corresponding distributions with neural networks. This technique can be used to more closely align with the physics goals of the ATLAS jet calibration.

9.7 Conclusion

This Chapter has presented a novel regression-based jet energy calibration procedure implemented with neural networks, called generalized numerical inversion. Generalized numerical inversion generalizes numerical inversion and in particular retains the key property of being independent of the underlying spectrum (Chapter 8). The new method is able to take into account multiple auxiliary jet features simultaneously, which is an improvement over the current sequential method - in particular, correlations in the response between various jet features can be corrected for. The new technique is shown to reduce the response difference between quark and gluon jets by using features associated with the jet provenance, and furthermore is shown to mitigate the effect of mismodeling of the underlying distribution of the features themselves. The simultaneous calibration offers similar benefits to the sequential calibration for the jet energy resolution and by construction has no impact on the

inclusive closure of the jet calibration. These improvements and new extensions of this technique can directly improve the performance of the searches presented in this Thesis (Chapter 6 and Chapter 7) and of searches and measurements in ATLAS in general.

Chapter 10

Conclusions

The unprecedented energies and luminosities provided by the LHC have allowed testing our physical theories in never-before-seen regimes and with remarkable precision. With data from collisions produced by the LHC gathered with the ATLAS detector[1], the final piece of the Standard Model, the Higgs boson, has been confirmed. The Standard Model has therefore withstood all experimental tests of its predictions, and there are no remaining undiscovered particles or forces predicted by the SM.

However, there are good motivations for believing the SM does not tell the full story, and it is almost certain that the LHC will play a major role in uncovering the next chapter. It is therefore vitally important to put the data provided by the LHC to as best use as possible, in order to advance our understanding of nature and lead to the next big discovery.

The use of jets in novel and interesting ways is one crucial way of making best use of LHC data - it is again almost certain that the next big discovery will involve jets in some way.

The original research efforts described in this Thesis contribute to providing means by which this endeavor can be fulfilled. Chapter 6 describes a search with multiple jets in the final state for new particles via their interactions with the Higgs boson. Tests of the precise properties of the Higgs boson like these present a major remaining test for the SM. The search in Chapter 7 uses the underlying features of jets in a novel way

[1] And CMS.

to perform a generic search for new resonant particles. These innovations in analysis techniques will enhance the potential of the LHC to be used to discover new unexpected physics phenomena. Finally, Chapters 8 and 9 present new understandings and methods in the reconstruction of jets. These improvements in the performance of jets enable their best use in measurements and searches with LHC data.

The search for new physics using jets continues, and the progress made in this Thesis will contribute to these efforts now and in the future.

Appendix A

A Search for a Beyond-the-Standard-Model Higgs Decay: Appendix

A.1 Existing Limits

We would like to place a limit on $BR(H \to aa \to \gamma\gamma jj) = 2BR(H \to aa) \times BR(a \to \gamma\gamma)BR(a \to jj)$, given limits on $BR(H \to aa \to 2\gamma 2\gamma)$ and $BR(H \to aa)$.

For brevity, let $B \equiv BR(H \to aa)$, $x \equiv BR(a \to \gamma\gamma)$, and $y \equiv BR(a \to jj)$.

With these conventions, we have that $B < 0.34$ from [320]; we also have that $Bx^2 < 0.001$ from [335]; and we would like to use this information to place a limit on $2Bxy$.

We note first that $y \leq 1 - x$, with the inequality satisfied when the a particle decays exclusively into photons and jets.

The following facts will be useful, which are true for any value of x.

First, for any $s \leq \frac{1}{2}$, if $x < s$ then $x(1-x) < s(1-s)$.

Second, $x(1-x) \leq \frac{1}{4}$.

We do not know the true value of B, so we consider two cases which cover all possibilities: $B < 0.004$, and $0.004 \leq B < 0.34$.

If $B < 0.004$, then:

$$\begin{aligned} 2Bxy &\leq 2Bx(1-x) \\ &< 2 \times 0.004 \times \frac{1}{4} \\ &= 0.002 \end{aligned}$$

If $0.004 \leq B < 0.34$, then:

$$\begin{aligned} Bx^2 &< 0.001 \\ \rightarrow x &< \sqrt{\frac{0.001}{B}} \\ &\leq \frac{1}{2} \\ \rightarrow x(1-x) &< \sqrt{\frac{0.001}{B}} \times \left(1 - \sqrt{\frac{0.001}{B}}\right) \end{aligned}$$

Then we have:

$$\begin{aligned} 2Bxy &\leq 2Bx(1-x) \\ &< 2 \times B \times \sqrt{\frac{0.001}{B}} \times \left(1 - \sqrt{\frac{0.001}{B}}\right) \\ &= 2(\sqrt{0.001B} - 0.001) \\ &< 2(\sqrt{0.001 \times 0.34} - 0.001) \\ &\approx 0.035 \end{aligned}$$

Thus no matter what B is, we have $2Bxy < 0.035$.

A.2 Jet Kinematics

The distributions of m_{jj}, $|m_{jj} - m_{\gamma\gamma}|$, and $|m_{\gamma\gamma jj}$ are shown for $m_a = 20$ GeV (Figure A.1); $m_a = 30$ GeV (Figure A.2); $m_a = 40$ GeV (Figure A.3); $m_a = 50$ GeV

(Figure A.4); and $m_a = 60$ GeV (Figure A.5).

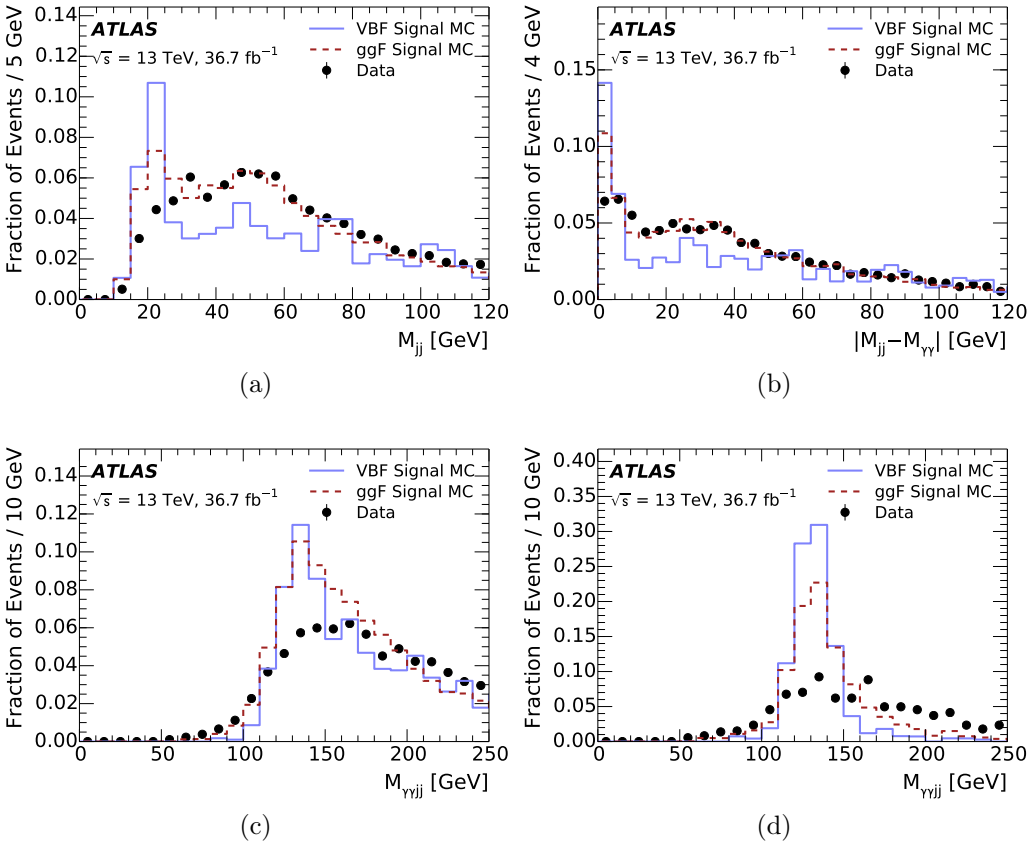


Figure A.1: Distributions of kinematic observables before the requirements on m_{jj}^{VBF}, leading VBF jet p_T, $m_{\gamma\gamma jj}$ and $|m_{jj} - m_{\gamma\gamma}|$ for: (a) m_{jj}; (b) $|m_{jj} - m_{\gamma\gamma}|$; (c) $m_{\gamma\gamma jj}$; and (d) $m_{\gamma\gamma jj}$ (with the additional requirement $|m_{jj} - m_{\gamma\gamma}| < x_R$ that defines the signal-enriched region). The quantities are shown separately for simulated signal events (with $m_a = 20$ GeV) produced in the VBF mode and compared with those produced in the ggF mode and the observed data.

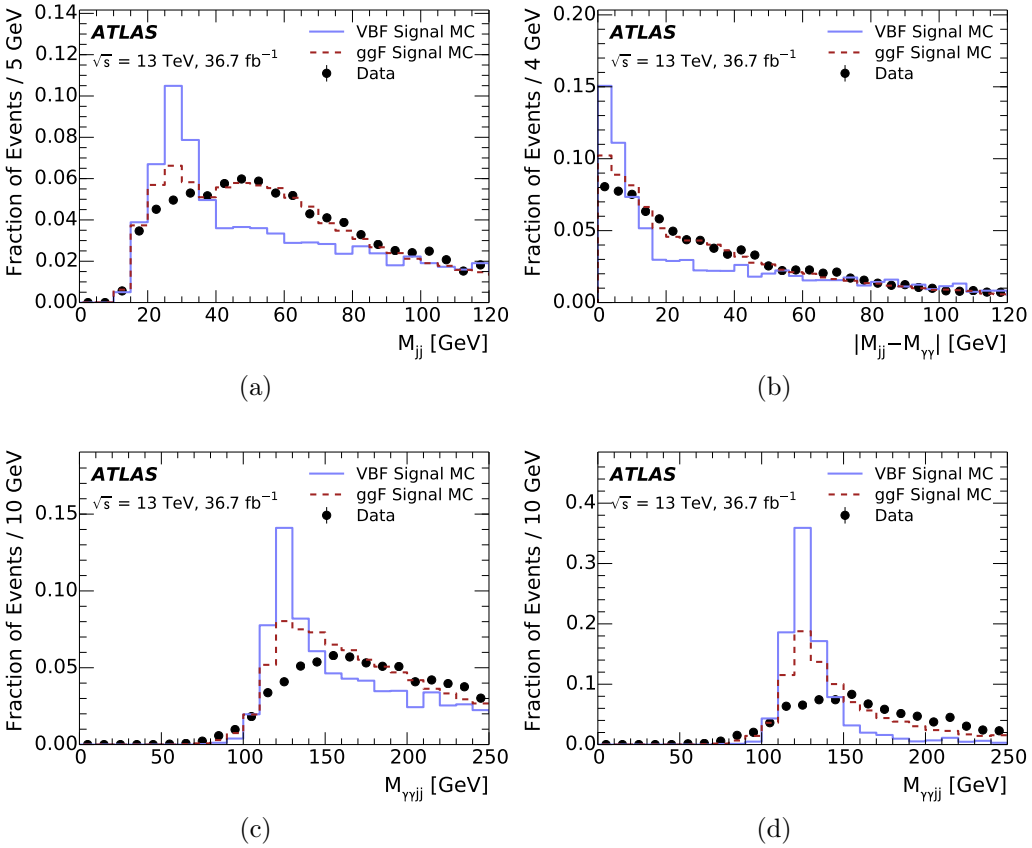


Figure A.2: Distributions of kinematic observables before the requirements on m_{jj}^{VBF}, leading VBF jet p_T, $m_{\gamma\gamma jj}$ and $|m_{jj} - m_{\gamma\gamma}|$ for: (a) m_{jj}; (b) $|m_{jj} - m_{\gamma\gamma}|$; (c) $m_{\gamma\gamma jj}$; and (d) $m_{\gamma\gamma jj}$ (with the additional requirement $|m_{jj} - m_{\gamma\gamma}| < x_R$ that defines the signal-enriched region). The quantities are shown separately for simulated signal events (with $m_a = 30$ GeV) produced in the VBF mode and compared with those produced in the ggF mode and the observed data.

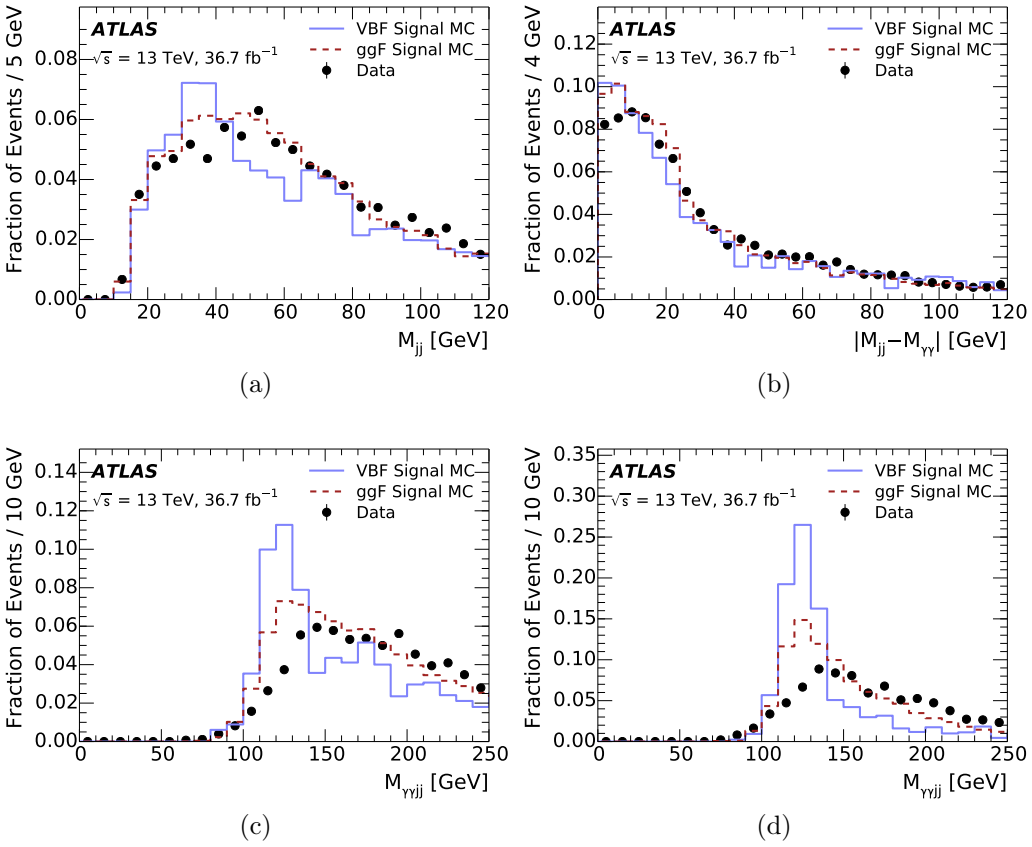


Figure A.3: Distributions of kinematic observables before the requirements on m_{jj}^{VBF}, leading VBF jet p_T, $m_{\gamma\gamma jj}$ and $|m_{jj} - m_{\gamma\gamma}|$ for: (a) m_{jj}; (b) $|m_{jj} - m_{\gamma\gamma}|$; (c) $m_{\gamma\gamma jj}$; and (d) $m_{\gamma\gamma jj}$ (with the additional requirement $|m_{jj} - m_{\gamma\gamma}| < x_R$ that defines the signal-enriched region). The quantities are shown separately for simulated signal events (with $m_a = 40$ GeV) produced in the VBF mode and compared with those produced in the ggF mode and the observed data.

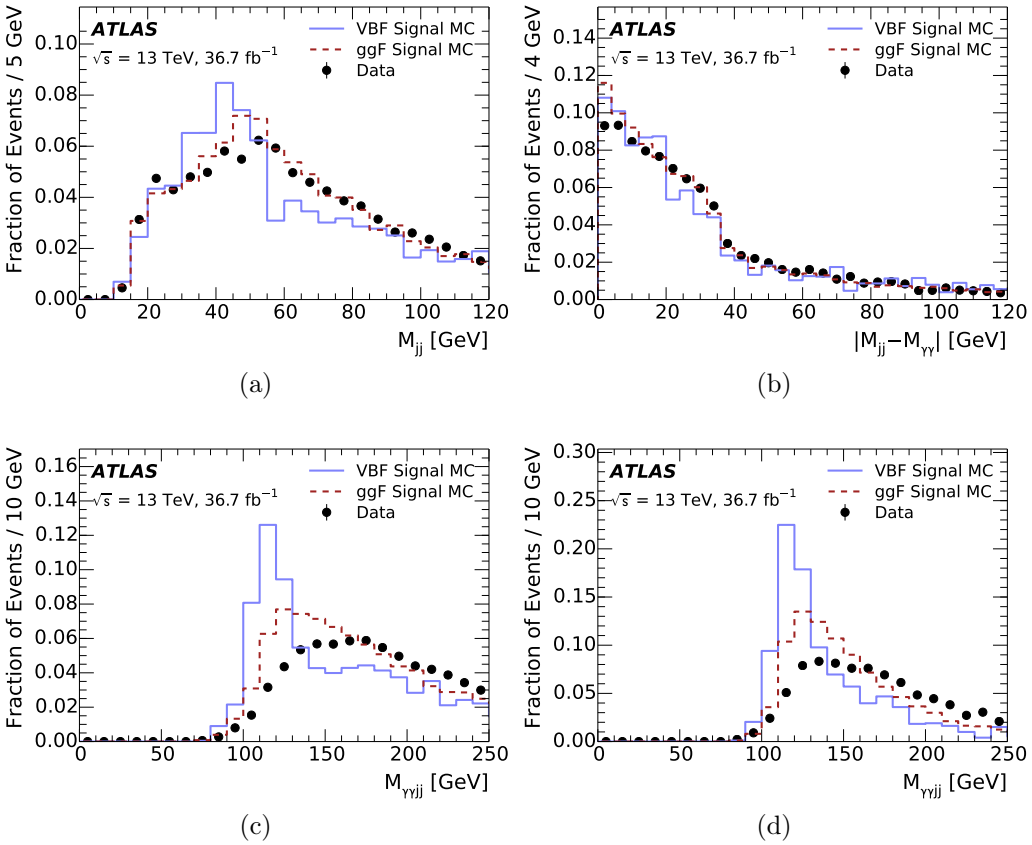


Figure A.4: Distributions of kinematic observables before the requirements on m_{jj}^{VBF}, leading VBF jet p_T, $m_{\gamma\gamma jj}$ and $|m_{jj} - m_{\gamma\gamma}|$ for: (a) m_{jj}; (b) $|m_{jj} - m_{\gamma\gamma}|$; (c) $m_{\gamma\gamma jj}$; and (d) $m_{\gamma\gamma jj}$ (with the additional requirement $|m_{jj} - m_{\gamma\gamma}| < x_R$ that defines the signal-enriched region). The quantities are shown separately for simulated signal events (with $m_a = 50$ GeV) produced in the VBF mode and compared with those produced in the ggF mode and the observed data.

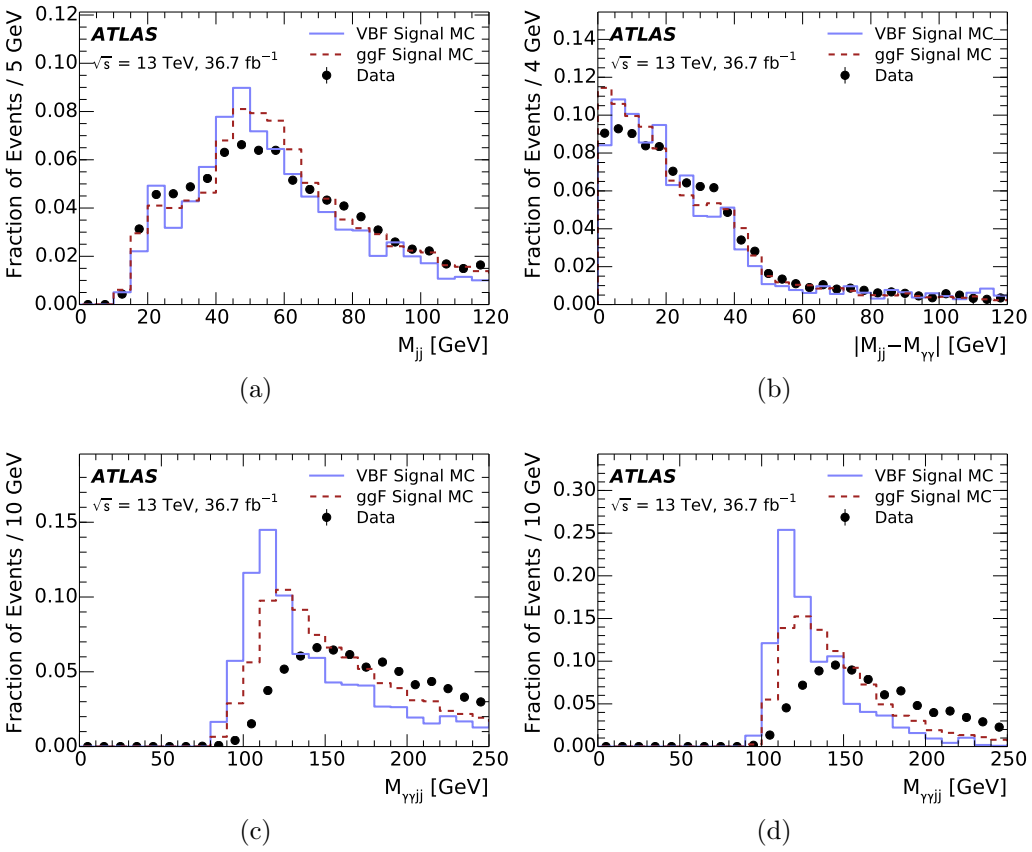


Figure A.5: Distributions of kinematic observables before the requirements on m_{jj}^{VBF}, leading VBF jet p_T, $m_{\gamma\gamma jj}$ and $|m_{jj} - m_{\gamma\gamma}|$ for: (a) m_{jj}; (b) $|m_{jj} - m_{\gamma\gamma}|$; (c) $m_{\gamma\gamma jj}$; and (d) $m_{\gamma\gamma jj}$ (with the additional requirement $|m_{jj} - m_{\gamma\gamma}| < x_R$ that defines the signal-enriched region). The quantities are shown separately for simulated signal events (with $m_a = 60$ GeV) produced in the VBF mode and compared with those produced in the ggF mode and the observed data.

Appendix B

A Search for a Beyond-the-Standard-Model Higgs Decay: Low Mass Feasibility Study

B.1 Low-mass Diphoton Trigger Selection

A study was performed to understand the limitations of the diphoton selection for low mass a particles, in particular $m_a \leq 20$ GeV which is the lowest m_a the search described in 6 has sensitivity to.

The reason the selection has low efficiency on these signals is because the a particle is boosted, causing its decay products to be collimated.

Figure B.1 shows the p_T of the leading and subleading a particles at truth level for some signals in this low mass region. The p_T of the a particle decreases slightly as m_a increases, because the energy from the Higgs decay is shared between the energy and mass of the a particle. The more prevalent effect relates to the opening angle of the decay products of the a. For a two-body decay from a particle with mass m and transverse momentum p_T, the opening angle of the decay products is generically roughly $\Delta R \sim \frac{2m}{p_T}$ [185]. So as m_a gets smaller, the ratio of m_a to the p_T of the a particle, and therefore the opening angle between the decay products, decreases sharply. The VBF events tend to have a broader distribution than the ggF events

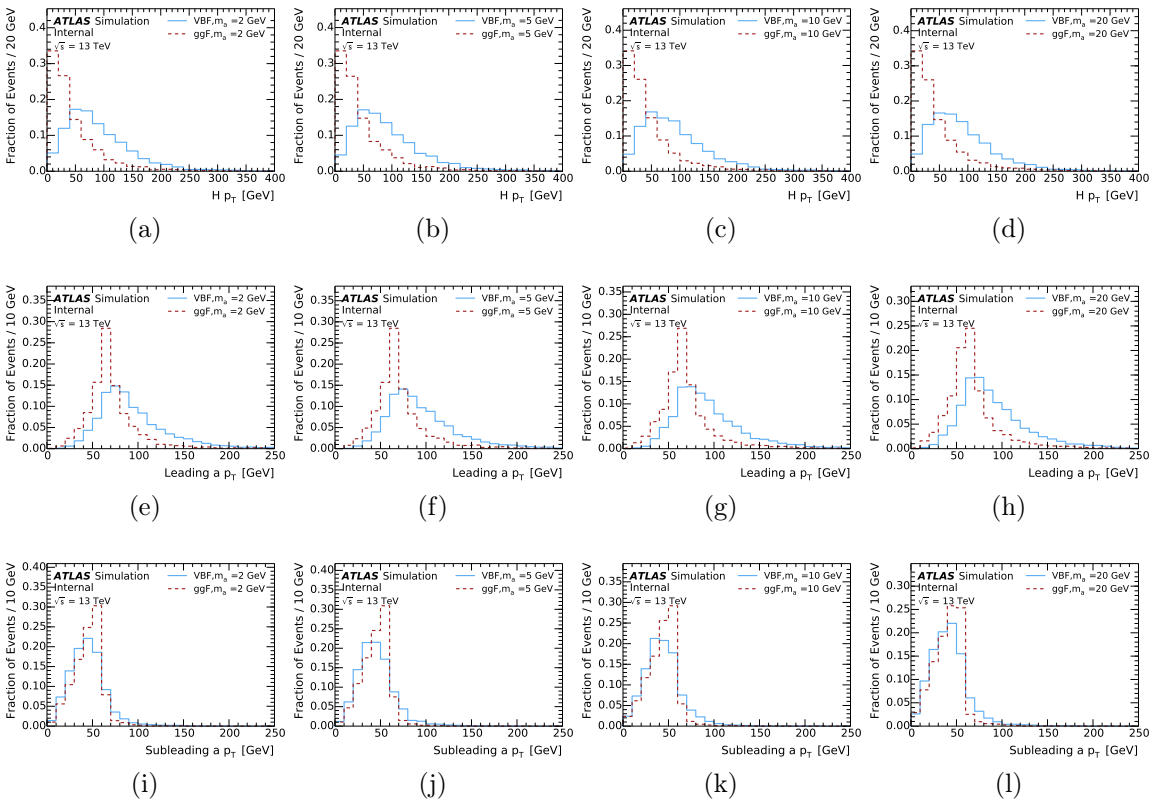


Figure B.1: Distributions of the p_T of the (top) Higgs; (middle) leading a; and (bottom) subleading a particle, at truth level. The quantities are shown for simulated signal events with (first column) $m_a = 2$ GeV; (second column) $m_a = 5$ GeV; (third column) $m_a = 10$ GeV; and (fourth column) $m_a = 20$ GeV. The distributions are shown separately for events produced in the VBF mode and those produced in the ggF mode.

simply because the kinematics of the mother Higgs particle are less constrained in the VBF topology (top row of Figure B.1).

This boosting can be seen in Figure B.2 for the photon decay products of the a and in Figure B.3 for the gluon decay products of the a. While the p_T of the photons and gluon decay products are relatively unaffected by the mass of the a, for low m_a the ΔR between the two decay products gets very small. This dependence of ΔR on m_a affects both the trigger efficiency (Section 6.3.1) and the offline diphoton selection efficiency (Section 6.3.2).

The value $\Delta R = 0.2$ is the distance metric for the photon reconstruction at the trigger level; below this value only one photon region of interest is identified [421,422]. For $m_a < 10$ GeV it is clear that this effect ruins the trigger efficiency, since a large fraction of the events have between the two photons $\Delta R < 0.2$. For $10 < m_a < 20$ GeV the effect is a little more nuanced. Some fraction of the photons do have $\Delta R < 0.2$ and therefore do not pass the trigger requirement. For the photons with $\Delta R > 0.2$, since the opening angle between the photons $\Delta R \sim \frac{2m_a}{p_T}$, the p_T of the originating a particle is $\sim \frac{2m_a}{\Delta R} < 10m_a$. Assuming the p_T of the two decay products are chosen basically uniformly between 0 and the p_T of the a particle, subject to conservation of energy, the p_T of the leading photon is expected to be on average twice the p_T of the subleading photon, and therefore the p_T of the subleading photon is about $\frac{1}{3}$ of the p_T of the a particle; this can be seen in Figure B.2 (and also in Figures B.3 for the gluon decay products). Thus, e.g. for $m_a \sim 10$ GeV and $\Delta R = 0.2$, the p_T of the subleading photon is on average (and therefore half the time less than) 33 GeV, therefore not passing the `HLT_g35_loose_g25_loose` trigger. Of course, for larger ΔR at fixed m_a this effect gets stronger.

In summary, there are two effects that cause the photon trigger efficiency to go down for small m_a. The first is simply that a larger fraction of events have the two photon decay products with $\Delta R < 0.2$. The second is that, for $\Delta R > 0.2$, for fixed m_a as ΔR gets larger the p_T of the a particle and therefore of the photons goes down inversely, and around $m_a = 10$ GeV the subleading photon is around the trigger threshold, causing inefficiencies. The trigger therefore requires m_a to be high enough such that if the p_T of the a particle is high enough for its decay products to pass the trigger threshold, the ΔR between the two photon decay products is also large enough to be reconstructed as two trigger-level photons.

At the offline level, $\Delta R = 0.4$ is the distance metric for the photon isolation [283]; below this value the photons are not reconstructed as isolated (ruining the offline selection efficiency). However, at offline level this isolation can be turned off, and since the triggers used in this analysis have no isolation applied (though such an isolation is possible [422]), the effects of the isolation at least are not limiting the sensitivity of this analysis. Of course, the isolation on the offline photons was kept

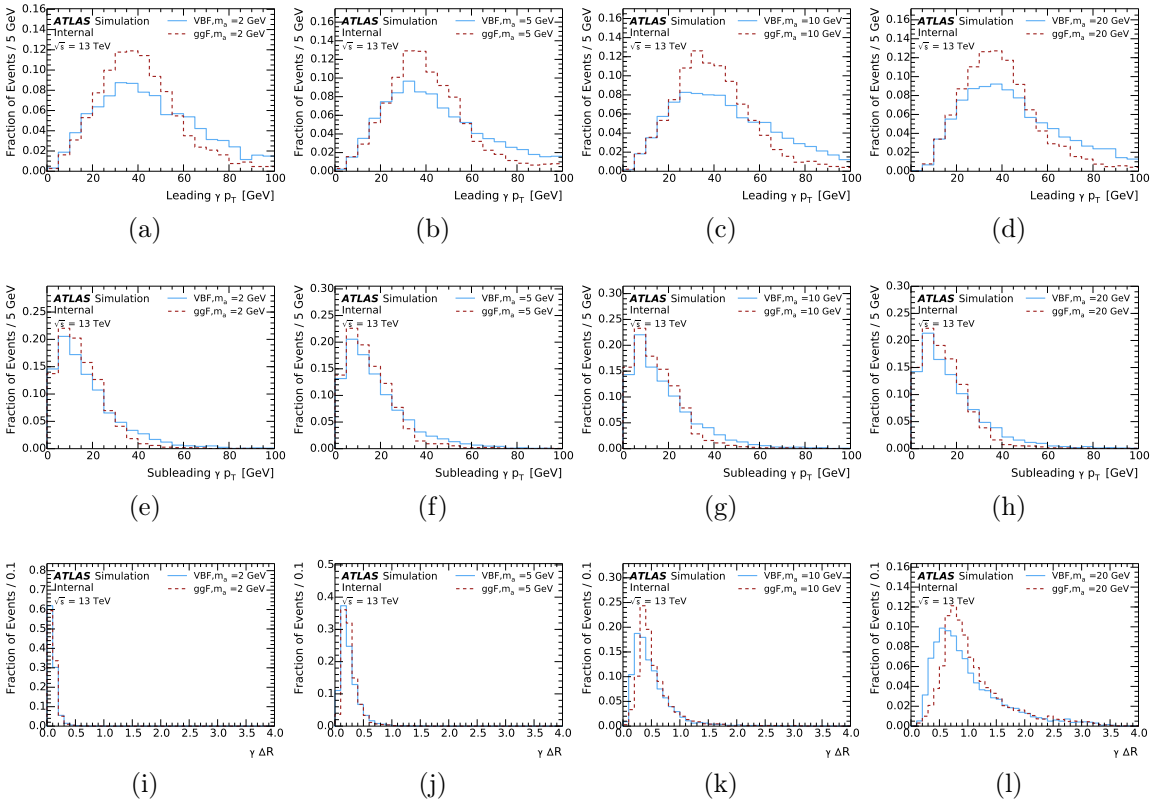


Figure B.2: Distributions of the p_T of the (top) leading photon; (middle) subleading photon; and (bottom) ΔR between the two photons, at truth level. The quantities are shown for simulated signal events with (first column) $m_a = 2$ GeV; (second column) $m_a = 5$ GeV; (third column) $m_a = 10$ GeV; and (fourth column) $m_a = 20$ GeV. The distributions are shown separately for events produced in the VBF mode and those produced in the ggF mode.

for this search; this is due to the trigger selection limiting the efficiency on the low masses, and the higher masses tending to have $\Delta R > 0.4$ so that the isolation has high signal efficiency and contributes to the background rejection.

Finally, the value $\Delta R = 0.4$ is the distance metric for the small-R jet algorithm; below this value both the photons and gluons are reconstructed in a single jet (making the offline selection significantly harder). An offline selection strategy taking advantage of the fact that the decay products are reconstructed in a single jet is discussed in Section B.2.

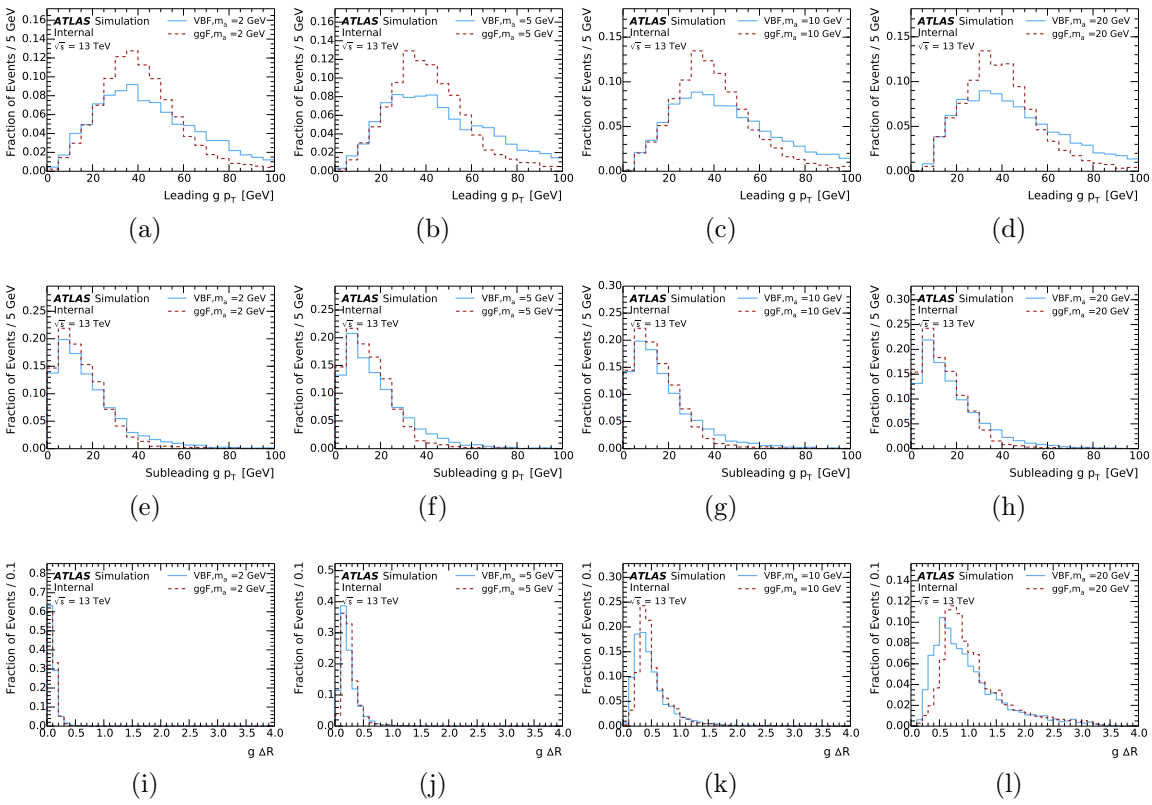


Figure B.3: Distributions of the p_T of the (top) leading signal gluon; (middle) subleading signal gluon; and (bottom) ΔR between the two gluons, at truth level. The quantities are shown for simulated signal events with (first column) $m_a = 2$ GeV; (second column) $m_a = 5$ GeV; (third column) $m_a = 10$ GeV; and (fourth column) $m_a = 20$ GeV. The distributions are shown separately for events produced in the VBF mode and those produced in the ggF mode.

However, there is an existing trigger which can be sensitive to these boosted topologies. This trigger is seeded off the L1 trigger that requires a single isolated photon with $E_T > 22$ GeV. In the high-level trigger, there are three additional requirements: a photon with $E_T > 25$ GeV; at least four jets with $E_T > 35$ GeV and $|\eta| < 4.9$; and at least one dijet pair with invariant mass greater than 1000 GeV. In ATLAS nomenclature, this trigger is called `HLT_g25_medium_4j35_0eta490_invm1000_L1_EM22VHI`. That this trigger has efficiency on our signals is an accident; it was originally intended

for background studies for the VBF $H \to bb$ + photon analysis [423]. However, because of this other analysis need, this trigger was running unprescaled in 2015-2018 data-taking periods [263–266, 424].

This trigger has potential to be efficient on our signals at low m_a. First, the photon isolation at L1 only goes out to $\Delta R = 0.2$ [422]; and furthermore, no isolation at all is applied if the p_T of the photon is > 50 GeV. As argued above, for fixed m_a as ΔR between the photons gets smaller the p_T of the leading photon gets larger, so that these two effects can mostly cancel out and allow our signals to pass the L1 trigger. The top row of Figure B.4 shows the efficiency of this trigger as a function of the leading offline photon p_T; the efficiency goes to 100% for $p_T > 30$ GeV for $m_a \geq 10$ GeV, indicating that the leading photon is passing the trigger requirement. For $m_a = 2$ GeV, the efficiency only goes to 100% for $p_T > 50$ GeV, in accord with the isolation requirements mentioned above.

Second, for the VBF signals the decay products of each of the two a particles are reconstructed as a single jet in the HLT (photons are reconstructed as jets both at trigger level and offline); these two jets plus the two VBF jets reach the requirement of 4 jets with $p_T > 35$ GeV, and in particular the p_T of each of the signal jets contains the full p_T of each of the a particles rather than just one of the decay products. The middle row of Figure B.4 shows the efficiency of this trigger as a function of the 4th-leading offline jet p_T (including jets overlapping photons); the efficiency goes to 100% for $p_T > 40$ GeV, indicating that the photons are being picked up as jets in the HLT.

Finally, the photon in the HLT requirement is met by the leading photon, since no isolation is required at HLT; and the invariant mass requirement is of course met in the VBF topology. The bottom row of Figure B.4 shows the efficiency of this trigger as a function of the maximum dijet invariant mass; the efficiency goes to 100% for invariant mass > 1100 GeV, indicating the level at which the turn-on occurs.

In summary, with an offline selection of at least one photon with $p_T > 30$ GeV, at least 4 jets (including any that overlap photons) with $p_T > 40$ GeV, and a maximum invariant mass of > 1100 GeV, the `HLT_g25_medium_4j35_0eta490_invm1000` trigger is fully efficiency. This offline selection is listed in Table B.1.

Table B.1: Preselection for low m_a analysis.

	Selection
Trigger	`HLT_g25_medium_4j35_0eta490_invm1000`
Photon Selection	≥ 1 photon with $p_T > 30$ GeV
Jet Selection	≥ 4 jets (not overlap-removed) with $p_T > 40$ GeV
Invariant Mass	≥ 1 pair of jets with invariant mass > 1100 GeV

The efficiency of this offline selection on some given values of m_a are shown in Table B.2. These efficiencies are pretty good, considering the very tight invariant

Table B.2: Efficiency of preselection for low m_a analysis.

m_a [GeV]	Efficiency
2	0.022
10	0.045
20	0.041

mass cut. For comparison, for the high m_a analysis, the efficiency through the trigger and VBF selection is roughly 0.02 (Figure 6.9).

B.2 Low-mass Diphoton Offline Selection

After the preselection detailed in Table B.1, there are additional selections offline that can be made to further distinguish the low mass signal from the background. As mentioned, no isolation should be required on the photons offline, in order to maintain the efficiency on the diphotons with $\Delta R < 0.4$; this increases the background presence, in particular hadronically-produced neutral pions decaying to two photons. For this separation, the decay products of the a particle end up in a single jet, so the signature

of this signal would be 2 jets each with a substructure of two hard subprongs. The photons-jet would look quite similar to the neutral pion decay mentioned above, but the gluons-jet would not look like any SM background process.

Jet substructure has a long history in high energy physics [69, 86]; its use can be traced back to LEP [188–191], Tevatron [192, 193], and HERA [194–196]. At the LHC, jet substructure for small-R jets focuses on distinguishing quark-initiated from gluon-initiated jets at ATLAS [87–90] and CMS [197–200]. Jets with two prongs are a signature of W-jet decays, and so most work with this signature target large-R jets intended to tag W decays at ATLAS [92, 94] and CMS [425]. There are no existing taggers that target small-R jets with two hard prongs.

The Author has advised on a project intended to identify collimated photon decays reconstructed in a single jet [426]. The tagging algorithm uses convolutional neural networks [315–317] to tag jet images [407,408], which have been used in ATLAS before for quark-gluon discrimination [90]. The tagging performance indicates a greatly improved signal/background discrimination power, enabling the offline selection for this analysis at low mass. These techniques could also be used to target the collimated gluon decays.

One drawback of using the direct jet image-based approach outlined in that project is the reliance on the simulation of the jet substructure in the signal and background used for training. The weak supervision analysis technique used in Chapter 7 could be useful as an alternative to this direct approach in order to derive a purely data-driven tagger. In particular, the masses of the two small-R jets and their substructure features can be used as features for tagging in the classification without labels, which is particularly useful since the masses m_a are unknown. The overall invariant mass of the two collimated jets can be used as the binning and bump hunt variable, with particular attention to the region around m_H. Such an approach would not only target these novel signatures but also both standard model processes like $H \to bb$ (if necessary, an anti-b tag could be applied in order to remove this standard model "background") and more exotic processes where the X, $m_X \neq m_H$ is the mother particle of the two a particles.

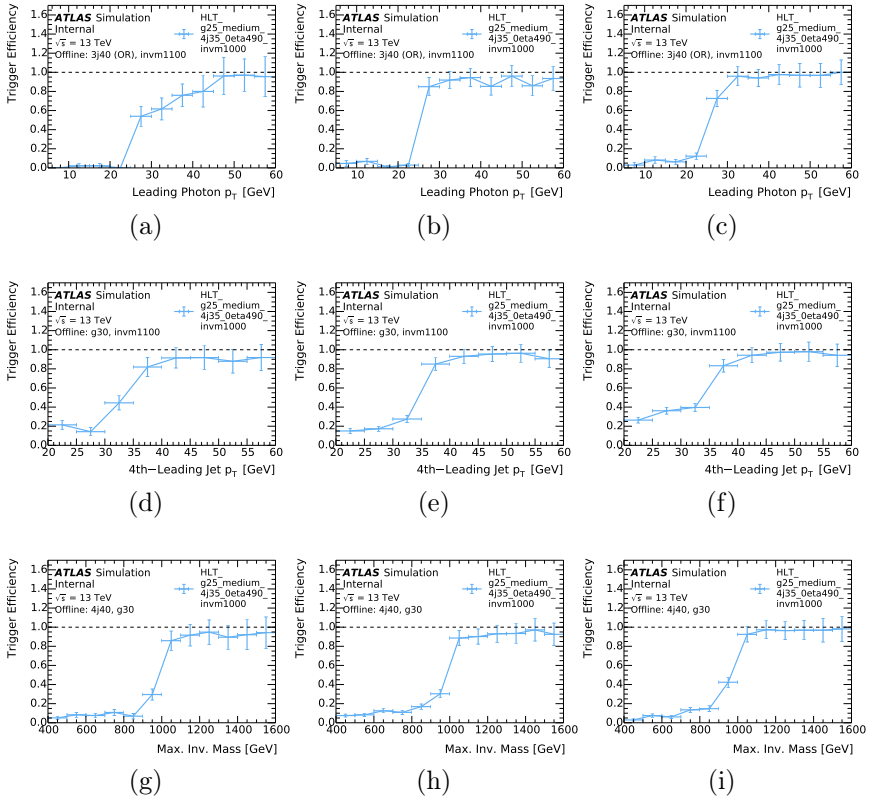


Figure B.4: Efficiency of the `HLT_g25_medium_4j35_0eta490_invm1000` trigger on VBF signal events with (left) $m_a = 2$ GeV; (middle) $m_a = 10$ GeV; and (right) $m_a = 20$ GeV. The efficiency is shown as a function of (top) the leading offline photon p_T; (middle) the 4th-leading offline jet including those overlapping photons; and (bottom) maximum invariant mass between any pair of offline jets. For each efficiency plot, all the other offline requirements shown in Table B.1 are included, so that the efficiency reaches 100% for large values.

Appendix C

A Generic Data-Driven Resonance Search with Weak Supervision: Appendix

C.1 Trials Factors and CWoLa

The search presented in Chapter 7 uses Classification Without Labels (CWoLa) to be generically sensitive to a broad class of signal models, meaning that it does not specify a signal model ahead of time, but is able to tag signals based on features that discriminate them from the background and therefore be sensitive to whatever signal happens to be in the data[1]. In the context of the search presented in Chapter 7, the features are the masses of the jets m_1, m_2[2]; it should also be kept in mind that the natural extension of this search would use a suite of jet substructure variables so that the feature space is high-dimensional.

There are two other ways to be generically sensitive to a broad class of signal models. The first approach is to not place any selections on the features at all - this approach could be called the "Inclusive Search" approach. The analog of the Inclusive Search in the context of the search in Chapter 7 would be the inclusive dijet search [19].

[1] Subject to the constraints of the method and setup, as discussed in Section 7.3.1.

[2] Here we are considering the search in a fixed m_{JJ} bin.

A second way to be generically sensitive is to have a dedicated search in every region of the feature space - this could be referred to as the "Direct Scanning" approach. The analog of the Direct Scanning approach would be a series of searches along the lines of the all-hadronic diboson resonance search [20][3], which targets a region around $(m_1, m_2) = (80, 80)$ GeV - in addition to that search, there could be another one targeting a region around $(m_1, m_2) = (200, 200)$ GeV, $(m_1, m_2) = (80, 400)$ GeV, etc.

It is crucially important to point out that CWoLa hunting differs fundamentally from these alternative approaches, in that it can be both more sensitive than the Inclusive Search, and in that it avoids suffering from a *trials factor*, also known as a *look-elsewhere effect* [12, 427–429], that the Direct Scanning approach would be subject to. This effect accounts for the fact that, with many statistically independent searches, there is a high probability of observing a large excess in at least one of the searches in the background-only case. Therefore, the size of the excess needed to claim a discovery in any one search needs to be increased in order to lower the probability of a false discovery [430–432], ultimately reducing sensitivity to true signals. The goal of this Section is to explain this effect in detail and how CWoLa hunting manages to avoid this effect, while still improving sensitivity compared to an inclusive search.

Suppose we collect a sample of N data points $\mathbf{x} = \{x_1, ..., x_N\}$, where each observation x_i could be multi-dimensional corresponding to $n \geq 1$ features in each observation[4]. $\mathbf{x}$ can be considered to be a sample from the random variable $\mathbf{X}$ which we observed by taking data. We construct a *test statistic* [107, 433–438] $Q(\mathbf{X})$ which is some mapping from the data to a real number, with the property that larger values of Q correspond to increasing tension with the null or background-only hypothesis H_0. The observed test statistic is $q = Q(\mathbf{x})$, and we define the observed p-value

$$p \equiv \text{Prob}(Q \geq q|H_0). \tag{C.1}$$

[3] In fact that search uses features other than just the jet mass to tag signals; including these other features and varying them with independent direct searches would be the analog of the natural extension of the search in Chapter 7 mentioned above.

[4] E.g., $x_i = (m_1^i, m_2^i)$ for the ith event, in which case $n = 2$.

A small p-value therefore indicates incompatibility with the background-only hypothesis, and a possible discovery of a new signal.

The discovery procedure is to specify a threshold value α such that, if the observed $p \leq \alpha$, a discovery is claimed. If the p-value is defined correctly, it should be the case that

$$\text{Prob}(p \leq \alpha | H_0) = \alpha, \tag{C.2}$$

i.e. the probability of making a Type I error (incorrectly rejecting the null hypothesis) is exactly α. If the probability is $< \alpha$ then the test is conservative, meaning that the probability of a Type I error is lower than expected; but if the probability is $> \alpha$ then the probability of a Type I error is higher than expected and the p-values should be recalibrated to account for this.

α is often mapped to a *significance* $Z = \Phi^{-1}(1-\alpha)$, where Φ is the cumulative distribution function (CDF) of the normal distribution.

In particle physics, since the null hypothesis is the Standard Model, α is typically set very low - the "5σ Standard" [439, 440] corresponds to $Z = 5$, or $\alpha \sim 3 \times 10^{-7}$, so that the probability of falsely ruling out the SM is very low.

We set up a toy model to demonstrate the background-only p-values and signal discovery potential using the various generic signal searches outlined above. In the toy model, the observed data are one-dimensional, $X \in [0,1]$[5]. Two samples will be considered: a background-only case (B only) and a background plus signal case ($B+S$). In the background-only case, there are expected B background events[6] with values drawn from a uniform distribution, $X_B \sim U(0,1)$. In the background plus signal case, there are in addition expected S signal events drawn from some very small but unknown region in this space, $X_S \sim \mathcal{N}(y_S, \delta)$, with $\delta \ll 1$[7]. In these examples S has

[5]The toy methods outlined below will find small regions in this space of size ϵ, corresponding to line segments $[y - \frac{\epsilon}{2}, y + \frac{\epsilon}{2}]$. The multidimensional analog would be balls or hypercubes of volume ϵ in a space where the n features have been mapped to $[0,1]^n$.

[6]I.e., the number of background events is Poisson-distributed with mean B.

[7]In n dimensions, the volume taken up by the signal in $[0,1]^n$ space would decrease quickly with n, so this assumption is realistic. Another way to say this is that, as the number of features increases, a selection with fixed signal efficiency can have a background efficiency $\delta \ll 1$.

been set to $\sqrt{B}$, i.e. a 1σ excess over the background-only hypothesis overall.

The first approach is the Inclusive Search, in which the test statistic Q is the total number of events observed. In the background-only hypothesis H_0, $Q \sim \text{Poiss}(B)$, with the p-values defined according to the Poisson CDF. Note that, in this case and in all cases outlined below, B is estimated from some external information; e.g. in the search in Chapter 7 B is estimated via the m_{JJ} fit interpolated in the m_{JJ} region being studied. Figure C.1 shows the probability distributions of the p-values. The background-only case has $\text{Prob}(p \leq \alpha|H_0) = \alpha$ as expected, and the background plus signal case has probability $> \alpha$, indicating tension with the background-only hypothesis. In particular, the probability of observing a $Z = 1$ significance is 50%, which makes sense since the injected signal corresponds to a 1σ excess. As a point of reference, the probability of observing a $Z = 2$ significance is ~ 0.16 and the probability of observing a $Z = 3$ significance is ~ 0.02 with this amount of signal.

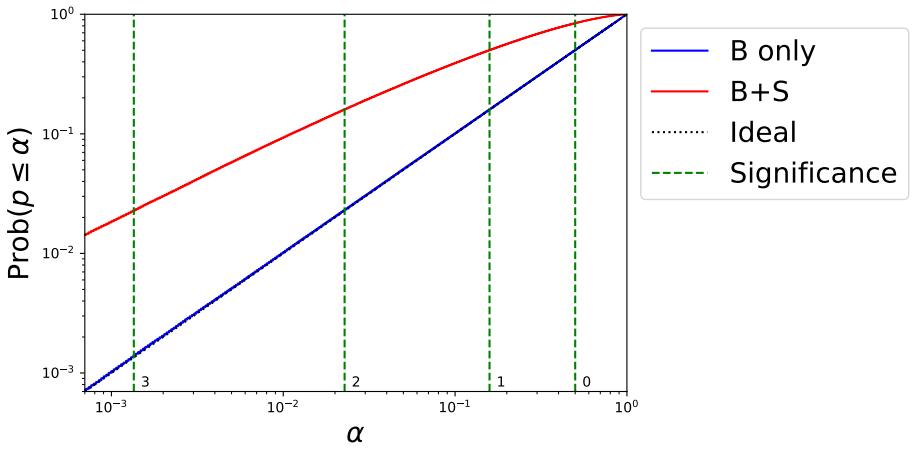


Figure C.1: Probability distributions of p-values for the Inclusive Search in the (blue) background-only and (red) background plus signal cases. The black dotted line indicates the ideal background-only case of $\text{Prob}(p \leq \alpha|H_0) = \alpha$. Then green dashed lines indicate the α thresholds corresponding to integer significances Z.

The second approach is the Direct Scanning method. In this approach, the $[0, 1]$ space is partitioned into $m = 10$ regions of size $\epsilon = 0.1$: $[0, \epsilon, 2\epsilon, ..., 1]$. In each region a signal discovery test is performed, corresponding to discrete searches covering all signal model hypotheses, with Q the number of events in the region and $Q \sim \text{Poiss}(\epsilon B)$

the background-only hypothesis. The overall p-value (also known as the *global* p-value) is taken as the minimum p-value over all the regions (i.e., $p \leq \alpha$ means that at least one of these searches claimed a discovery with significance $Z(\alpha)$). Figure C.2 shows the probability distributions of the p-values. The background-only case has $\text{Prob}(p \leq \alpha|H_0) > \alpha$, since there is a high probability that at least one of the tests has a low p-value. This effect is exactly what is referred to as a trials factor or look-elsewhere effect. In principle this global p-value can be corrected analytically if the number of tests is fixed [430–432] or calibrated by simulating the entire procedure in some sample intended to produce the background-only p-value distribution. In this example, the analytic correction would require $\alpha \to \hat{\alpha} \approx \frac{\alpha}{m}$ [8] , for the same significance level; in the simulation it can also be seen that $\text{Prob}(p \leq \hat{\alpha}|H_0) \approx \alpha$. However requiring this more stringent test reduces the signal sensitivity - for example, the probability of observing a $Z = 2$ significance ($\alpha \approx 0.02 \to \hat{\alpha} \approx 0.002$) is approximately 0.6, as opposed to ~ 0.9 as the naive p-values would indicate. This is better than for the Inclusive Search, since the true signal does lie in one of the regions and there is some chance that all the other regions have high p-values. However, we will see the improvement is worse than the improvement that can be made with CWoLa.

In any case, there are good reasons to avoid having to re-calibrate the background-only p-values: most often the disparate searches will not be so clear-cut as to allow an analytic correction; and a simulation of the background-only p-values is often hard to come by. Because of this, it's desirable to have a method where the background-only p-values are defined correctly.

We turn now to the ideas that enable CWoLa to be more sensitive than the Inclusive Search without biasing the background-only p-values. The first thing to point out is that the trials factor can be removed by splitting the data randomly into a train set comprising $(1-f)$ of the data and a test set comprising f of the data. In the *train* set the Direct Scanning method is performed to find the region with minimum p-value, but the test statistic Q is then defined to be the number of events in the

[8] The more precise correction is actually $\alpha = 1 - (1 - \hat{\alpha})^m \approx m\hat{\alpha}$, which can easily be derived as $\text{Prob}(p \leq \hat{\alpha}|H_0) = 1 - \bigcup_{i=1}^{m} \text{Prob}(p_i > \hat{\alpha}|H_0) = 1 - (1 - \hat{\alpha})^m$.

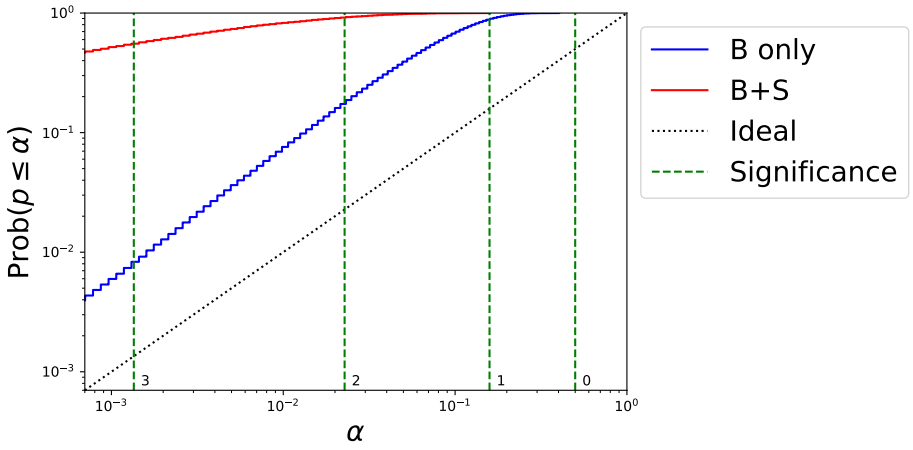


Figure C.2: Probability distributions of p-values for the Direct Scanning approach in the (blue) background-only and (red) background plus signal cases. The black dotted line indicates the ideal background-only case of $\text{Prob}(p \leq \alpha|H_0) = \alpha$. Then green dashed lines indicate the α thresholds corresponding to integer significances Z.

test set in the region indicated by the train set, which under the background-only hypothesis should be distributed as $Q \sim \text{Poiss}(f\epsilon B)$. Since the train and tests sets are statistically independent in the background-only case, just because there's an excess in the train set that does not imply that there will be an excess in the test set, and the background-only p-values should be unbiased.

Figure C.3 shows the probability distributions of the p-values with a $f = \frac{1}{2}$ train/test split. The background-only case has $\text{Prob}(p \leq \alpha|H_0) \leq \alpha$[9]. The signal sensitivity is better than the Inclusive Search, with the probability of a $Z = 2$ excess about 0.4, and the probability of a $Z = 3$ excess about 0.15. However, there are two factors which could be addressed to make the signal sensitivity better. First, since $(1 - f) < 1$, it is harder to find the region in which the signal lays than without the train/test split. Second, since $f < 1$, even in cases where the correct region is found, the significance of the excess in that region is reduced by a factor of $\sim \sqrt{f}$[10].

Both of these factors can be addressed by having a k-fold cross-testing. In this

[9]There is a small effect that, given there is an excess in the train set in one of the regions, there are fewer events to go around for all the other regions (train and test), so Q is actually biased low.

[10]The significance in the correct region is roughly $\frac{S}{\sqrt{\epsilon B}}$. If both S and B are multiplied by a factor of $f < 1$, the significance is multiplied by $\sqrt{f} < 1$.

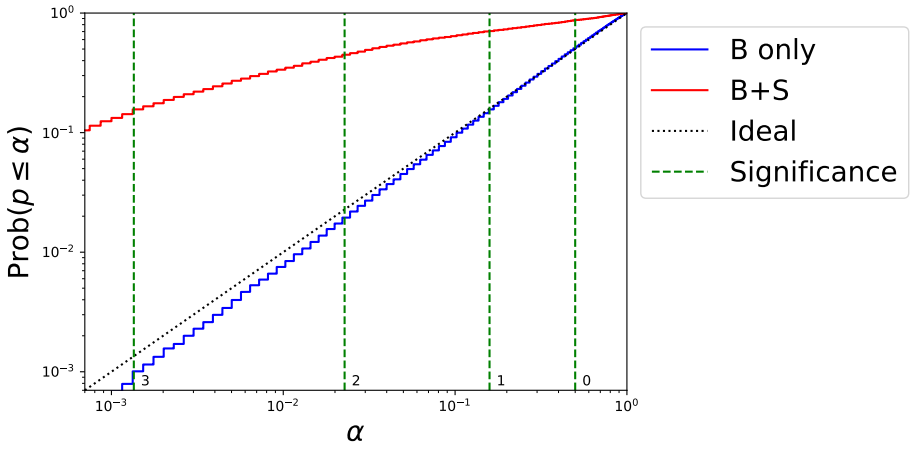


Figure C.3: Probability distributions of p-values for the Direct Scanning approach with a $f = \frac{1}{2}$ train/test split in the (blue) background-only and (red) background plus signal cases. The black dotted line indicates the ideal background-only case of $\text{Prob}(p \leq \alpha|H_0) = \alpha$. Then green dashed lines indicate the α thresholds corresponding to integer significances Z.

method, the data are split randomly into k sets. One of the k sets is set as the test set, comprising $f = \frac{1}{k}$ of the data, and the remaining $1 - f = \frac{k-1}{k}$ of the data is set as the train set. The Direct Scanning method with the train/test split described above is then performed. Then, the whole process is repeated with each of the k sets designated as the test set in turn. The test statistic Q is the sum across the k test sets of the number of events in the chosen region from the respective train set, which in the background-only hypothesis is distributed as $Q \sim \text{Poiss}(\epsilon B)$. Both of the factors mentioned above are addressed, as the train set comprises $\frac{k-1}{k} \sim 1$ of the data, and the sum of the test set sizes is not reduced by a factor of f.

Figure C.4 shows the probability distributions of the p-values with a $k = 5$ k-fold cross-testing. Unfortunately, the background-only p-values are biased, with $\text{Prob}(p \leq \alpha|H_0) > \alpha$. This is due to the following effect. In cases where there is a large (fake) excess in one of the regions, even when splitting into the k sets, there will still be a large excess in each of the k train sets in that region. This means that each of the k trainings will choose the same region to tag in the test set, which is the region in which there is a large excess overall. Interestingly, this effect goes away when

there are no large excesses in any of the regions, which can be seen in Figure C.4 as $\text{Prob}(p \leq \alpha|H_0) \leq \alpha$ for $\alpha \gtrsim 0.3$. This effect is made more clear in the case that $k \gg 1$ - in that case, the training is basically the same every time, so that the same region is always chosen for the test set.

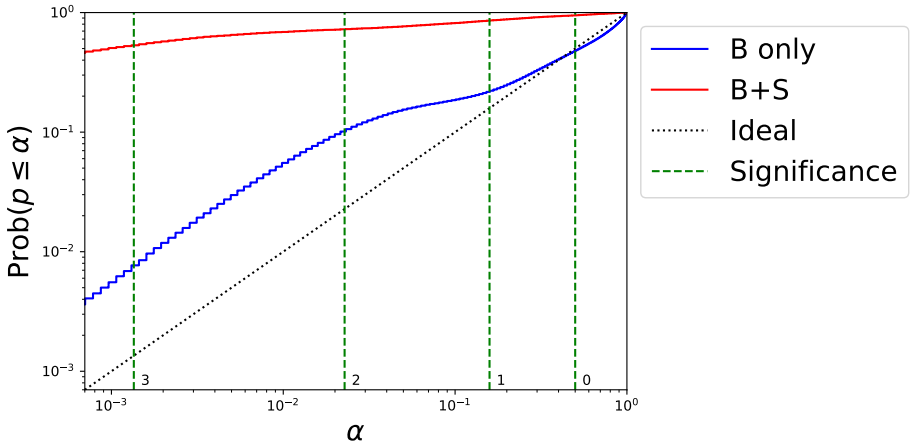


Figure C.4: Probability distributions of p-values with a $k = 5$ k-fold cross-testing in the (blue) background-only and (red) background plus signal cases. The black dotted line indicates the ideal background-only case of $\text{Prob}(p \leq \alpha|H_0) = \alpha$. Then green dashed lines indicate the α thresholds corresponding to integer significances Z.

CWoLa hunting makes use of the above ideas with k-fold cross-testing. However, in addition, CWoLa uses a neural network (NN) to learn what region to tag in the train set before tagging in the test set. The NN features a key difference to the Direct Scanning method outlined above in that the NN score has certain regularization requirements which lead to smoothness in the NN output [391–396][11]. The NN attempts to approximate the likelihood distribution of X [441] with a smooth function, by finding small regions $\epsilon' < \epsilon$ with overdensities, assuming that corresponds to a local maximum of the probability distribution, and smoothly interpolating from there. In the toy model, this behavior is encapsulated by finding the region with length $\epsilon' = \frac{\epsilon}{10} = 0.01$ that has the highest density in the train set, and then designating the

[11] In the search in Chapter 7, there is also a validation set which serves to prevent overtraining the network, and the scores from multiple networks across different validation permutations are averaged together, further smoothing the output.

region of size ϵ with the overdensity at its center as the tagging region for the test set.

To demonstrate the advantages of the NN over the Direct Scanning method, this mock NN is demonstrated **without k-fold cross-testing**, but rather training and testing on the same set. The test statistic Q is the number of events observed in the tagging region, which in the background-only hypothesis should be distributed as $Q \sim \text{Poiss}(\epsilon B)$.

The probability distributions of the p-values with this mock NN are shown in Figure C.5. Despite training and testing on the same set, as in the original Direct Scanning approach (Figure C.2), the background-only p-values already look better, albeit still biased (since there is an overdensity in the tagged region). In addition, the p-values in the background plus signal case are smaller (higher probability that $p \leq \alpha$ for fixed α) than in the Direct Scanning approach, indicating that if the background-only p-values were calibrated, the signal sensitivity would be better.

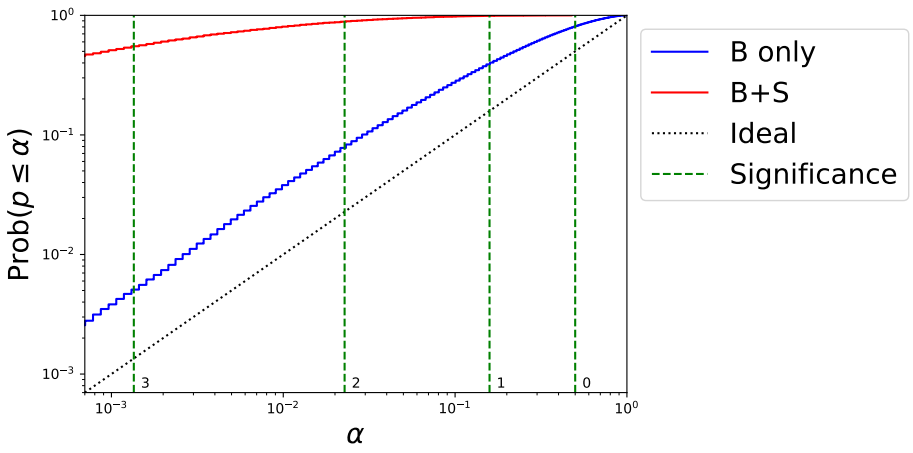


Figure C.5: Probability distributions of p-values with a mock NN with no k-fold cross-testing in the (blue) background-only and (red) background plus signal cases. The black dotted line indicates the ideal background-only case of $\text{Prob}(p \leq \alpha | H_0) = \alpha$. Then green dashed lines indicate the α thresholds corresponding to integer significances Z.

The CWoLa hunting method makes use of both a NN and a k-fold cross-testing ($k = 5$), combining the advantages of both innovations over the Direct Scanning

method. The test statistic Q is the number of events observed in the tagging region summed across the test sets, which in the background-only hypothesis should be distributed as $Q \sim \text{Poiss}(\epsilon B)$.

The probability distributions of the p-values can be seen in Figure C.6. It can be seen that the background-only p-values are approximately unbiased, with $\text{Prob}(p \leq \alpha|H_0) \approx \alpha$ across the whole range[12]. In addition, the background plus signal p-values are significantly smaller than in the Inclusive Search - the probability of a $Z = 2$ excess is about 0.9, and the probability of a $Z = 3$ excess is greater than 50%. This indicates that the signal sensitivity is significantly better than in the Inclusive Search, without paying a penalty of higher Type I errors (or recalibrating the p-values in order to set the rate of Type I errors correctly).

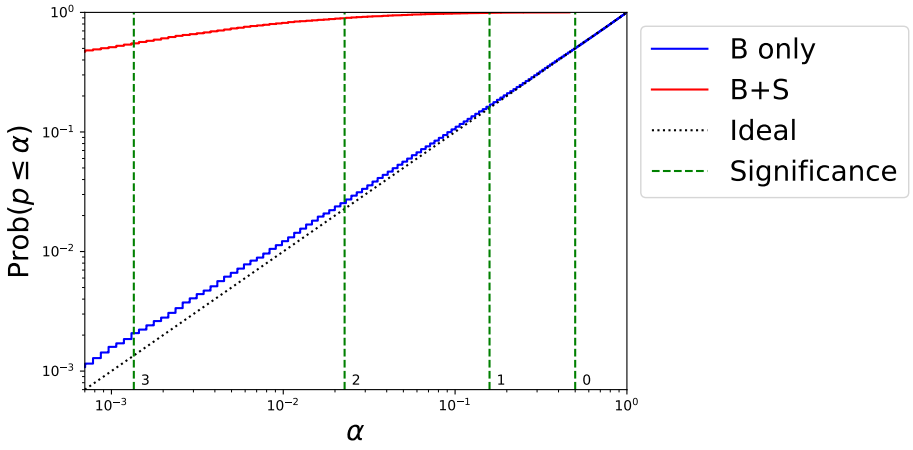


Figure C.6: Probability distributions of p-values with a mock NN and with $k = 5$ k-fold cross-testing in the (blue) background-only and (red) background plus signal cases. The black dotted line indicates the ideal background-only case of $\text{Prob}(p \leq \alpha|H_0) = \alpha$. Then green dashed lines indicate the α thresholds corresponding to integer significances Z.

It should be emphasized that the location of the signal y_S was never specified, and so the NN is able to find the signal regardless of where it lies in the feature space. Thus, the CWoLa hunting method is generically sensitive without paying a large trials factor, as advertised.

[12]There is a small bias towards smaller p-values, but correcting for this would not significantly change the signal sensitivity.

In this toy model the size of the signal δ was set to be smaller than ϵ' in order to demonstrate how CWoLa hunting can find signals without biasing the background-only p-values. However, this is not necessary in general, as long as the presence of the signal causes a significant overdensity in a small region, which the NN can tag according to its regularization requirements.

In addition, the above discussion points out that the k-fold cross-testing is necessary for the CWoLa hunting method to work as expected, but the use of a regularized NN is equally important, which is maybe not as obvious. The various complicated steps of the NN training (Section 7.4.4) serve not only to prevent overtraining the network but also to smooth the NN output.

Finally, the above discussion is simply a toy model intended to demonstrate how CWoLa hunting can work, but not necessarily that it does, which depends on the NN architecture. Ultimately the background-only p-values have to be validated as an empirical test of the setup. These tests are included in the various steps of the validation (Sections 7.5 and 7.6).

C.2 Classification Without Labels Optimality Proof

The proceeding discussion and proof follows [1]. In a classification problem, there is a sample of events from a signal process S and a sample of events from a background process B, and the goal is to determine based on observable features which process a given event comes from. Let $\mathbf{x}$ be a set of features which are useful for discriminating signal and background, and let $p_S(\mathbf{x})$ and $p_B(\mathbf{x})$ be the probability distributions of $\mathbf{x}$ for signal and background events, respectively. A classifier $h : \mathbf{x} \to \mathcal{R}$ is designed to distinguish between signal and background events. For a given selection $h > c$, the efficiency on the signal is $\epsilon_S = \int d\mathbf{x} p_S(\mathbf{x}) \Theta(h(\mathbf{x}) - c)$, and the efficiency on the background is $\epsilon_B = \int d\mathbf{x} p_B(\mathbf{x}) \Theta(h(\mathbf{x}) - c)$, with Θ the Heaviside step function. It is clear that a monotonic rescaling of h corresponds only to a monotonic rescaling of the threshold c, and that the performance of the classifier is determined entirely by the function $\epsilon_B(\epsilon_S)$, also known as the receiver operating characteristic (ROC) curve [442].

A classifier h is considered to be *optimal* if for all other classifier h', $\epsilon_B^{h'}(\epsilon_S) \geq \epsilon_B^h(\epsilon_S)$. By the Neyman-Pearson Lemma [441], an optimal classifer is the likelihood ratio $h^{\text{optimal}} = p_S(\mathbf{x})/p_B(\mathbf{x})$. The goal is therefore to derive a classifier that is the likelihood ratio, up to monotonic rescalings.

In a fully supervised learning problem, the labels of data in the training set are known, and a classifier is trained to distinguish between the signal and background samples. With a suitable loss function, a flexible enough classifier architecture, and enough training data, the derived classifier approximates the optimal likelihood ratio classifier.

In Classification Without Labels (CWoLa), there are only two mixed samples, M_1 and M_2, with signal fractions f_1 and f_2 ($f_1 > f_2$), respectively, which are unknown. The key insight of the CWoLa method is that the optimal classifier for distinguishing M_1 and M_2 is the same as the optimal classifier for distinguishing S and B. That is to say, training a classifier to distinguish M_1 and M_2 will, in the asymptotic limit, result in an optimal classifier for distinguishing signal and background.

The proof is as follows. The likelihood distributions of the mixed samples are

$$p_{M_1}(\mathbf{x}) = f_1 p_S(\mathbf{x}) + (1 - f_1) p_B(\mathbf{x}), \tag{C.3}$$

$$p_{M_2}(\mathbf{x}) = f_2 p_S(\mathbf{x}) + (1 - f_2) p_B(\mathbf{x}). \tag{C.4}$$

The optimal classifier for distinguishing M_1 and M_2 is the likelihood ratio $L_{M_1/M_2} = p_{M_1}/p_{M_2}$, and the optimal classifier for distinguishing S and B is the likelihood ratio $L_{S/B} = p_S/p_B$.

We see that

$$L_{M_1/M_2} = \frac{p_{M_1}}{p_{M_2}} = \frac{f_1 p_S + (1 - f_1) p_B}{f_2 p_S + (1 - f_2) p_B} = \frac{f_1 L_{S/B} + (1 - f_1)}{f_2 L_{S/B} + (1 - f_2)}. \tag{C.5}$$

Then, with $f_1 > f_2$, we have

$$\frac{\partial}{\partial L_{S/B}} L_{M_1/M_2} = \frac{f_1 - f_2}{(f_2 L_{S/B} + (1 - f_2))^2} > 0, \tag{C.6}$$

so that L_{M_1/M_2} is a monotonic rescaling of $L_{S/B}$[13]. Therefore, L_{M_1/M_2} and $L_{S/B}$ define the same optimal classifier. □

C.3 Alternate Ideas for Validation

Section 7.3.3 discusses the difficulties present in this analysis for constructing a validation sample with low signal efficiency to test the analysis pipeline. The primary validation sample used in this analysis is formed by inverting the delta rapidity cut. Some other ideas for forming a validation sample are below, which could be useful for a future analysis based on classification without labels, e.g. one that does not want to impose a delta rapidity cut or one targeting a different topology entirely.

C.3.1 Swapped Dataset

One possibility for a validation sample which is derived from data, and still retains the property that the signal contamination is low, is called the *jet swapped* dataset. In this dataset, leading and subleading jets are swapped between random pairs of events. It is expected that this sample has low signal contamination, because it is unlikely that two random events chosen from the original sample will both be signal events. In particular, suppose that there is some number of signal, S, and some number of background, B, in the original sample with no cuts. We can suppose that the signal fraction $\frac{S}{B} = p \ll 1$, due to existing limits from the inclusive dijet search [19] (note that the limits from previous searches set a stronger bound, that the significance $\frac{S}{\sqrt{B}} = p\sqrt{B} \lesssim 1$). Then the signal contamination in the swapped dataset is expected to be on the order of $p^2 \ll p$ (and therefore with significance $p^2\sqrt{B} \ll 1$).

The m_{JJ} distribution in the swapped dataset is different than in the original sample, and some correlations between the ensemble of jet features and the m_{JJ} value will not be preserved. However, any correlations between the individual jet features and the jet p_T will be preserved, and thus some significant part of the correlations between the individual jet features and the m_{JJ} value. The swapped dataset therefore

[13] With $f_1 < f_2$, the classifier is simply reversed.

serves as an entirely data-derived reasonable proxy for the background spectrum of the features and of m_{JJ}, and can be used as a validation region for testing the validity of the background shape fit, for demonstrating that the learned features do not sculpt the m_{JJ} distribution and therefore violate Assumption 2, and for demonstrating the sensitivity of the method to (unswapped) injected signal.

C.3.2 Anti-tagged Dataset

After learning the neural network scores to distinguish events in the signal region from events in the sideband regions, cuts are placed at some efficiency $\epsilon < 1$ in order to find the most signal-like events. That is to say, every event i has a score $0 \leq S_i \leq 1$ based on the neural network output, with larger S_i indicating an event more signal-like, and moreover the scores are scaled such that a cut $S_i > 1 - \epsilon$ has efficiency exactly ϵ in the signal region bin.

There is a concept of anti-tagging in this framework: by applying a cut $S_i < \epsilon$, the ϵ fraction of events that are the least signal-like are chosen. However, this anti-tagged dataset may still be contaminated by signal, in particular if the true signal is actually mostly in the sideband region rather than the signal region i.e. Case 4b; in this case the scores S_i are signal anti-taggers in the first place, and so the anti-tagged dataset is in fact anti-anti-tagging signal, i.e. positively tagging signal.

C.3.3 Median Dataset

Another possibility is to consider the *median dataset*, where the cut that is placed is $|S_i - 0.5| < \frac{\epsilon}{2}$. These are the set of events that the neural network has decided it is agnostic about being in the signal region or the sideband regions; it is therefore expected that this dataset has little signal contamination, because regardless of whether the true signal was in the signal or sideband regions, it would not end up with this median score. It is not expected that the m_{JJ} distribution will be exactly the same in the median dataset as in the signal-tagged dataset, even in the case there is no true signal, since there are some residual correlations between the features and m_{JJ}. However, the median dataset can be used to test the validity of the background

shape fit, and for looking at jet kinematics in a blinded way.

C.4 Analysis Software

The `mc16-xAOD-ntuple-maker` (https://gitlab.cern.ch/acukierm/mc16-xAOD-ntuple-maker) package is used to create ntuples used for the analysis. The machine learning code can be found in the `cwola-hunting-learning` (https://gitlab.cern.ch/cwola-hunting/cwola-hunting-learning) package.

C.5 Fitting Software

The statistical analysis uses the `ResonanceFinder` package from the DBL group. This code is based on RooFit [443], RooStats [444], and HistFactory [445]. Significant modifications for the background fitting were made for this analysis, which can be found in the dedicated branch (https://gitlab.cern.ch/cwola-hunting/fitting/tree/minimal).

C.6 Bin Offset Test

A study of the effect of the bin positioning relative to the signal center m_A is shown in Figure C.7. A fixed signal is injected with $m_A = 3000$ GeV, so that the m_{JJ} distribution of the signal lies mostly in signal region 5. For this study only, the signal region bins are shifted in units of 0.25 the current bin size, so that after 4 shifts the bins are exactly the same as before with the numbering changed by 1. The NN efficiency on the signal at $\epsilon = 0.1$ is shown as a function of the center of signal region 5, when the signal region used for training is 4, 5, and 6. It can be seen that, regardless of where the bin definitions are, there is some signal region for which the NN efficiency is high (> 0.4). Importantly, the NN maintains very high signal efficiency (> 0.8) in 3 out of the 4 bin positions, and dips lower in only the final 1 out of 4 positions. This indicates that if the signal m_A lies in roughly 75% of the m_{JJ} kinematic space, the NN performance is unaffected, while in the remaining 25% of the space the NN can

still learn to tag the signal albeit at lower efficiency; regardless, CWoLa is sensitive to these new signals.

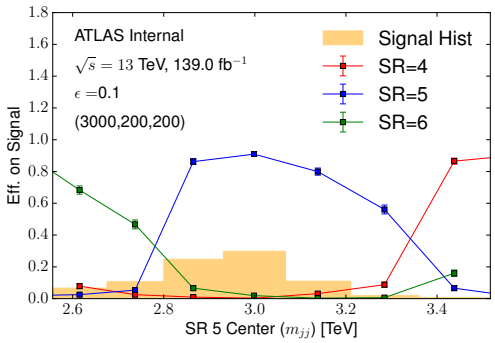


Figure C.7: The efficiency of the NN on the signal (m_A, m_B, m_C = (3000, 200, 200) GeV) at cut $\epsilon = 0.1$ at different values of the center of signal region 5. The efficiency of the NN is shown when training on signal region 4 (in red); 5 (in blue); and 6 (in green). The distribution of m_{JJ} in the signal is also shown (in orange) for reference.

C.7 Validation Analysis: No Signal Fits

The fit on the validation data with no signal injected is shown for signal region 6 (Figure C.8); signal region 7 (Figure C.9); signal region 8 (Figure C.10); and signal region 9 (Figure C.11). These fits generally indicate problems with the $\epsilon = 1.0$ and $\epsilon = 0.25$ fits, especially at lower masses, which motivates the decision to limit the analysis to only $\epsilon = 0.1$ and $\epsilon = 0.01$ for the full unblinded analysis.

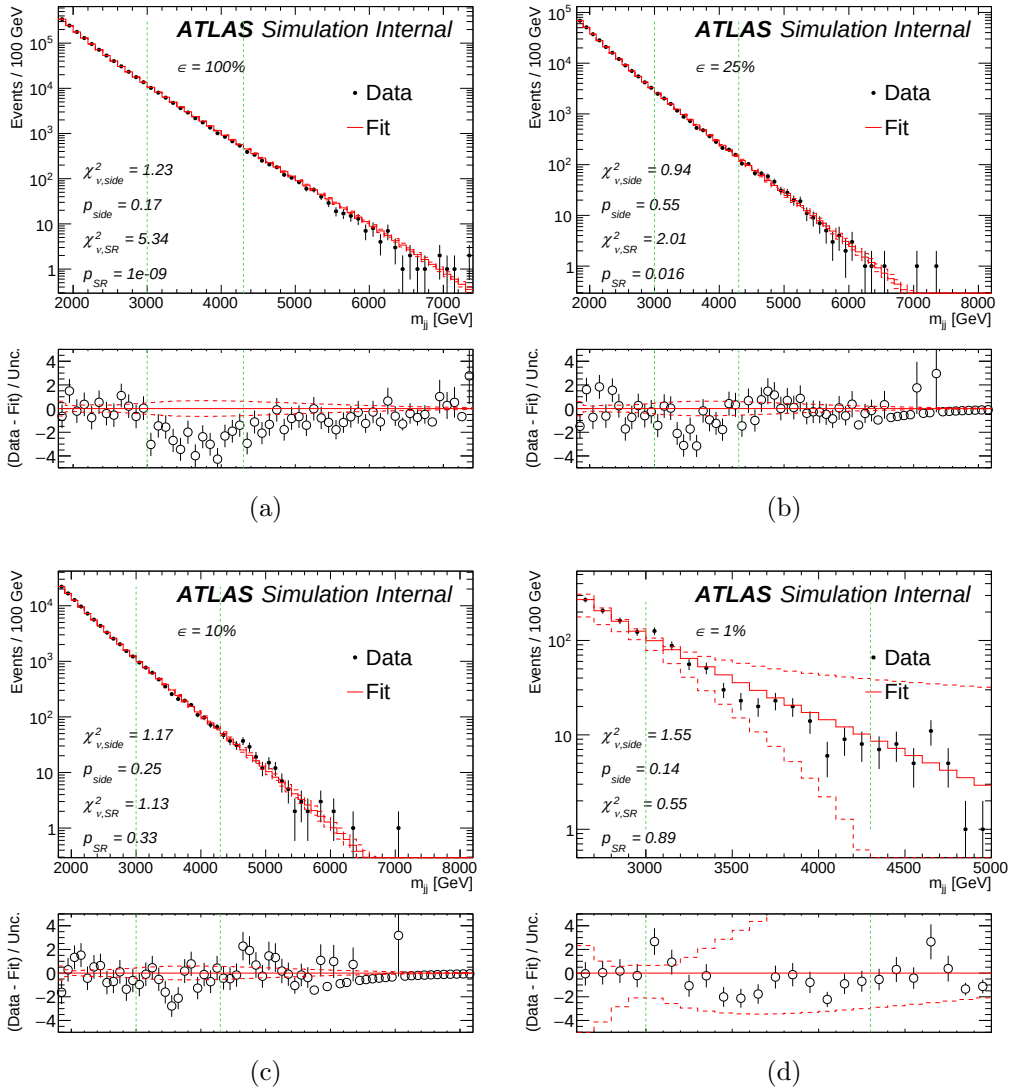


Figure C.8: The background fit when there is no injected signal, in signal region 6, for various efficiency points ϵ. Note that these are data and not simulation, using the inverted rapidity cut data selection.

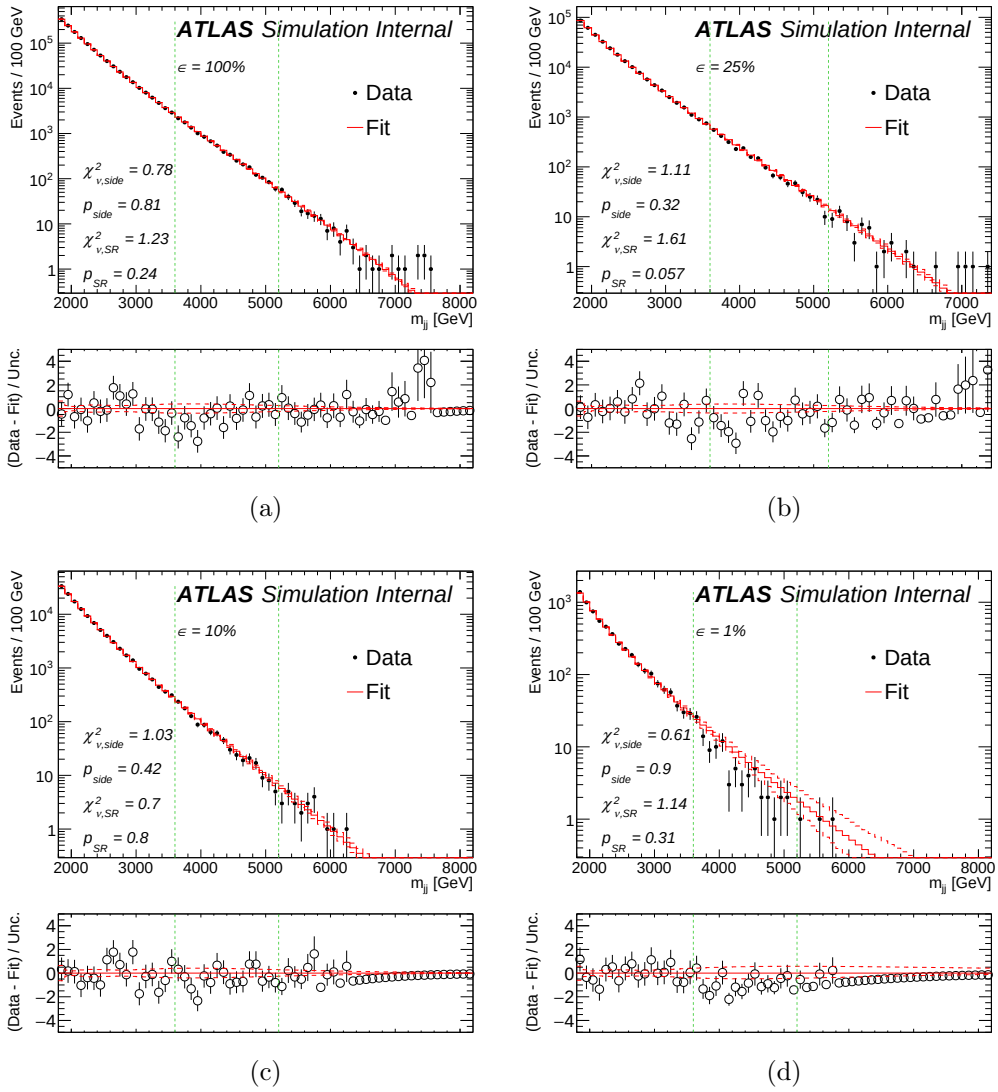


Figure C.9: The background fit when there is no injected signal, in signal region 7, for various efficiency points ϵ. Note that these are data and not simulation, using the inverted rapidity cut data selection.

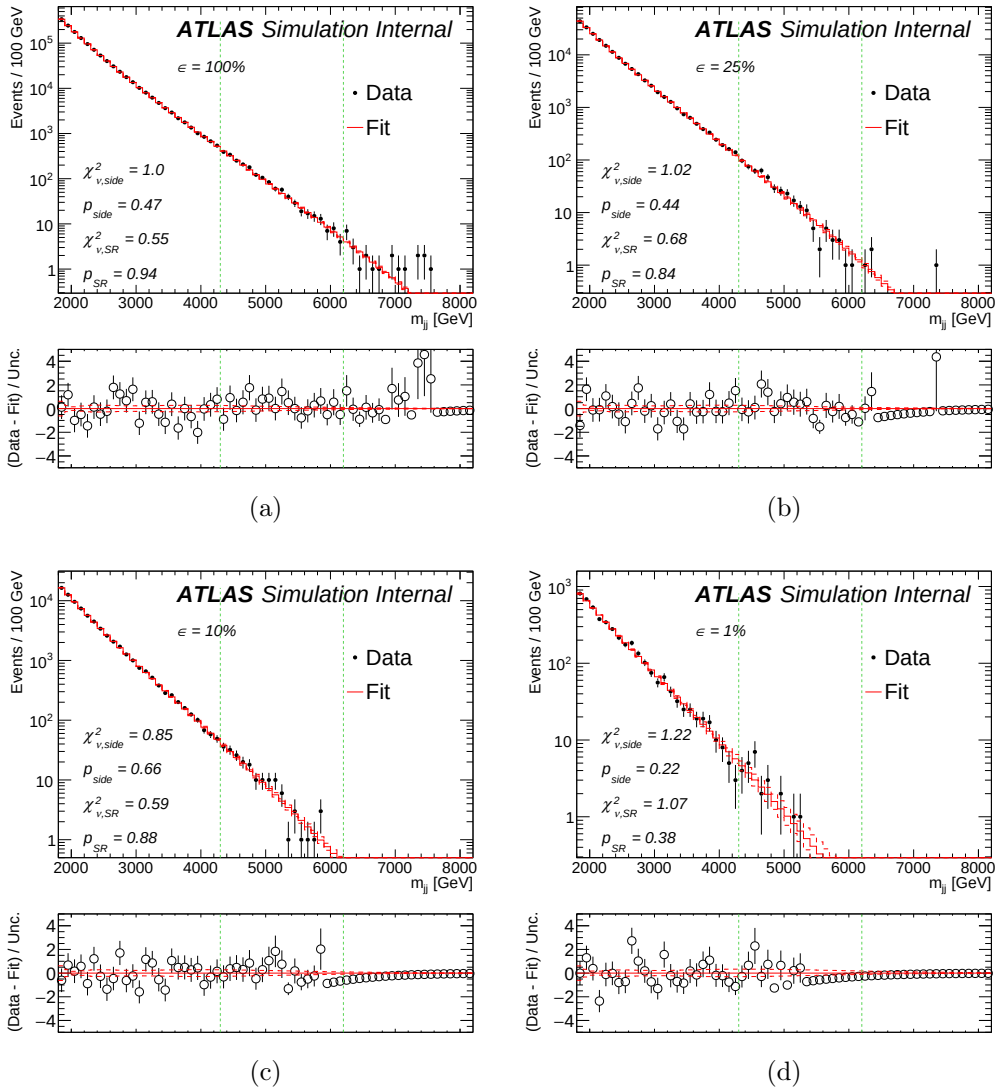


Figure C.10: The background fit when there is no injected signal, in signal region 8, for various efficiency points ϵ. Note that these are data and not simulation, using the inverted rapidity cut data selection.

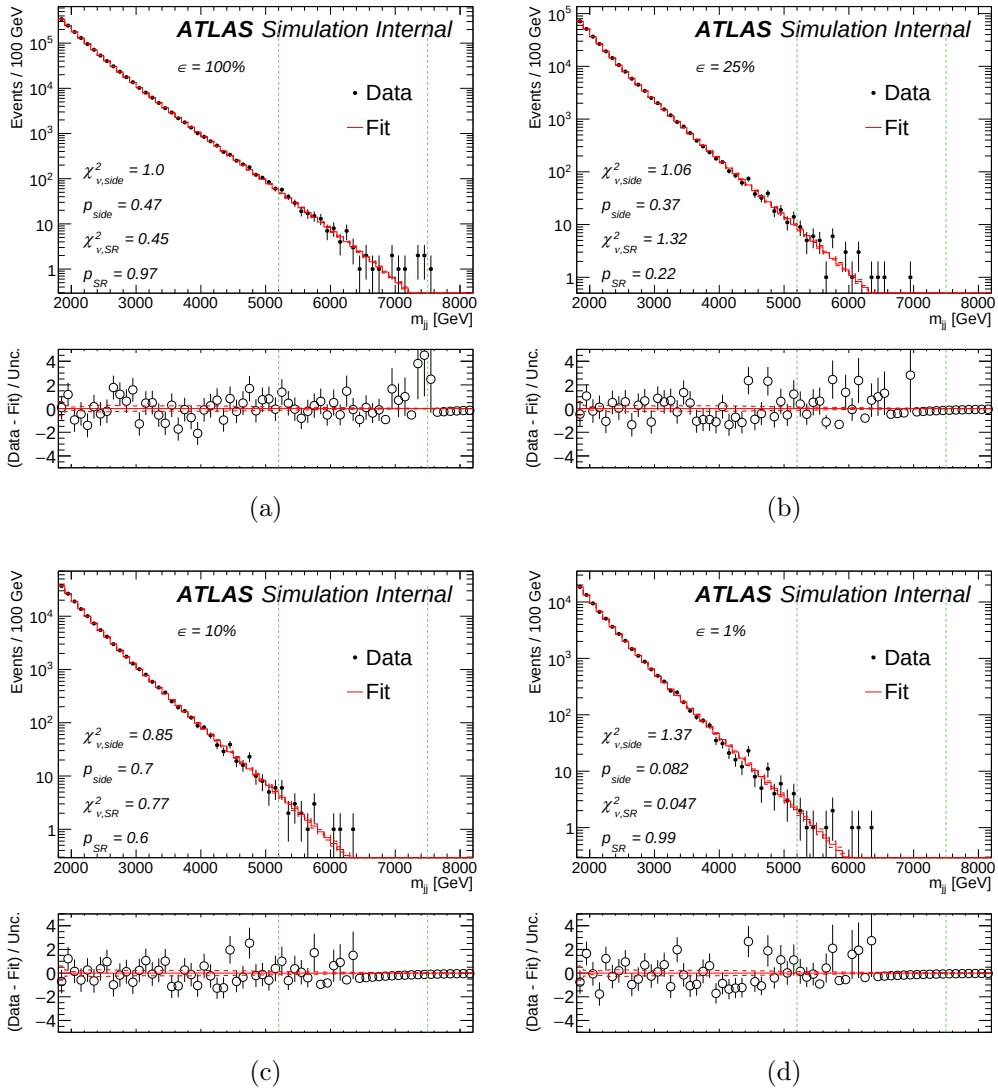


Figure C.11: The background fit when there is no injected signal, in signal region 9, for various efficiency points ϵ. Note that these are data and not simulation, using the inverted rapidity cut data selection.

C.8 Unblinded Analysis: Signal Injection Tests

The effect of an injected signal on the fit is studied. Figure C.12 shows the dependence of the maximum likelihood signal strength $\hat{\mu}$ on the injected signal strength $\mu = \mathcal{L} \times \sigma$, where $\mathcal{L}$ is the data luminosity and σ is the cross section for the production of the given signal. Also indicated is the final 95% CL exclusion limit on the given signal. It can be seen that the fitted signal strength is not consistent with the injected signal strength for values below the 95% CL exclusion limit; above that value there is some bias towards smaller values due to the fit process outlined in Section 7.7.4.

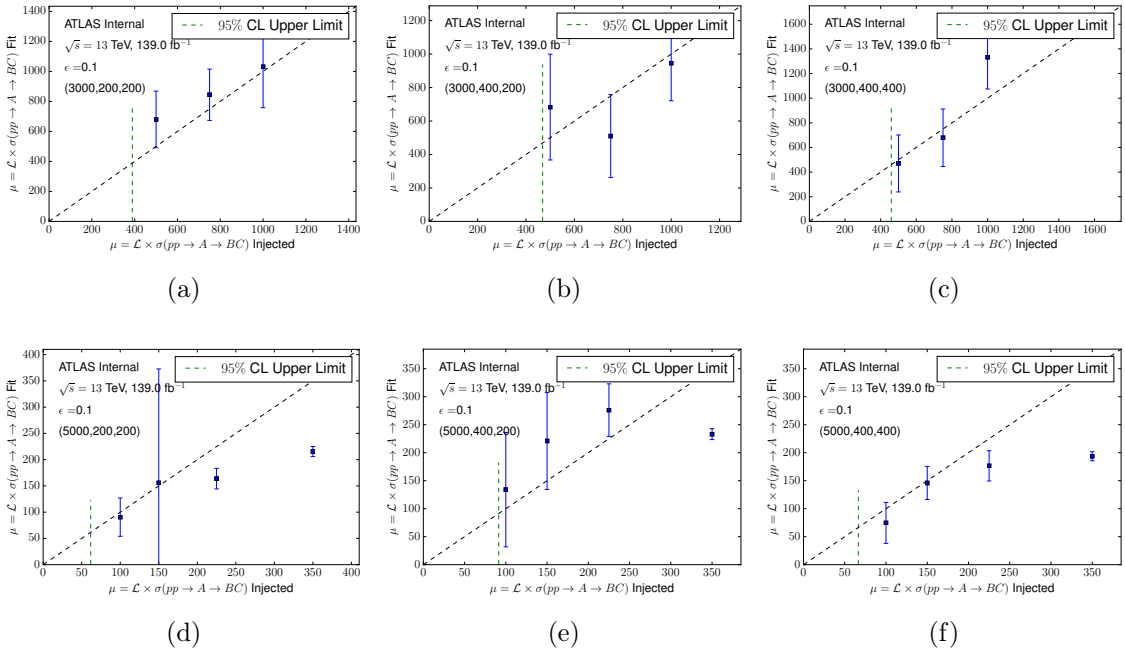


Figure C.12: The dependence of the fitted signal strength $\hat{\mu}$ on the injected signal strength μ with a NN cut at efficiency $\epsilon = 0.1$ trained on (a,b,c) signal region 5 and (d,e,f) signal region 8 for a signal with m_A, m_B, m_C equal to (a) (3000,200,200); (b) (3000,400,200); (c) (3000,400,400); (d) (5000,200,200); (e) (5000,400,200); and (f) (5000,400,400) GeV. Also shown is the 95% CL exclusion limit for the given signal.

C.9 Unblinded Analysis: Neural Network Dependence on μ

The NN efficiency on the given signals as a function of μ is given in Figure C.13 for signals with $m_A = 3000$ GeV and in Figure C.14 for signals with $m_A = 5000$ GeV. The NN output for each of the 5 different random samplings of the signal is included. The NN with the median efficiency on the signal is also indicated; note that this does not necessarily correspond to the NN that gives rise to the median expected limit, because the shape of the background may change, and so this is only indicated as an aesthetic choice. The envelope of the outputs across the different random samplings tends to be small when the efficiency is very high or very low, while the envelope widens at the μ values where the efficiency is middling; these are exactly the transition regions where the NN can find the signal but does not always. The median NN tends to be smoothly rising with μ, while single samplings from the envelope may not be; thus, the choice to run the analysis with different random samplings serves as a smoothing procedure. For $m_A = 5000$ GeV, for the best-performing signals at high m_B, m_C, the NN efficiency actually goes down with increasing μ. This is simply due to the fact that at these values of μ the amount of signal is comparable to the 10% or 1% of the signal plus background remaining after the NN tagging, so that it is mathematically impossible to have a higher efficiency on the signal and retain the overall efficiency on all the events in the signal region.

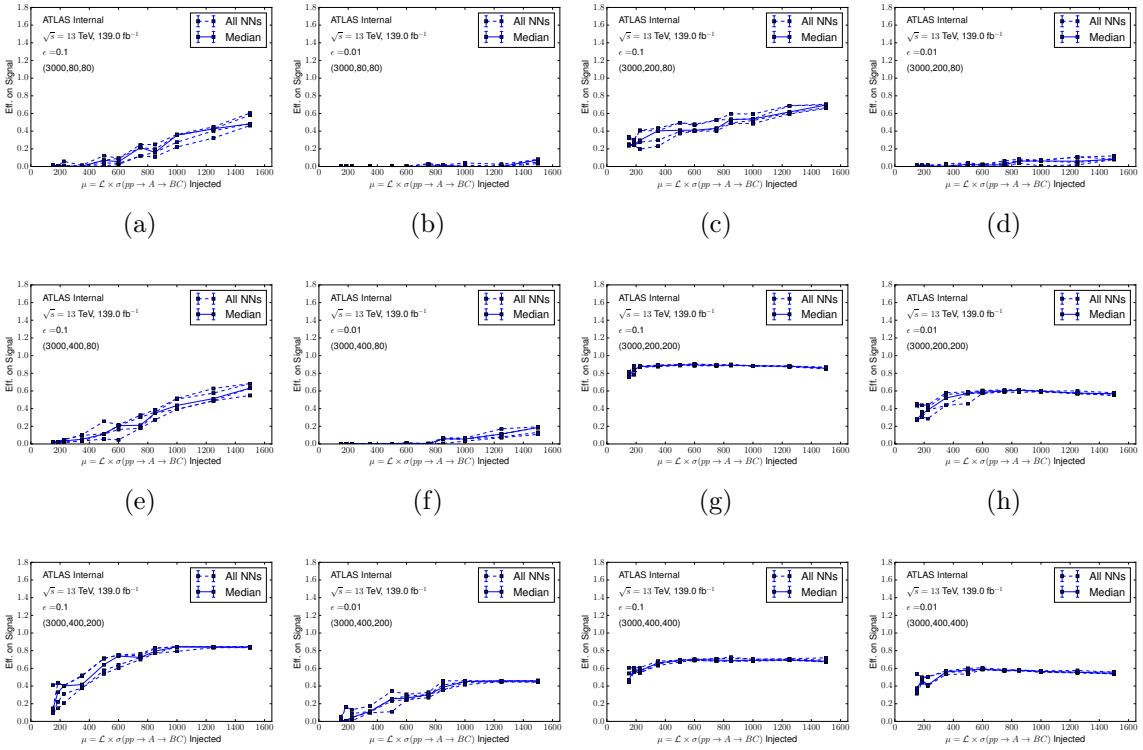


Figure C.13: The efficiency of the NN as a function of μ on the injected signal at $m_A = 3000$, in signal region 5, for (a,c,e,g,i,k) $\epsilon = 0.1$ and (b,d,f,h,j,l) $\epsilon = 0.01$. There are 5 lines corresponding to the 5 different random samplings of the signal in the training of the NN; the network with the median efficiency is also marked. Each signal is labeled by (m_B, m_C) in GeV. (a,b) (80,80); (c,d) (80,200); (e,f) (80,400); (g,h) (200,200); (i,j) (200,400); (k,l) (400,400).

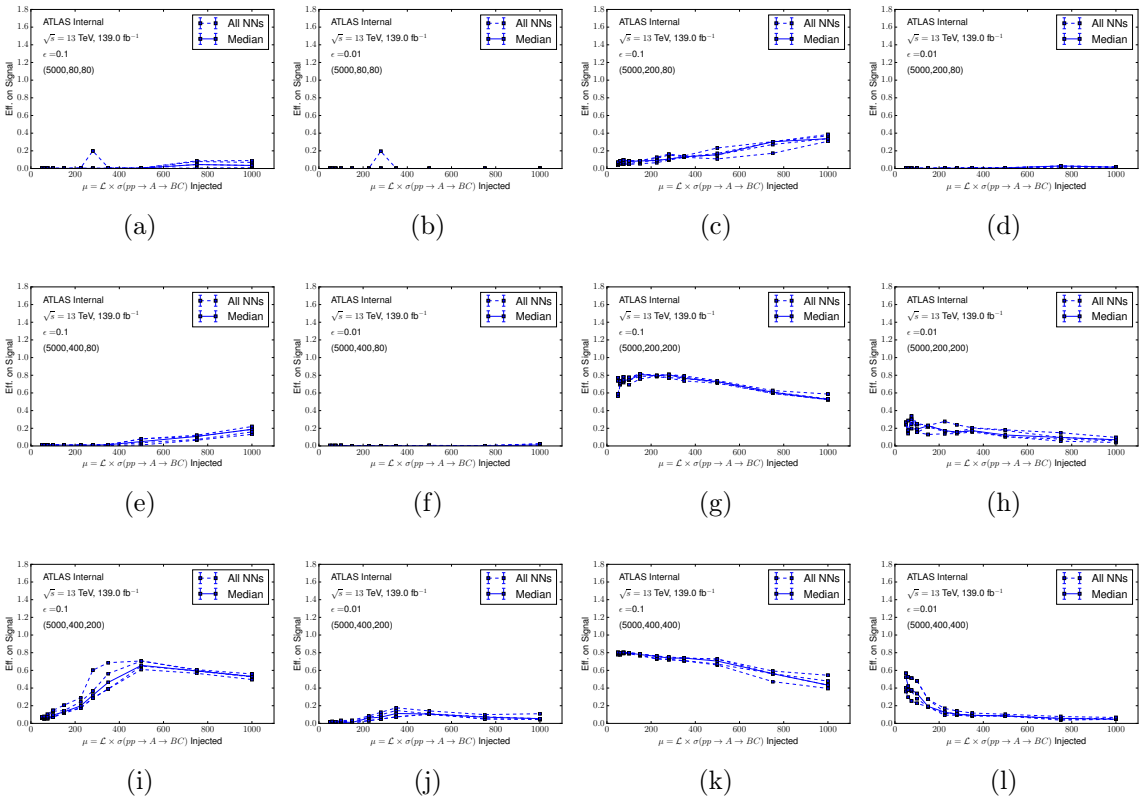


Figure C.14: The efficiency of the NN as a function of μ on the injected signal at $m_A = 5000$, in signal region 8, for (a,c,e,g,i,k) $\epsilon = 0.1$ and (b,d,f,h,j,l) $\epsilon = 0.01$. There are 5 lines corresponding to the 5 different random samplings of the signal in the training of the NN; the network with the median efficiency is also marked. Each signal is labeled by (m_B, m_C) in GeV. (a,b) (80,80); (c,d) (80,200); (e,f) (80,400); (g,h) (200,200); (i,j) (200,400); (k,l) (400,400).

C.10 Unblinded Analysis: Fit Correction

It is found that the distribution of significances of data with respect to the background fit both has a global (negative) offset and also has some dependence on m_{JJ}. This is verified in the validation (inverted rapidity cut) data, as can be seen in Figure C.15. Only bins with fit values greater than 10 are included, as the distribution of significances is not expected to be Normal for bins with low counts; see [446]. The effect is measured both when restricting the fit to just the nominal fit functions (1 and 2) and when restricting the fit to just the UA2 fit function (3).

Some key statistics of the distribution of significances when using the nominal fit functions are given in Table C.1. The fit values are used to correct the background

Table C.1: Key statistics of distribution of significances in validation selection data.

Statistic	Value
Mean	-0.16
Std. Dev.	1.02
Mean Std. Err.	0.06
m_{JJ} Fit Slope	-0.20 TeV^{-1}
m_{JJ} Fit Slope Std. Err.	0.06 TeV^{-1}
m_{JJ} Fit Intercept	0.49

fit bin-by-bin given the m_{JJ} value in that bin; i.e., the correction is exactly the green dashed line in Figure C.15b. Since there is uncertainty on the estimate of the offset and mean from the validation dataset, and in addition an uncertainty on whether the fit derived in the validation dataset applies to the signal selection dataset, an uncertainty is applied on the offset and the slope independently according to the standard error of the estimate.

The question of how exactly to apply this correction is non-trivial. The correction is derived on the significance of the data with respect to the background fit in the validation selection data, but it would be improper in general to set the median background fit value to the value that changes the significance by the correction amount, since then this correction would depend on the observed data in each bin.

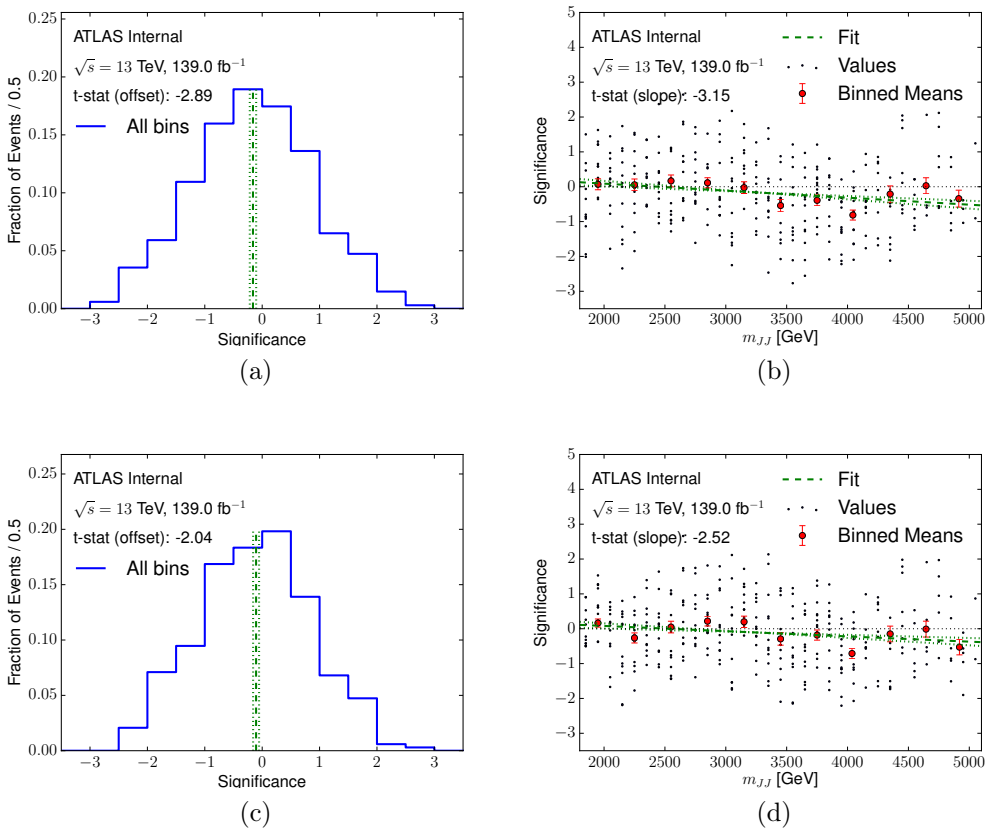


Figure C.15: Distribution of significances in validation (inverted rapidity cut) data. (a,c) Overall distribution. The green dashed line shows the mean significance and the green dotted lines show the standard error on the estimate of the mean. The t-stat of this estimate is shown on the plot as well. (b,d) Dependence on m_{JJ}. The dots show individual observations across the 6 signal region bins at each value of m_{JJ} included in the fits. The red points with error bars bin these values and show the average in order to reduce the noise. The green dashed line shows the line of best fit, and the green dotted lines show the standard error on the fit. The t-stat associated with the estimate of the slope of the fit line is shown on the plot as well. (a,b) Using the nominal fit function (1 and 2); (c,d) Using the UA2 fit function (3).

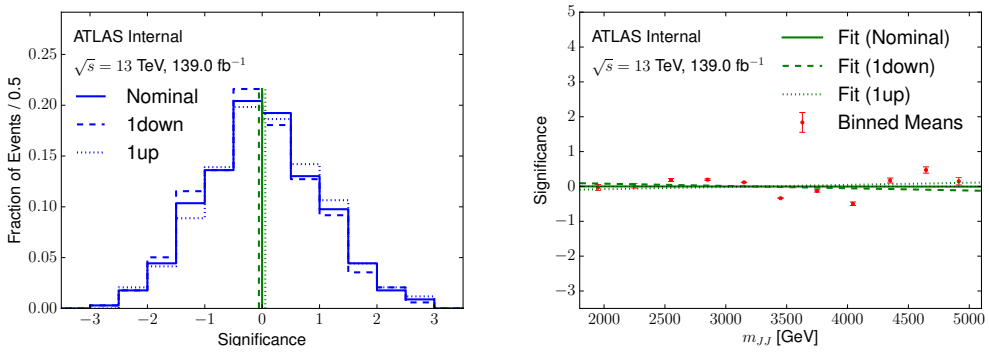
Instead, the correction is applied to change the (approximate) median of the expected distribution of event counts under the Poisson hypothesis to have a significance of the corrected value. In other words, given the background fit value E_i and correction

value c, the new fit value E_i' is given by:

$$S(E_i', \lfloor E_i + 1 \rfloor) = S(E_i, \lfloor E_i + 1 \rfloor) - c, \tag{C.7}$$

where $S(E_i, O_i)$ is the significance as defined in Equation 7.13. For the variations of the offset and the slope according to their uncertainties, new values of E_i' are calculated according to the up and down variations of each. The additional uncertainty on the background fit value E_i' is given as the sum in quadrature of the differences due to these new values.

After applying this correction, the distribution of significances in the validation selection (inverted rapidity cut) data is shown in Figure C.16. After the correction, the distribution of significances is consistent with mean 0 across the entire range, indicating that the correction (Equation C.7) is working as intended.



(a) (b)

Figure C.16: Distribution of significances in validation (inverted rapidity cut) data after applying the background fit correction. (a) Overall distribution, showing the up and down variations of the offset term. The green solid line shows the nominal mean significance and the green dashed/dotted lines show the down/up variations, respectively. (b) Dependence on m_{JJ}. The green solid line shows the line of best fit to the nominal significances, and the green dashed/dotted lines show the down/up variations on the slope term, respectively. The red points bin the individual nominal values and show the average, while the error bars indicate the same for the down/up variations of the slope term.

The correction is validated in the signal selection (no inverted rapidity cut) data

in the sidebands of the fit, since after the NN selection it is expected there will not be any significant signal presence in the sidebands, and the background fit should describe the data there. This validation is shown in Figure C.17. It can be seen both that before the correction, the distribution of significances and dependence on m_{JJ} is consistent with that observed in the validation (inverted rapidity cut) data. Some key statistics of the distribution are shown in Table C.2. Both the mean significance and the slope of the significances with respect to m_{JJ} are slightly higher than the values observed in the validation data.

Table C.2: Key statistics of distribution of significances in signal selection data.

Statistic	Value
Mean	-0.10
Std. Dev.	1.06
Mean Std. Err.	0.06
m_{JJ} Fit Slope	-0.09 TeV^{-1}
m_{JJ} Fit Slope Std. Err.	0.07 TeV^{-1}
m_{JJ} Fit Intercept	0.19

After the correction the distribution of significances is consistent with mean 0 across the entire range to within the uncertainties, again with the nominal slightly higher since the correction is kept constant from the derivation in the validation selection data. This indicates that the correction can also be applied to the bins in the signal regions.

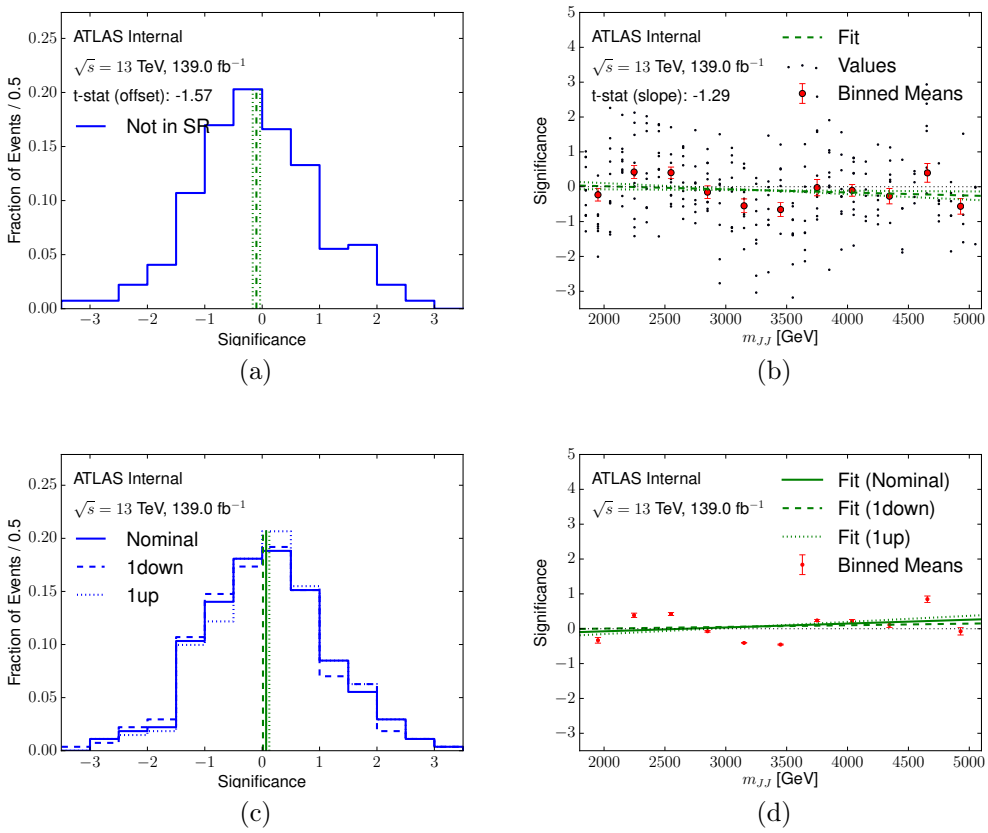


Figure C.17: Distribution of significances in signal selection (no inverted rapidity cut) data before (a,b) and after (c,d) applying the background fit correction. (a) Overall distribution. The green dashed line shows the mean significance and the green dotted lines show the standard error on the estimate of the mean. The t-stat of this estimate is shown on the plot as well. (b) Dependence on m_{JJ}. The dots show individual observations across the 6 signal region bins at each value of m_{JJ} included in the fits. The red points with error bars bin these values and show the average in order to reduce the noise. The green dashed line shows the line of best fit, and the green dotted lines show the standard error on the fit. The t-stat associated with the estimate of the slope of the fit line is shown on the plot as well. (c) Overall distribution, showing the up and down variations of the offset term. The green solid line shows the nominal mean significance and the green dashed/dotted lines show the down/up variations, respectively. (d) Dependence on m_{JJ}. The green solid line shows the line of best fit to the nominal significances, and the green dashed/dotted lines show the down/up variations on the slope term, respectively. The red points bin the individual nominal values and show the average, while the error bars indicate the same for the down/up variations of the slope term.

C.11 Unblinded Analysis: Global Distribution of Significances

The distribution of significances observed in the unblinded data in the given signal regions with no signal injected (Figure 7.36 is studied under a toy Gaussian model. In each bin, the significance S_i that is being shown in the Figures is the same significance that goes into the χ^2 calculation (Equation 7.12):

$$S_i = \frac{O_i - E_i}{E_i} \tag{C.8}$$

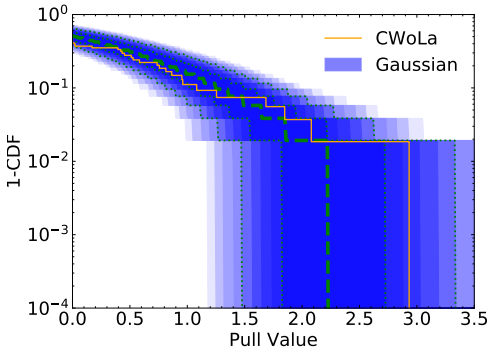
Since the toy model being used is Gaussian, we only examine the significances in bins in which the background prediction $E_i > 5$. The empirical CDF $\Phi_O(x)$ of the observed significances S_i is formed as:

$$\Phi_O(x) = \sum_{i=1}^{N} \frac{\mathbb{1}(x \leq S_i)}{N}$$

where $\mathbb{1}$ is the indicator function, and the sum goes over all N significances being considered; in this case $N = 53$.

The toy model predictions are calculated by generating, in each toy, $N = 53$ samples from a standard normal distribution and calculating the empirical toy CDF in the same way as for the observed. $N_{toys} = 20000$ empirical toy CDFs are generated in this way and quantiles across the toys are then calculated.

Figure C.18 shows the results of this test. The observed empirical CDF lies well within the Gaussian expectation out to significance $x < 2$. For the most extreme excess observed in the data, this value or larger is observed in the toys at around the 1.3-σ level, or $\sim 10\%$ of the time.



(a)

Figure C.18: The CDF at each value x for the observed data (orange) and for the toys (median is bold green dashed line and the green dotted lines correspond to (1,2)-σ quantiles, respectively; the blue shadings correspond to finer quantiles at the 0.25σ level).

C.12 Unblinded Analysis: Fits with Injected Signal

The fit results for all the signals at the injected μ value that gives rise to the limits given in Section 7.7.5 can be found in Figure C.19 for $m_A = 3000$ GeV and in Figure C.20 for m_A=5000 GeV. For signals with no limit set (because the NN did not find that signal at any value of μ), the fit at the maximum injected value (7.9) is shown.

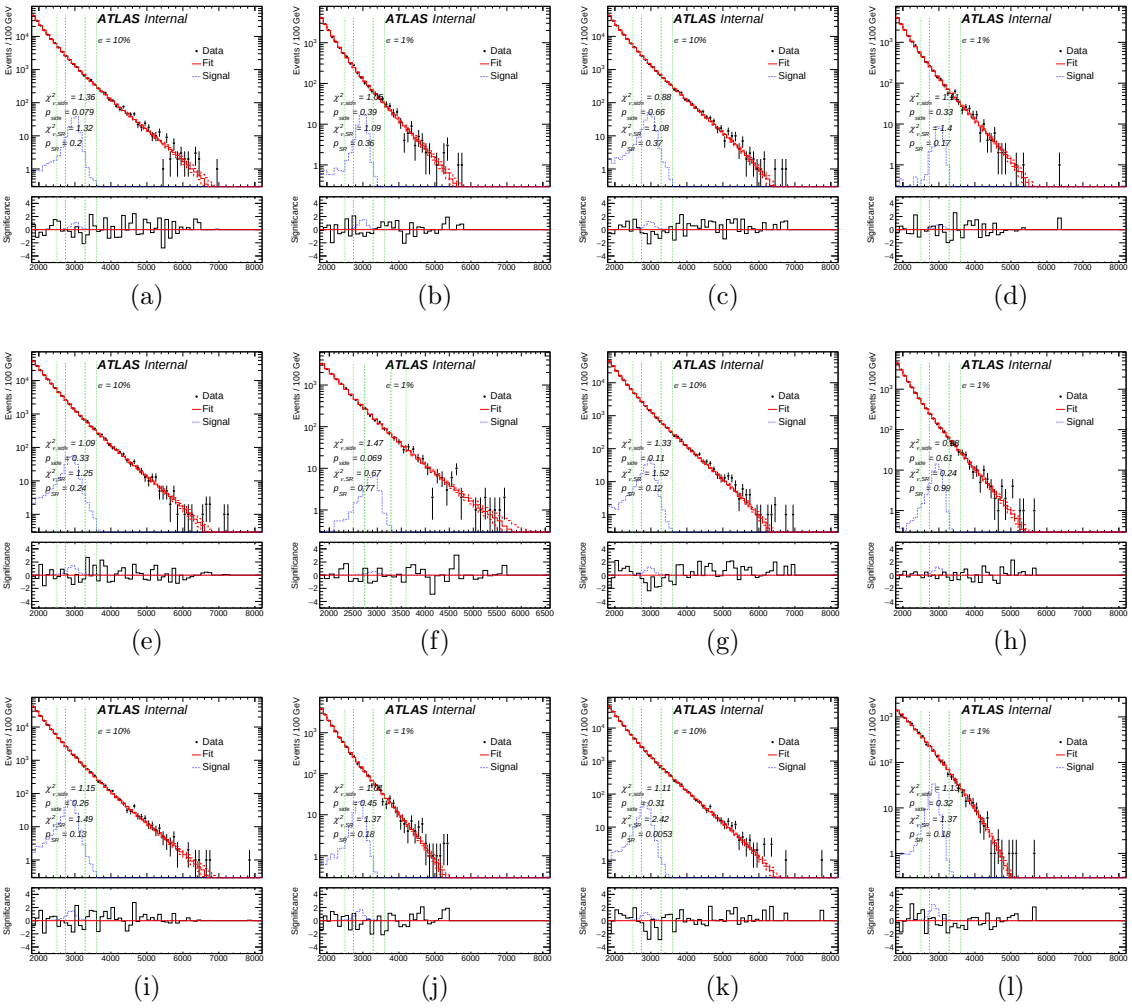


Figure C.19: The fit with an injected signal at $m_A = 3000$, in signal region 5, for (a,c,e,g,i,k) $\epsilon = 0.1$ and (b,d,f,h,j,l) $\epsilon = 0.01$. The strength of the signal is the injected μ value that gives rise to the limits given in Section 7.7.5, or the maximum injected μ if no limits are set. Each signal is labeled by (m_B, m_C) in GeVand μ for the two ϵ values. (a,b) (80,80), μ =(1250,1500); (c,d) (80,200), μ =(750,1250); (e,f) (80,400), μ =(1000,850); (g,h) (200,200), μ =(350,225); (i,j) (200,400), μ =(600,600); (k,l) (400,400), μ =(500,350). The red dashed lines indicate the fit uncertainty.

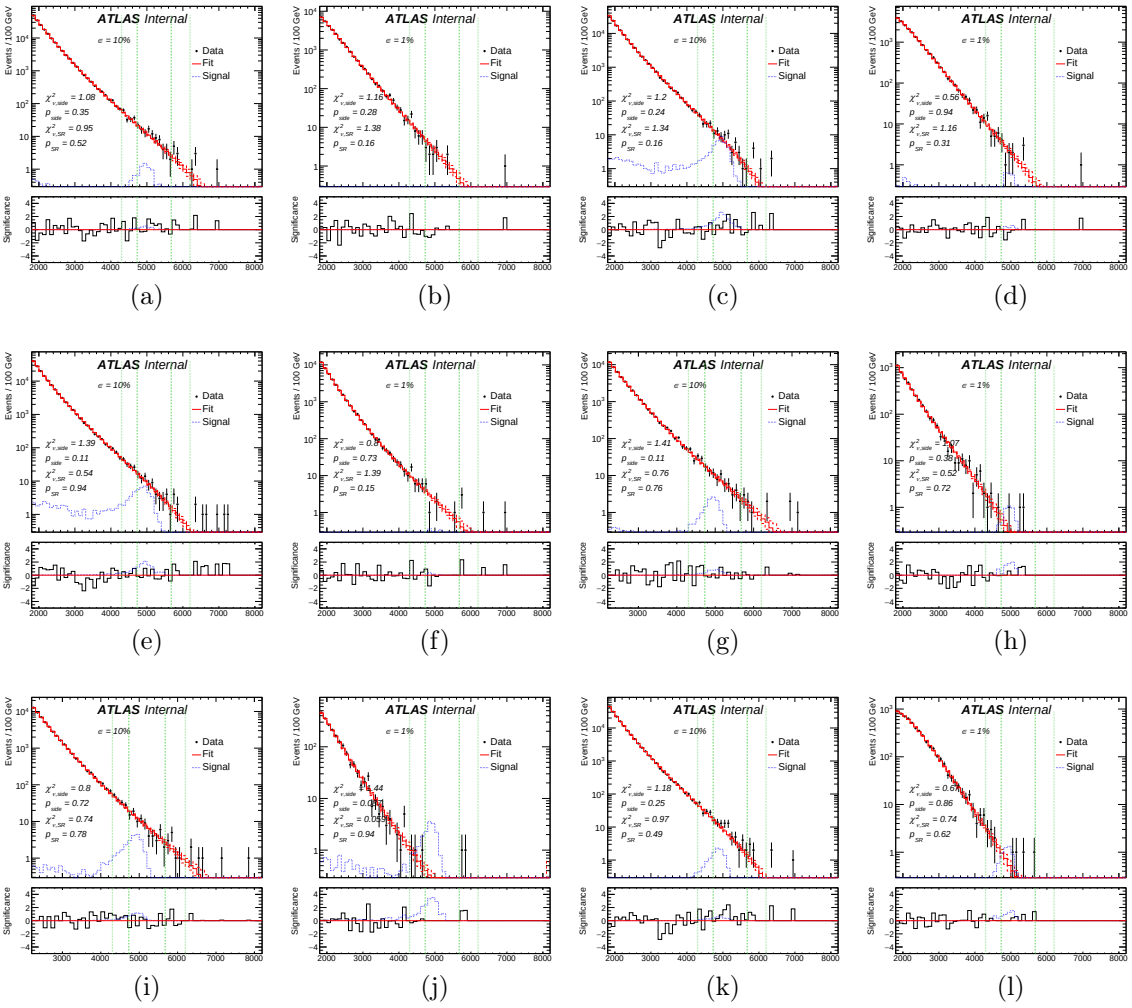


Figure C.20: The fit with an injected signal at $m_A = 5000$, in signal region 8, for (a,c,e,g,i,k) $\epsilon = 0.1$ and (b,d,f,h,j,l) $\epsilon = 0.01$. The strength of the signal is the injected μ value that gives rise to the limits given in Section 7.7.5, or the maximum injected μ if no limits are set. Each signal is labeled by (m_B, m_C) in GeVand μ for the two ϵ values. (a,b) (80,80), μ =(750,280); (c,d) (80,200), μ =(750,750); (e,f) (80,400), μ =(1000,350); (g,h) (200,200), μ =(75,75); (i,j) (200,400), μ =(280,500); (k,l) (400,400), μ =(50,50). The red dashed lines indicate the fit uncertainty.

C.13 Unblinded Analysis: Dependence of Limits on μ

The 95% CL exclusion limits on the given signals as a function of μ are given in Figure C.21 for signals with $m_A = 3000$ GeV and in Figure C.22 for signals with $m_A = 5000$ GeV. The limits for each of the 5 different random samplings of the signal is included. The network that gives rise to the median limit on the signal is also indicated, and the $\pm 1\sigma$ and $\pm 2\sigma$ bands and observed limit are indicated for this network; this is the network that is used to derive the final limits. The envelope of the outputs across the different random samplings tends to be small when the NN efficiency is very high (when it is very low the limits are large and worse than existing limits), while the envelope widens at the μ values where the efficiency is middling; these are exactly the transition regions where the NN can find the signal but does not always. The median limit tends to be smoothly falling with μ (before taking the max with the dotted line μ), while single samplings from the envelope may not be; thus, the choice to run the analysis with different random samplings serves as a smoothing procedure. For $m_A = 5000$ GeV, for the best-performing signals at high m_B, m_C, the limits actually get worse with increasing μ; this is related to the fact that the NN efficiency gets worse with increasing μ for these signals (Appendix C.9).

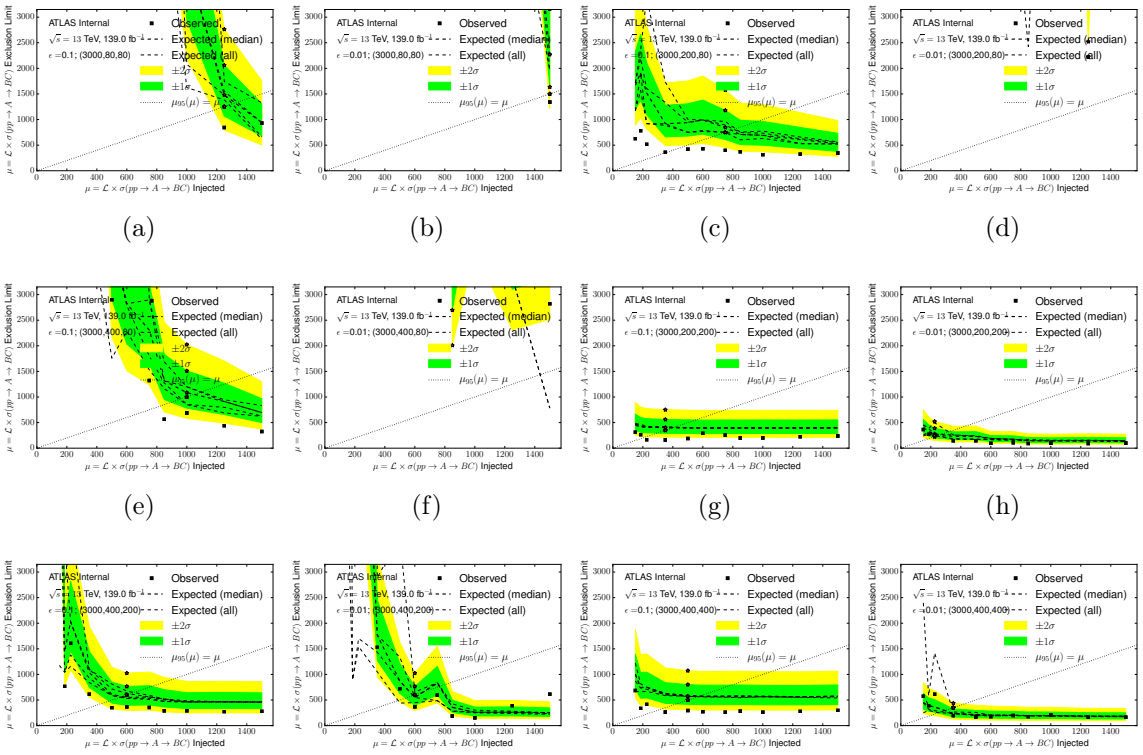


Figure C.21: The limits $\mu_{95}(\mu)$ as a function of μ on the injected signal at $m_A = 3000$, in signal region 5, for (a,c,e,g,i,k) $\epsilon = 0.1$ and (b,d,f,h,j,l) $\epsilon = 0.01$. There are 5 lines corresponding to the expected limit for the 5 different random samplings of the signal in the training of the NN; the network with the median expected limit is also marked. The $\pm 1\sigma$ and $\pm 2\sigma$ bands and the observed limit are given for the network that gives rise to the median expected limit. The red stars indicate the expected, observed, and bands of the limit after taking the max of the limit and μ. Each signal is labeled by (m_B, m_C) in GeV. (a,b) (80,80); (c,d) (80,200); (e,f) (80,400); (g,h) (200,200); (i,j) (200,400); (k,l) (400,400).

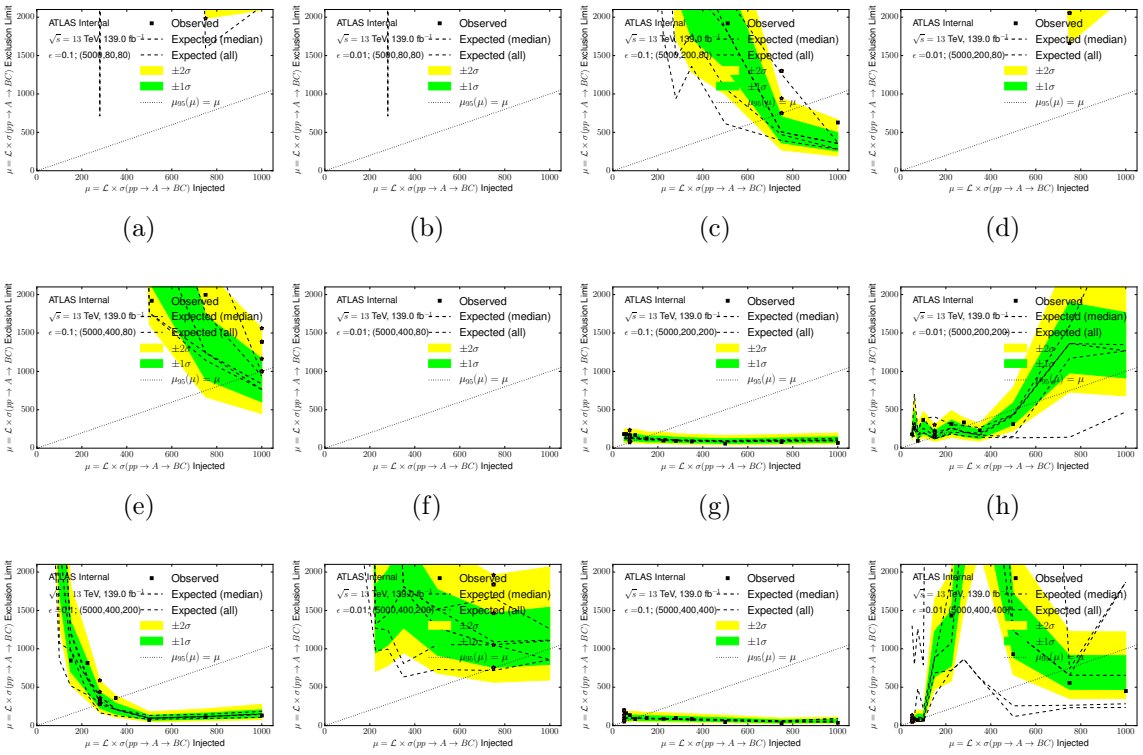


Figure C.22: The limits $\mu_{95}(\mu)$ as a function of μ on the injected signal at $m_A = 5000$, in signal region 8, for (a,c,e,g,i,k) $\epsilon = 0.1$ and (b,d,f,h,j,l) $\epsilon = 0.01$. There are 5 lines corresponding to the expected limit for the 5 different random samplings of the signal in the training of the NN; the network with the median expected limit is also marked. The $\pm1\sigma$ and $\pm2\sigma$ bands and the observed limit are given for the network that gives rise to the median expected limit. The red stars indicate the expected, observed, and bands of the limit after taking the max of the limit and μ. Each signal is labeled by (m_B, m_C) in GeV. (a,b) (80,80); (c,d) (80,200); (e,f) (80,400); (g,h) (200,200); (i,j) (200,400); (k,l) (400,400).

C.14 Computing Resources

The time for training depends on the signal region. The initial factor of 3 networks to find the one with the lowest validation loss is done sequentially in a single job, but everything else is parallelized. For signal region 4 (the largest), a single training takes $\mathcal{O}$(hours), and decreases proportionally according to the number of events in the signal region + sidebands. For a single mass point, μ value, and signal region, there are $4 \times 5 = 20$ such jobs. For the no-signal networks, there are therefore 20×5 jobs, some of which run very quickly.

There are far more jobs with injected signal. We look at 12 signal mass points, and $\mathcal{O}(10)$ μ values for each, and for a given signal hypothesis only in a single signal region. For the full unblinded analysis each training is repeated 5 times with random samplings of the signal. Therefore, for the full unblinded analysis, there are $4 \times 5 \times 12 \times 10 \times 5 = 12000$ jobs. The Author is especially grateful for the computing resources at SLAC (https://atlas.slac.stanford.edu/using-the-slac-computing-resources), which provide a batch CPU system on which a single user can reliably run about 400 jobs simultaneously. Altogether it takes $\mathcal{O}$(days-weeks) in compute time to reproduce all the results in this analysis.

Each job stores $\mathcal{O}(100)$ MB of data in output, since there are tens of millions of events for which a few key features have to be stored, e.g. the NN output and the values of the features and m_{JJ}. Probably wastefully, each step of the validation combination process is stored as a new file, in order to reduce computation time, increasing the amount of storage by a factor of ~ 2. Therefore the whole analysis uses a few TB in storage space, which again gratitude is due to the SLAC computing facilities for providing and maintaining.

Appendix D

Numerical Inversion for Jet Calibrations: Appendix

D.1 Gaussian Invariance Lemma

Let $X \sim \mathcal{N}(\mu, \sigma)$ *and* f *be some function such that* $f'(x) > 0$*. Then,* $f(X) \sim \mathcal{N}(\mu', \sigma')$ *if and only if* $f(x)$ *is linear in* x*.*

Proof. The converse is a well-known result, and can be obtained directly from application of Equation 8.12.

Now suppose that $f(X) \sim \mathcal{N}(\mu', \sigma')$. Let $Y = (X - \mu)/\sigma$ and define

$$g(y) = \frac{f(\sigma y + \mu) - \mu'}{\sigma'}, \tag{D.1}$$

so that Y and $Z = g(Y)$ both have a standard normal distribution. Furthermore,

$$g'(y) = \frac{\sigma}{\sigma'} f'(\sigma y + \mu) > 0, \tag{D.2}$$

so g is monotonic.

We then can write for any c:

$$\begin{aligned}\Phi(c) = \Pr(Y < c) &= \Pr(g(Y) < g(c)) \\ &= \Pr(Z < g(c)) \\ &= \Phi(g(c)),\end{aligned} \tag{D.3}$$

Where $\Phi(x)$ is the normal distribution cumulative distribution function. Since Φ is invertible, we then have that $g(c) = c$. Inserting the definition of g then gives us the final result:

$$f(x) = \frac{\sigma'}{\sigma}(x - \mu) + \mu'. \qquad \square \tag{D.4}$$

D.2 Closure of the Mean

The closure of jets reconstructed from truth jets with $E_T = x$ *and* $f(x) = f_{me}(x)$ *is given to first order by* $C \approx 1 - \frac{1}{2}\frac{f''(x)}{f'(x)^3}\frac{\sigma(x)^2}{x}$.

Derivation. We begin by Taylor expanding $f^{-1}(y)$ about $y = f(x)$:

$$\begin{aligned}f^{-1}(y) &= \sum_{n=0}^{\infty} \frac{1}{n!}\left(f^{-1}\right)^{(n)}(f(x)) \cdot (y - f(x))^n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} g_n(x) \cdot (y - f(x))^n,\end{aligned} \tag{D.5}$$

where $g_n(x) \equiv (f^{-1})^{(n)}(f(x))$ means the nth derivative of $f^{-1}(y)$, evaluated at $y =$

$f(x)$. Plugging this into Equation 8.13, we have

$$\begin{aligned} C(x) &= \frac{1}{x}\int dy \rho_{Y|X}(y|x) f^{-1}(y) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}\frac{g_n(x)}{x}\int dy \rho_{Y|X}(y|x)\,(y-f(x))^n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}\frac{g_n(x)}{x}\mu_n(x), \end{aligned} \tag{D.6}$$

where $\mu_n(x)$ are the standard central moments $\mu_n(x) = \mathbb{E}\left[(Y - \mathbb{E}[Y])^n | X = x\right]$, since by definition $f(x) = \mathbb{E}[Y|X = x]$.

The first few central moments are independent of the distribution $\rho_{Y|X}$. In particular, $\mu_0 = 1$ is the normalization, and $\mu_1 = 0$. Writing these terms out, we have

$$C(x) = \frac{g_0(x)}{x} + \sum_{n=2}^{\infty} \frac{1}{n!}\frac{g_n(x)}{x}\mu_n(x). \tag{D.7}$$

Noting that $g_0(x) = f^{-1}(f(x)) = x$,

$$C(x) = 1 + \sum_{n=2}^{\infty} \frac{1}{n!}\frac{g_n(x)}{x}\mu_n(x). \tag{D.8}$$

We see that, if f is linear, then so is f^{-1}, and so $g_n = 0$ for all $n \geq 2$. Then Equation D.8 reduces to $C = 1$, and numerical inversion closes, as was found in Equation 8.14.

It will be instructive to expand out the first few terms of Equation D.8. We note that, by definition, $\mu_2(x) = \sigma(x)^2$ is the variance, and $\mu_3(x) = \sigma(x)^3\gamma_1$ defines the skew γ_1. Then we have

$$C(x) = 1 + \frac{1}{2}\frac{g_2(x)}{x}\sigma(x)^2 + \frac{1}{6}\frac{g_3(x)}{x}\sigma(x)^3\gamma_1 + \sum_{n=4}^{\infty} \frac{1}{n!}\frac{g_n(x)}{x}\mu_n(x). \tag{D.9}$$

Suppose we are given an arbitrary distribution specified by its moments $\mu_n(x)$. Then the requirement that closure is satisfied in the form of the right hand side of

Equation D.8 converging to 1 exactly imposes strict constraints on the function $g(x)$, so that only for a highly specific choice of g and therefore f is closure achieved. Thus in general we do not expect closure to be satisfied for an arbitrary initial distribution $\rho_{Y|X}$.

We note that, since we expect the derivatives $g_n(x)$ and the moments $\mu_n(x)$ to grow considerably slower than $n!$ for functions f and distributions $\rho_{Y|X}$ encountered at the LHC, we expect Equation D.8 to converge, and Equation D.9 gives the dominant contributions to the non-closure, i.e.

$$C(x) \approx 1 + \frac{1}{2}\frac{g_2(x)}{x}\sigma(x)^2 + \frac{1}{6}\frac{g_3(x)}{x}\sigma(x)^3\gamma_1. \tag{D.10}$$

If $\rho_{Y|X}$ is symmetric or near-symmetric, or if the third derivative of g is small, such that $g_3(x)\sigma(x)\gamma_1 \ll g_2(x)$, then the dominant contribution to the non-closure is just

$$C(x) \approx 1 + \frac{1}{2}\frac{g_2(x)}{x}\sigma(x)^2. \tag{D.11}$$

We further note that

$$\begin{aligned} g_2(x) = (f^{-1})^{(2)}(f(x)) &= -\frac{f''(x)}{f'(x)^3} \\ \rightarrow C(x) \approx 1 - \frac{1}{2}\frac{f''(x)}{f'(x)^3}\frac{\sigma(x)^2}{x}. &\qquad \square \end{aligned} \tag{D.12}$$

D.3 Calibrated Resolution of the Mean

The calibrated resolution of jets reconstructed from truth jets with $E_T = x$ *and* $f(x) = f_{me}(x)$ *is given to first order by* $\frac{\sigma(x)}{f'(x)}$.

Derivation. We note that, expanding $f^{-1}(y)$ about $y = f(x)$ out to one derivative, and using the definitions of $g_n(x)$ and $\mu_n(x)$ from the previous section,

$$(f^{-1}(y))^2 \approx g_0(x)^2 + 2g_0(x)g_1(x)(y - f(x)) + g_1(x)^2(y - f(x))^2, \tag{D.13}$$

so that

$$\begin{aligned}\mathbb{E}\left[Z^2\middle|X=x\right] &= \int dy \rho_{Y|X}(y|x)(f^{-1}(y))^2 \\ &\approx \int dy \rho_{Y|X}(y|x)\left(g_0(x)^2 + 2g_0(x)g_1(x)(y-f(x)) + g_1(x)^2(y-f(x))^2\right) \\ &= g_0(x)^2\mu_0(x) + 2g_0(x)g_1(x)\mu_1(x) + g_1(x)^2\mu_2(x) \\ &= g_0(x)^2 + g_1(x)^2\sigma(x)^2. \quad (\mu_1 = 0 \text{ by construction}) \end{aligned} \tag{D.14}$$

Out to one derivative we also have that (as derived in the previous section)

$$\begin{aligned}\mathbb{E}\left[Z|X=x\right]^2 &\approx g_0(x)^2 \\ \rightarrow \sigma\left[Z|X=x\right]^2 &= \mathbb{E}\left[Z^2\middle|X=x\right] - \mathbb{E}\left[Z|X=x\right]^2 \\ &\approx g_1(x)^2\sigma(x)^2. \end{aligned} \tag{D.15}$$

Then,

$$\begin{aligned} g_1(x) = (f^{-1})'(f(x)) &= \frac{1}{f'(x)} \\ \rightarrow \sigma\left[Z|X=x\right]^2 &\approx \frac{\sigma(x)^2}{f'(x)^2} \\ \rightarrow \hat{\sigma}(x) = \sigma\left[Z|X=x\right] &\approx \frac{\sigma(x)}{f'(x)}. \qquad \square \end{aligned} \tag{D.16}$$

D.4 Closure of the Mode

The closure of jets reconstructed from truth jets with $E_T = x$ *and* $f(x) = f_{mo}(x)$ *is given to first order by* $C \approx 1 + \frac{f''(x)}{f'(x)^3}\frac{\tilde{\sigma}(x)^2}{x}$.

Derivation. As a reminder for the reader, for brevity, we will let $\rho_Y(y) = \rho_Y(y|x)$ and $\rho_Z(z) = \rho_Z(z|x)$, and let the parameter x be understood.

We begin by supposing that the closure is not much different than 1, so that we can examine $\rho_Z(z)$ in the vicinity of $z = x$ to find the mode z^*. Expanding Equation 8.12

about to second order in $(z-x)$:

$$\begin{aligned}\rho_Z(z) &= f'(z)\rho_Y(f(z)) \\ &\approx \left[f'(x) + (z-x)f''(x) + \frac{(z-x)^2}{2}f'''(x)\right] \\ &\times \left[\rho_Y(f(x)) + (z-x)\rho_Y'(f(x))f'(x) + \frac{(z-x)^2}{2}\rho_Y''(f(x))f'(x)^2\right]. \end{aligned} \tag{D.17}$$

We note from the condition Equation 8.23 that $\rho_Y'(f(x)) = 0$, so

$$\begin{aligned}\rho_Z(z) &\approx \left[f'(x) + (z-x)f''(x) + \frac{(z-x)^2}{2}f'''(x)\right] \\ &\times \left[\rho_Y(f(x)) + \frac{(z-x)^2}{2}\rho_Y''(f(x))f'(x)^2\right] \\ &\approx f'(x)\rho_Y(f(x)) + (z-x)f''(x)\rho_Y(f(x)) \\ &+ \frac{(z-x)^2}{2}\left[f'''(x)\rho_Y(f(x)) + f'(x)^3\rho_Y''(f(x))\right], \end{aligned} \tag{D.18}$$

so that

$$\rho_Z'(z) \approx f''(x)\rho_Y(f(x)) + (z-x)\left[f'''(x)\rho_Y(f(x)) + f'(x)^3\rho_Y''(f(x))\right]. \tag{D.19}$$

Then the closure condition Equation 8.24 gives

$$\begin{aligned}&\rho_Z'(z^*) = 0 \\ &\rightarrow z^* \approx x - \frac{f''(x)\rho_Y(f(x))}{f'''(x)\rho_Y(f(x)) + f'(x)^3\rho_Y''(f(x))}, \end{aligned} \tag{D.20}$$

i.e. the mode of $\rho_Z(z)$ occurs at $z = z^*$. Then the closure is

$$\begin{aligned} C(x) &= \frac{z^*}{x} \\ &\approx 1 - \frac{1}{x}\frac{f''(x)\rho_Y(f(x))}{f'''(x)\rho_Y(f(x)) + f'(x)^3\rho_Y''(f(x))} \\ &= 1 - \frac{1}{x}\frac{f''(x)\frac{\rho_Y(f(x))}{\rho_Y''(f(x))}}{f'''(x)\frac{\rho_Y(f(x))}{\rho_Y''(f(x))} + f'(x)^3} \\ &= 1 + \frac{f''(x)}{f'(x)^3 - \tilde{\sigma}(x)^2 f'''(x)}\frac{\tilde{\sigma}(x)^2}{x}. \end{aligned} \tag{D.21}$$

In practice we find that for typical response functions, higher derivatives of f tend to vanish. A comparison between the two terms in the denominator of Equation D.21 can be found in Figure D.1 for the toy model considered in Appendix D.9; we find that $f'(x)^3 \gg \tilde{\sigma}(x)^2 f'''(x)$. Thus, in practice we recommend the approximation

$$C(x) \approx 1 + \frac{f''(x)}{f'(x)^3}\frac{\tilde{\sigma}(x)^2}{x}. \qquad \square \tag{D.22}$$

The agreement between the actual and estimated closure in Figure 8.4 also confirms this approximation. Thus, in the body of this text Equation D.22 is presented as the result, even though Equation D.21 is technically more precise.

D.5 Resolution of the Mode

The resolution of jets reconstructed from truth jets with $E_T = x$ *and* $f(x) = f_{mo}(x)$ *is given to first order by* $\hat{\tilde{\sigma}}(x) \approx \frac{\tilde{\sigma}(x)}{f'(x)}$.

Derivation. From Equation D.19 we have

$$\rho_Z''(z) \approx f'''(x)\rho_Y(f(x)) + f'(x)^3\rho_Y''(f(x)). \tag{D.23}$$

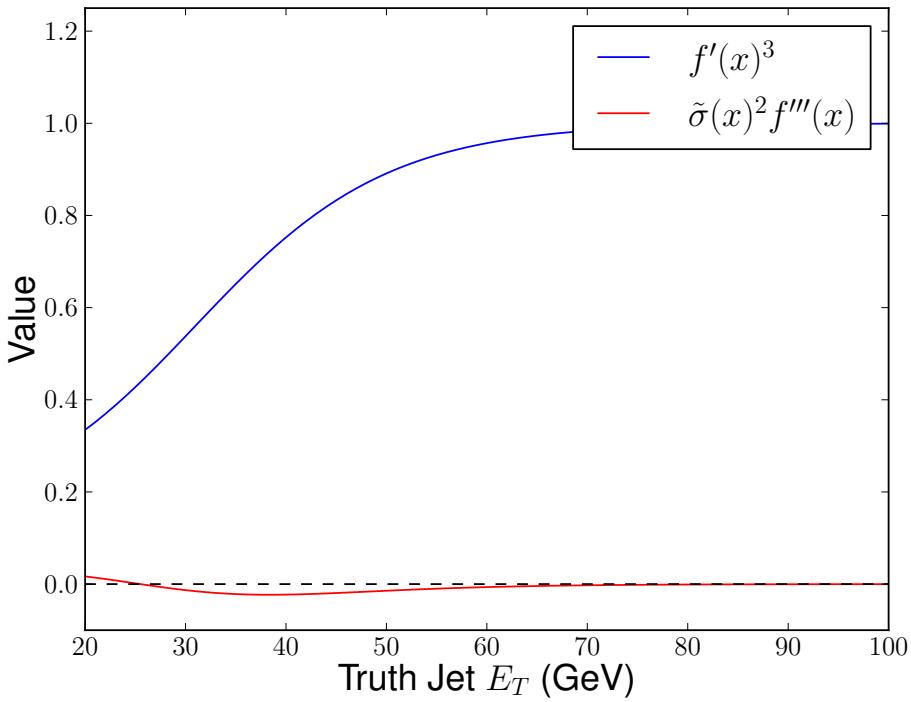


Figure D.1: A comparison of derivative values using a toy model similar to conditions in ATLAS or CMS. In blue, $\mathtt{f}'(\mathtt{x})^3$. In red, $\tilde{\sigma}(\mathtt{x})^2\mathtt{f}'''(\mathtt{x})$. For details of the model, see Appendix D.9.

Then the resolution is given as

$$\begin{aligned}\hat{\tilde{\sigma}}(x)^2 &= -\frac{\rho_Z(z^*)}{\rho_Z''(z^*)} \\ &\approx -\frac{f'(x)\rho_Y(f(x))}{f'''(x)\rho_Y(f(x)) + f'(x)^3\rho_Y''(f(x))} \\ &= \frac{f'(x)\tilde{\sigma}(x)^2}{f'(x)^3 - f'''(x)\tilde{\sigma}(x)^2}. \end{aligned} \tag{D.24}$$

Following the discussion in Appendix D.4, we simplify the denominator to get the approximation

$$\begin{aligned}\tilde{\sigma}(x)^2 &\approx \frac{\tilde{\sigma}(x)^2}{f'(x)^2} \\ \rightarrow \tilde{\sigma}(x) &\approx \frac{\tilde{\sigma}(x)}{f'(x)}. \qquad \square \end{aligned} \tag{D.25}$$

D.6 Iterated Numerical Inversion Calculation

The closure $C_{new}(x)$ after iterating numerical inversion is not necessarily closer to 1 than the closure $C(x)$ after performing numerical inversion once.

Derivation. We limit ourselves to the case that we are using the modes of the distributions $Y|X = x$ and $Z|X = x$ to calibrate, as in practice that is what is used at ATLAS and CMS for numerical inversion.

We use the estimation of the closure of the mode Equation 8.28:

$$\begin{aligned} C(x) &\approx 1 + \frac{f''(x)}{f'(x)^3}\frac{\tilde{\sigma}(x)^2}{x} \\ \rightarrow |C(x) - 1| &\approx \left|\frac{f''(x)}{f'(x)^3}\frac{\tilde{\sigma}(x)^2}{x}\right|. \end{aligned} \tag{D.26}$$

We use the iterated numerical inversion response

$$\begin{aligned} f_{new}(x) &= C(x)x \\ &\approx x + \frac{f''(x)}{f'(x)^3}\tilde{\sigma}(x)^2 \qquad \text{(D.27)} \\ \rightarrow f'_{new}(x) &\approx 1 - 3\frac{f''(x)^2}{f'(x)^4}\tilde{\sigma}(x)^2 \qquad \text{(D.28)} \\ \rightarrow f''_{new}(x) &\approx 12\frac{f''(x)^3}{f'(x)^5}\tilde{\sigma}(x)^2. \qquad \text{(D.29)} \end{aligned}$$

Where we have ignored higher derivatives of $f(x)$[1] and derivatives of $\sigma(x)$[2]. We also have the estimation of the resolution of the calibrated distribution Equation 8.32

$$\hat{\tilde{\sigma}}(x) \approx \frac{\tilde{\sigma}(x)}{f'(x)}, \qquad \text{(D.30)}$$

So that we can estimate the closure after iterating numerical inversion as

$$\begin{aligned} C_{new}(x) &\approx 1 + \frac{f''_{new}(x)}{f'_{new}(x)^3}\frac{\hat{\tilde{\sigma}}(x)^2}{x} \\ &\approx 1 + 12\frac{f''(x)^3}{f'(x)^5}\tilde{\sigma}(x)^2\frac{\tilde{\sigma}(x)^2}{f'(x)^2}\frac{1}{x} \\ &= 1 + \frac{12}{x}\frac{f''(x)^3}{f'(x)^7}\tilde{\sigma}(x)^4 \qquad \text{(D.31)} \\ \rightarrow |C_{new}(x) - 1| &\approx \left|\frac{12}{x}\frac{f''(x)^3}{f'(x)^7}\tilde{\sigma}(x)^4\right| \qquad \text{(D.32)} \\ \rightarrow \frac{|C_{new}(x) - 1|}{|C(x) - 1|} &\approx \frac{12f''(x)^2\tilde{\sigma}(x)^2}{f'(x)^4}. \qquad \text{(D.33)} \end{aligned}$$

If the ratio in Equation D.33 is greater than 1, then the closure gets worse after a second iteration of numerical inversion. □

[1] See, e.g., Figure D.1.

[2] For this specific counterexample, we are examining the case that $\sigma'(x) = 0$, which is realistic for high pile-up conditions.

D.7 Corrected Numerical Inversion Calculation

With $Y \mapsto Z_{\text{corr}} = g^{-1}(Y)$, we will get a corrected calibrated distribution $\rho_{Z_{\text{corr}}|X}(z|x)$. For brevity, let $\rho_{Z_{\text{corr}}}(z) = \rho_{Z_{\text{corr}}|X}(z|x)$, where it is understood we are examining the distributions around a particular value of x. We will again require that $g'(x) > 0$, so that

$$\rho_{Z_{\text{corr}}}(z) = g'(z)\rho_Y(g(z)). \tag{D.34}$$

The closure condition is then equivalent to the condition

$$\rho'_{Z_{\text{corr}}}(x) = 0, \tag{D.35}$$

i.e., the mode of the distribution $Z_{\text{corr}}|X = x$ occurs at x. We have that

$$\rho'_{Z_{\text{corr}}}(z) = g''(z)\rho_Y(g(z)) + g'(z)^2\rho'_Y(g(z)), \tag{D.36}$$

so that the closure condition requires

$$\begin{aligned} 0 &= \rho'_{Z_{\text{corr}}}(x) \\ &= g''(x)\rho_Y(g(x)) + g'(x)^2\rho'_Y(g(x)) \\ \rightarrow 0 &= g''(x) + g'(x)^2\frac{\rho'_Y(g(x))}{\rho_Y(g(x))}. \end{aligned} \tag{D.37}$$

We suppose that $g(x)$ is close to $f(x)$, $g(x) = f(x) + \alpha(x)$, with $|\alpha(x)| \ll \tilde{\sigma}(x)$. Then we have directly from the supposition that the distribution $Y|X = x$ is approximately Gaussian about its mode $f(x)$ with width $\tilde{\sigma}(x)$ that

$$\frac{\rho'_Y(g(x))}{\rho_Y(g(x))} = -\frac{(g(x) - f(x))}{\tilde{\sigma}(x)^2}. \tag{D.38}$$

Then, the closure condition gives

$$\begin{aligned} 0 &= g''(x) + g'(x)^2 \frac{\rho_Y'(g(x))}{\rho_Y(g(x))} \\ &= g''(x) - g'(x)^2 \frac{g(x) - f(x)}{\tilde{\sigma}(x)^2}. \end{aligned} \tag{D.39}$$

D.8 Corrected Numerical Inversion Parameterization

We parameterize the corrected calibration function $g(x) = g(x; f(x); a_1, ..., a_n)$. For the toy model used in this chapter, we use the parameterization

$$g(x) = f(x) + \frac{a_1}{1 + \exp(\frac{x - a_2}{a_3})}. \tag{D.40}$$

In the model considered here, and for the response functions at the LHC, the closure goes to 1 for large x and moves away from 1 for small x, a natural result of Equation 8.28. Thus, the parameterization in Equation D.40 includes a "turn-off" to recover $g(x) = f(x)$ at large x (with $a_3 > 0$).

In practice, there is some smallest value $x = x'$ which is being studied, and which per the discussion in the above paragraph tends to have the largest non-closure. The value $x' = 20$ GeV is used in this chapter, which is the lowest calibrated E_T at current conditions at the LHC. For the corrected calibration curve shown in Figure 8.5, the parameters a_1, a_2, a_3 are scanned over to minimize the non-closure at this value x'. For the corrected calibration curve shown in Figure 8.5, the values $a_2 = a_3 = x' = 20$ GeV and $a_1 = 5$ GeV were used.

D.9 Toy Model of the ATLAS/CMS Response Function

All the "Proofs" quoted in Chapter 8 are valid in general, regardless of the response function $R(x)$ and the underlying distributions $Y|X = x$ (within the assumptions

outlined in Section 8.2.3). We also expect that the "Derivations", which are all approximate formulas, to apply in a wide variety of cases. In order to visualize some of the results, and verify the approximations, a particular model was needed in order to get numerical values. All figures made in this chapter were derived from a simple model of the ATLAS or CMS jet E_T response function[3]. After specifying $f(x)$ and the distributions $Y|X = x$, the calibrated distributions were constructed using the analytic form of the calibrated distributions Equation 8.12. Then the various moments were found numerically for the calibrated distribution at each value x.

The response function was guided both by physical intuition and by the intention to reasonably simulate response functions published by ATLAS [81] and CMS [410, 411]. When there is only a small amount of energy already in a detector cell, the detector only reconstructs a small fraction of the energy put into it, because of noise thresholds and the non-compensating nature of the ATLAS and CMS detectors. Whereas if there is already a lot of energy in a detector cell, the detector reconstructs almost all of the energy put into it. Thus $f'(x)$ was designed to be low at low values of x and then to rise steadily to 1 at high values of x. This intuition does not directly apply to jets that directly use tracking information (e.g. particle-flow jets in CMS), but for the sake of simplicity only one (calorimeter) jet definition is used for illustration.

$f'(x)$ was then integrated to get $f(x)$ and divided by x to get $R(x)$. The resulting $R(x)$ function approximately corresponds to the $R = 0.4$ anti-k_t [171] central jet response at the EM scale available in Ref. [81] (e.g. Fig. 4a). The shapes of $f'(x)$ and $R(x)$ in this model can be seen in Figure D.2.

In this simplified model, the distributions $Y|X = x \sim \mathcal{N}(f(x), \sigma(x))$ were used. In ATLAS and CMS, $Y|X = x$ is approximately Gaussian. The constant value of $\sigma(x) = 7$ GeV was used, corresponding to a calibrated resolution (Fig. 8.3) of about 50% at $E_T = 20$ GeV. This is consistent with e.g. [85] and has the property that $\sigma'(x) = 0$, which should be the case if pile-up is the dominant contributor to the

[3]Energies are measured with calorimeters and momenta are measured with tracking detectors. In-situ corrections using momentum balance techniques constrain the momentum. For small-radius QCD jets, the E_T and p_T are nearly identical. Since the simulation-based correction of calorimeter jets is used here as a model, the E_T is used throughout.

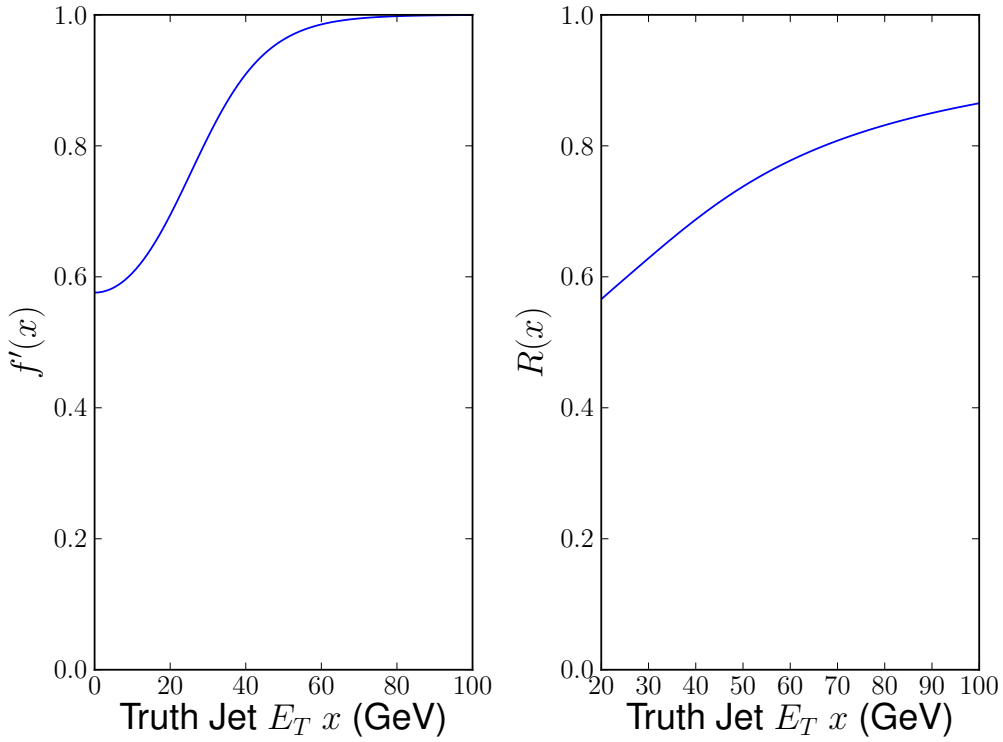


Figure D.2: The toy model used in this chapter to simulate conditions in ATLAS or CMS. The left plot shows $\mathsf{f}'(\mathsf{x})$ and the right plot shows $\mathsf{R}(\mathsf{x})$.

resolution of low $\mathsf{E_T}$ jets.

Appendix E

Improving Jet Calibrations with Machine Learning: Appendix

E.1 Correcting for Auxiliary Variables

In the following, note that the description of the jet energy correction applies both to the currently used method of correction in the ATLAS GSC (Section 9.3) for a single variable and to generalized numerical inversion (Section 9.4). The difference between the two methods is only in the details of the derivation of $f_\theta(x)$; the former does a binned fit and is only used with a single variable at a time, while the latter does an unbinned fit and allows for fitting multiple variables at once.

Following the notation of Chapter 8, let X be a random variable representing p_T^{true} and Y be a random variable representing p_T^{reco}. We also let θ represent some auxiliary variable on which the jet energy depends. For demonstrative purposes suppose this effect is linear,

$$f_\theta(x) = \alpha(x)\theta + \beta(x). \tag{E.1}$$

We note that, using the law of total expectation,

$$f(x) = \mathbb{E}_\theta \left[f_\theta(x) \right] \tag{E.2}$$

$$= \alpha(x)\mu_\theta(x) + \beta(x) \tag{E.3}$$

where $\mu_\theta(x) \equiv \mathbb{E}\left[\theta | X = x\right]$. Here we use the shorthand $\mathbb{E}_\theta[\cdot] = \mathbb{E}[\cdot | X = x]$ to indicate the expectation taken over θ while keeping X fixed.

Assuming that after the inclusive correction the calibration closes overall ($f(x) = x$) further imposes the form

$$f_\theta(x) = \alpha(x)(\theta - \mu_\theta(x)) + x. \tag{E.4}$$

We now turn to the resolution of the jets before and after the correction. We define

$$\sigma(x)^2 \equiv \mathbb{E}\left[(Y - f(x))^2 | X = x \right], \tag{E.5}$$

$$\sigma(x,\theta)^2 \equiv \mathbb{E}\left[(Y - f_\theta(x))^2 | X = x, \theta \right], \tag{E.6}$$

$$\sigma_\theta(x)^2 \equiv \mathbb{E}\left[(\theta - \mu_\theta(x))^2 | X = x \right]. \tag{E.7}$$

I.e., $\sigma(x)^2$ is the overall variance of Y given $X = x$, which is simply the (square of the) resolution after the inclusive calibration but before the correction for the auxiliary variable; $\sigma(x,z)^2$ is the variance of Y given both $X = x$ and θ; and $\sigma_\theta(x)$ is the variance of θ itself given $X = x$.

Before the correction,

$$\sigma(x)^2 = \mathbb{E}\left[Y^2 | X = x \right] - f(x)^2 \tag{E.8}$$

$$= \mathbb{E}_\theta \left[\mathbb{E}[Y^2 | X = x, \theta] \right] - x^2 \tag{E.9}$$

$$= \mathbb{E}_\theta \left[\sigma(x,\theta)^2 + f(x,\theta)^2 \right] - x^2 \tag{E.10}$$

$$= \mathbb{E}_\theta \left[\sigma(x,\theta)^2 \right] + \mathbb{E}_\theta \left[\alpha(x)^2 (\theta - \mu_\theta(x)^2 \right] \tag{E.11}$$

$$= \mathbb{E}_\theta \left[\sigma(x,\theta)^2 \right] + \alpha(x)^2 \sigma_\theta(x)^2 \tag{E.12}$$

The last term, including the variance of θ, comes from the fact that we have a dependence on the external variable. This is exactly the term that gives rise to the intuition that there is an increase in the spread of the jet energy due to the dependence on θ and the spread of θ itself, and is the impetus for correcting this dependence.

For deriving the resolution after the correction, it will be important to understand the derivative of $f_\theta(x)$:

$$f'_\theta(x) = 1 + \alpha'(x)(\theta - \mu_\theta(x)) - \alpha(x)\mu'_\theta(x). \tag{E.13}$$

The correction is $Y \mapsto \hat{Y} = f_\theta^{-1}(Y)$, so that after the correction:

$$\begin{aligned}
\hat{f}_\theta(x) &\equiv \mathbb{E}\left[\hat{Y}|X = x, \theta\right] && \text{(E.14)}\\
&\approx f_\theta(f_\theta^{-1}(x)) = x, && \text{(E.15)}\\
\hat{\sigma}(x,\theta)^2 &\equiv \mathbb{E}\left[\left(\hat{Y} - \hat{f}_\theta(x)\right)^2 |X = x, \theta\right] && \text{(E.16)}\\
&\approx \frac{\sigma(x,\theta)^2}{f'_\theta(x)^2}, && \text{(E.17)}
\end{aligned}$$

where both approximations come from Chapter 8 (Equation 8.20 and Equation 8.22).

The response overall has closure:

$$\begin{aligned}
\hat{f}(x) &\equiv \mathbb{E}\left[\hat{Y}|X = x\right] && \text{(E.18)}\\
&= \mathbb{E}_\theta\left[\hat{f}_\theta(x)\right] && \text{(E.19)}\\
&= x && \text{(E.20)}
\end{aligned}$$

And the resolution overall is

$$\begin{aligned}
\hat{\sigma}(x)^2 &\equiv \mathbb{E}\left[(\hat{Y} - \hat{f}(x))^2 | X = x\right] && \text{(E.21)}\\
&= \mathbb{E}_\theta\left[\mathbb{E}\left[(\hat{Y} - \hat{f}_\theta(x))^2 | X = x, \theta\right]\right] && \text{(E.22)}\\
&= \mathbb{E}_\theta\left[\hat{\sigma}(x,\theta)^2\right] && \text{(E.23)}\\
&= \mathbb{E}_\theta\left[\frac{\sigma(x,\theta)^2}{f'(x,\theta)^2}\right] && \text{(E.24)}\\
&= \mathbb{E}_\theta\left[\frac{\sigma(x,\theta)^2}{(1 + \alpha'(x)(\theta - \mu_\theta(x)) - \alpha(x)\mu'_\theta(x))^2}\right] && \text{(E.25)}
\end{aligned}$$

The resolution without the correction is always worse than $\mathbb{E}_\theta\left[\sigma(x,\theta)^2\right]$, due to the spread of the response due to θ, and this is the value we should compare the corrected resolution to.

To gain intuition, we examine a few simple cases. $\alpha'(x)$ captures whether the dependence of the jet energy on θ changes with x, and $\mu'_\theta(x)$ captures whether θ itself is correlated with X rather than just the fluctuations in Y around X.

The first thing to note is that if both $\alpha'(x) = 0$ and $\mu'_\theta(x) = 0$, then the resolution is exactly $\mathbb{E}_\theta\left[\sigma(x,z)^2\right]$ and so strictly gets better. However, in this case, numerical inversion is unnecessary, as $f_\theta(x) = \alpha(\theta - \mu_\theta) + x$ and so a simple correction $Y \mapsto Y - \alpha(\theta - \mu_\theta)$ is sufficient to remove this effect. In effect this is what is done in the pile-up subtraction stage of the jet calibration (Section 5.5.4.1), which is a simple offset and does not use numerical inversion.

Suppose then that $\mu'_\theta(x) = 0$ and $\alpha'(x) \neq 0$. If the spread of $\alpha'(x)(\theta - \mu_\theta)$ is small, then $\mathbb{E}_\theta\left[\frac{1}{(1+\alpha'(x)(\theta-\mu_\theta))^2}\right] \approx \mathbb{E}_\theta\left[1 - 2\alpha'(x)(\theta - \mu_\theta)\right] = 1$ and the resolution is exactly $\mathbb{E}_\theta\left[\sigma(x,z)^2\right]$; so the resolution after correction is always better. However, the resolution suffers depending on the support of θ. If there are even a few values of θ for which $\alpha'(x)(\theta - \mu_\theta)$ is near -1, then the derivative goes to 0 at those values and those values are calibrated to plus or minus infinity. Because of this, the calibrated resolution blows up, even as the uncalibrated resolution is finite. Even when using the trimmed Gaussian width instead of the standard deviation, if there is enough support of θ at these extreme values then the resolution can blow up.

As another case, suppose that $\alpha'(x) = 0$ but $\mu'_\theta(x) \neq 0$. Then the resolution depends on the sign of $\alpha\mu'_\theta(x)$. Taking $\alpha > 0$ for concreteness, this could either be because θ gives information about the fluctuations in $Y-X$, in which case $\mu'_\theta(x) < 0$; or it could be because θ contains information about X directly, in which case $\mu'_\theta(x) > 0$. In the former case, $\alpha\mu'_\theta(x) < 0$, and correcting for the dependence on θ makes the resolution better, because these fluctuations are being accounted for. In the latter case, $\alpha\mu'_\theta(x) > 0$, and correcting for the dependence on θ makes the resolution worse, because information about the correlation between Y and X is actually being removed by the correction. The story is the same, *mutatis mutandis*, with $\alpha < 0$.

As an extreme example of the degradation of the resolution if $\alpha\mu'_\theta(x) > 0$, suppose $\mu_\theta(x) = x$, so $\mu'_\theta(x) = 1$, and $\alpha = 1$. Then $f_\theta(x) = (\theta - x) + x = \theta$. So after correcting for the dependence of Y on θ, all the correlation between Y and X has been removed, and the resolution of Y given X is therefore infinite. In fact, in this case it makes more sense to simply use θ as the measurement of the jet energy and calibrate that with respect to X, an idea which is beyond the scope of this Thesis.

Bibliography

[1] E. M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: Learning from mixed samples in high energy physics*, *JHEP* **10** (2017) 174 [1708.02949].

[2] J. Collins, "CWoLa Hunting: Data-driven anomaly detection for new physics at the LHC." https://indico.cern.ch/event/558411/contributions/3420889/, May, 2019.

[3] **CMS** Collaboration, *Model Unspecific Search for New Physics in pp Collisions at sqrt(s) = 7 TeV*, .

[4] **CMS** Collaboration, *MUSiC, a Model Unspecific Search for New Physics, in pp Collisions at* $\sqrt{s} = 8$ TeV, Tech. Rep. CMS-PAS-EXO-14-016, CERN, Geneva, 2017.

[5] **ATLAS** Collaboration, *A general search for new phenomena with the ATLAS detector in pp collisions at sort(s)=7 TeV.*, Tech. Rep. ATLAS-CONF-2012-107, CERN, Geneva, Aug, 2012.

[6] **ATLAS** Collaboration, *A general search for new phenomena with the ATLAS detector in pp collisions at* $\sqrt{s} = 8$ *TeV*, Tech. Rep. ATLAS-CONF-2014-006, CERN, Geneva, Mar, 2014.

[7] **ATLAS** Collaboration, M. Aaboud *et. al.*, *A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment*, *Eur. Phys. J.* **C79** (2019), no. 2 120 [1807.07447].

[8] M. Farina, Y. Nakai and D. Shih, *Searching for New Physics with Deep Autoencoders*, 1808.08992.

[9] T. Heimel, G. Kasieczka, T. Plehn and J. M. Thompson, *QCD or What?*, *SciPost Phys.* **6** (2019), no. 3 030 [1808.08979].

[10] T. S. Roy and A. H. Vijay, *A robust anomaly finder based on autoencoder*, 1903.02032.

[11] O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu and J.-R. Vlimant, *Variational Autoencoders for New Physics Mining at the Large Hadron Collider*, *JHEP* **05** (2019) 036 [1811.10276].

[12] A. Blance, M. Spannowsky and P. Waite, *Adversarially-trained autoencoders for robust unsupervised new physics searches*, *JHEP* **10** (2019) 047 [1905.10384].

[13] J. Hajer, Y.-Y. Li, T. Liu and H. Wang, *Novelty Detection Meets Collider Physics*, 1807.10261.

[14] B. M. Dillon, D. A. Faroughy and J. F. Kamenik, *Uncovering latent jet substructure*, *Phys. Rev.* **D100** (2019), no. 5 056002 [1904.04200].

[15] B. Nachman and D. Shih, *Anomaly Detection with Density Estimation*, *Phys. Rev. D* **101** (2020) 075042 [2001.04990].

[16] A. Andreassen, B. Nachman and D. Shih, *Simulation Assisted Likelihood-free Anomaly Detection*, *Phys. Rev. D* **101** (2020), no. 9 095004 [2001.05001].

[17] J. H. Collins, K. Howe and B. Nachman, *Anomaly Detection for Resonant New Physics with Machine Learning*, *Phys. Rev. Lett.* **121** (2018), no. 24 241803 [1805.02664].

[18] J. H. Collins, K. Howe and B. Nachman, *Extending the search for new resonances with machine learning*, *Phys. Rev.* **D99** (2019), no. 1 014038 [1902.02634].

[19] **ATLAS** Collaboration, G. Aad *et. al.*, *Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **03** (2020) 145 [1910.08447].

[20] **ATLAS** Collaboration, G. Aad *et. al.*, *Search for diboson resonances in hadronic final states in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **09** (2019) 091 [1906.08589].

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

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

[23] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys.Lett.* **12** (1964) 132–133.

[24] G. Guralnik, C. Hagen and T. Kibble, *Global Conservation Laws and Massless Particles*, *Phys.Rev.Lett.* **13** (1964) 585–587.

[25] 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 [1207.7214].

[26] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [1207.7235].

[27] Nobel Media AB 2020, "The nobel prize in physics 2013." https://www.nobelprize.org/prizes/physics/2013/summary/, 2020.

[28] V. Trimble, *Existence and Nature of Dark Matter in the Universe*, *Ann. Rev. Astron. Astrophys.* **25** (1987) 425–472.

[29] J. Silk *et. al.*, *Particle Dark Matter: Observations, Models and Searches*. Cambridge Univ. Press, Cambridge, 2010.

[30] A. G. Riess, "Dark matter." https://www.britannica.com/science/dark-matter.

[31] C. CsÃąki and P. Tanedo, *Beyond the Standard Model*, in *2013 European School of High-Energy Physics*, pp. 169–268, 2015. 1602.04228.

[32] A. de Gouvea, D. Hernandez and T. M. P. Tait, *Criteria for Natural Hierarchies*, *Phys. Rev. D* **89** (2014), no. 11 115005 [1402.2658].

[33] H. P. Nilles, M. Srednicki and D. Wyler, *Weak Interaction Breakdown Induced by Supergravity*, *Phys. Lett.* **120B** (1983) 346.

[34] M. Dine, *TASI lectures on the strong CP problem*, in *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2000): Flavor Physics for the Millennium*, pp. 349–369, 6, 2000. hep-ph/0011376.

[35] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440–1443.

[36] J. Wess and J. Bagger, *Supersymmetry and supergravity*. World Publishing Corporation, 2009.

[37] J. Wess, *SUPERSYMMETRY AND SUPERGRAVITY*, in *22nd International Conference on High-Energy Physics*, p. II.84, 7, 1984.

[38] J. D. Lykken, *Introduction to supersymmetry*, in *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 96): Fields, Strings, and Duality*, pp. 85–153, 6, 1996. hep-th/9612114.

[39] F. Quevedo, S. Krippendorf and O. Schlotterer, *Cambridge Lectures on Supersymmetry and Extra Dimensions*, 1011.1491.

[40] M. J. Strassler, *An Unorthodox introduction to supersymmetric gauge theory*, in *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2001): Strings, Branes and EXTRA Dimensions*, pp. 561–638, 9, 2003. hep-th/0309149.

[41] M. E. Peskin, *Duality in supersymmetric Yang-Mills theory*, in *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 96): Fields, Strings, and Duality*, pp. 729–809, 2, 1997. hep-th/9702094.

[42] N. Seiberg, *Electric - magnetic duality in supersymmetric nonAbelian gauge theories*, *Nucl. Phys. B* **435** (1995) 129–146 [hep-th/9411149].

[43] K. A. Intriligator and N. Seiberg, *Lectures on Supersymmetry Breaking*, *Class. Quant. Grav.* **24** (2007) S741–S772 [hep-ph/0702069].

[44] V. Baluni, *CP Violating Effects in QCD*, *Phys. Rev. D* **19** (1979) 2227–2230.

[45] R. Crewther, P. Di Vecchia, G. Veneziano and E. Witten, *Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics*, *Phys. Lett. B* **88** (1979) 123. [Erratum: Phys.Lett.B 91, 487 (1980)].

[46] R. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons*, *Phys. Rev. D* **16** (1977) 1791–1797.

[47] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223–226.

[48] F. Wilczek, *Problem of Strong* P *and* T *Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279–282.

[49] J. D. Lykken, *Beyond the Standard Model*, 5, 2010. 1005.1676.

[50] H. M. Lee, *Lectures on Physics Beyond the Standard Model*, 7, 2019. 1907.12409.

[51] J. Ellis, *Physics Beyond the Standard Model*, *Nucl. Phys. A* **827** (2009) 187C–198C [0902.0357].

[52] J. Ellis, *Outstanding questions: Physics beyond the Standard Model*, *Phil. Trans. Roy. Soc. Lond. A* **370** (2012) 818–830.

[53] T. Virdee, *Beyond the standard model of particle physics*, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **374** (08, 2016) 20150259.

[54] E. Halkiadakis, G. Redlinger and D. Shih, *Status and Implications of Beyond-the-Standard-Model Searches at the LHC*, *Ann. Rev. Nucl. Part. Sci.* **64** (2014) 319–342 [1411.1427].

[55] D. Curtin *et. al.*, *Exotic decays of the 125 GeV Higgs boson*, *Phys. Rev. D* **90** (2014), no. 7 075004 [1312.4992].

[56] **ATLAS** Collaboration, G. Aad *et. al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.

[57] "LHC Guide." https://cds.cern.ch/record/2255762, Mar, 2017.

[58] **LHC Study Group** Collaboration, T. S. Pettersson and P. Lefèvre, *The Large Hadron Collider: conceptual design*, Tech. Rep. CERN-AC-95-05-LHC, Oct, 1995.

[59] L. R. Evans, *The Large Hadron Collider Project*, .

[60] *LHC Machine*, *JINST* **3** (2008) S08001.

[61] M. Gell-Mann, *Quarks*, *Acta Phys. Austriaca Suppl.* **9** (1972) 733–761.

[62] H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, *Phys. Lett. B* **47** (1973) 365–368.

[63] H. Fritzsch and M. Gell-Mann, *Current algebra: Quarks and what else?*, *eConf* **C720906V2** (1972) 135–165 [hep-ph/0208010].

[64] A. Ali and G. Kramer, *Jets and QCD: A Historical Review of the Discovery of the Quark and Gluon Jets and its Impact on QCD*, *Eur. Phys. J. H* **36** (2011) 245–326 [1012.2288].

[65] S. Drell, D. J. Levy and T.-M. Yan, *A Theory of Deep Inelastic Lepton-Nucleon Scattering and Lepton Pair Annihilation Processes. 3. Deep Inelastic electron-Positron Annihilation*, *Phys. Rev. D* **1** (1970) 1617–1639.

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

[67] R. Blankenbecler, S. J. Brodsky and J. Gunion, *Inclusive Processes at High Transverse Momentum*, *Phys. Lett. B* **42** (1972) 461–465.

[68] **ATLAS, CMS** Collaboration, G. Pásztor, *Precision tests of the Standard Model at the LHC with the ATLAS and CMS detectors*, *PoS* **FFK2019** (2020) 005.

[69] R. Kogler *et. al.*, *Jet Substructure at the Large Hadron Collider: Experimental Review*, *Rev. Mod. Phys.* **91** (2019), no. 4 045003 [1803.06991].

[70] **ATLAS** Collaboration, G. Aad *et. al.*, *Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment*, *Phys. Rev. D* **101** (2020), no. 1 012002 [1909.02845].

[71] A. Canepa, *Searches for Supersymmetry at the Large Hadron Collider*, *Rev. Phys.* **4** (2019) 100033.

[72] **ATLAS** Collaboration, J. Tam, *Summary of results on exotic searches from ATLAS*, *EPJ Web Conf.* **95** (2015) 04065.

[73] **ATLAS** Collaboration, B. Meirose, *Overview of dark matter searches at the ATLAS experiment*, *Int. J. Mod. Phys. Conf. Ser.* **43** (2016) 1660196.

[74] G. Soyez, *Pileup mitigation at the LHC: A theorist's view*. PhD thesis, IPhT, Saclay, 2018. 1801.09721.

[75] **ATLAS** Collaboration, G. Aad *et. al.*, *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 [1306.4945].

[76] B. Nachman, P. Nef, A. Schwartzman, M. Swiatlowski and C. Wanotayaroj, *Jets from Jets: Re-clustering as a tool for large radius jet reconstruction and grooming at the LHC*, *JHEP* **02** (2015) 075 [1407.2922].

[77] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084 [0912.1342].

[78] *Performance of Boosted W Boson Identification with the ATLAS Detector*, Tech. Rep. ATL-PHYS-PUB-2014-004, CERN, Geneva, Mar, 2014.

[79] **ATLAS Collaboration** Collaboration, *Impact of Alternative Inputs and Jet Grooming on Large-R Jet Performance*, Tech. Rep. ATL-PHYS-PUB-2019-027, CERN, Geneva, Jul, 2019.

[80] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Phys. Rev. D* **96** (2017), no. 7 072002 [1703.09665].

[81] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, *Eur. Phys. J. C* **73** (2013) [1112.6426].

[82] **ATLAS** Collaboration, M. Aaboud *et. al.*, *In situ calibration of large-radius jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector*, *Eur. Phys. J.* **C79** (2019), no. 2 135 [1807.09477].

[83] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett.* **B659** (2008) 119–126 [0707.1378].

[84] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04** (2008) 005 [0802.1188].

[85] **ATLAS** Collaboration, G. Aad *et. al.*, *Performance of pile-up mitigation techniques for jets in* pp *collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, *Eur. Phys. J.* **C76** (2016), no. 11 581 [1510.03823].

[86] A. J. Larkoski, I. Moult and B. Nachman, *Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning*, *Phys. Rept.* **841** (2020) 1–63 [1709.04464].

[87] **ATLAS** Collaboration, *Quark versus Gluon Jet Tagging Using Charged Particle Multiplicity with the ATLAS Detector*, Tech. Rep. ATL-PHYS-PUB-2017-009, CERN, Geneva, May, 2017.

[88] **ATLAS** Collaboration, G. Aad *et. al.*, *Light-quark and gluon jet discrimination in* pp *collisions at* $\sqrt{s} = 7$ TeV *with the ATLAS detector*, *Eur. Phys. J.* **C74** (2014), no. 8 3023 [1405.6583].

[89] **ATLAS** Collaboration, *Discrimination of Light Quark and Gluon Jets in* pp *collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2016-034, CERN, Geneva, Jul, 2016.

[90] **ATLAS** Collaboration, *Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector*, Tech. Rep. ATL-PHYS-PUB-2017-017, CERN, Geneva, Jul, 2017.

[91] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, *Eur. Phys. J.* **C77** (2017), no. 9 580 [1705.02211]. [Erratum: Eur. Phys. J.C77,no.10,712(2017)].

[92] **ATLAS** Collaboration, G. Aad *et. al.*, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at* $\sqrt{s} = 8$ *TeV*, *Eur. Phys. J.* **C76** (2016), no. 3 154 [1510.05821].

[93] **ATLAS** Collaboration, G. Aad *et. al.*, *Identification of high transverse momentum top quarks in* pp *collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, *JHEP* **06** (2016) 093 [1603.03127].

[94] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC*, *Eur. Phys. J.* **C79** (2019), no. 5 375 [1808.07858].

[95] **ATLAS** Collaboration, G. Aad *et. al.*, *Identification of boosted Higgs bosons decaying into* b*-quark pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J.* **C79** (2019), no. 10 836 [1906.11005].

[96] **ATLAS** Collaboration, *Standard Model Summary Plots Summer 2019*, Tech. Rep. ATL-PHYS-PUB-2019-024, CERN, Geneva, Jul, 2019.

[97] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3, 2014.

[98] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[99] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 6, 2005.

[100] D. Griffiths, *Introduction to Elementary Particles*. Wiley, 2 ed., 2008.

[101] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, *Phys. Rev.* **96** (Oct, 1954) 191–195.

[102] Nobel Media AB 2020, "The nobel prize in physics 1957." https://www.nobelprize.org/prizes/physics/1957/summary/, 2020.

[103] S. L. Glashow, *The renormalizability of vector meson interactions*, *Nucl. Phys.* **10** (1959) 107–117.

[104] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968) 367–377.

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

[106] Nobel Media AB 2020, "The nobel prize in physics 1979." https://www.nobelprize.org/prizes/physics/1979/summary/, 2020.

[107] **Particle Data Group** Collaboration, M. Tanabashi *et. al.*, *Review of particle physics*, *Phys. Rev. D* **98** (Aug, 2018) 030001.

[108] D. Kazakov, *Beyond the standard model: In search of supersymmetry*, in *2000 European School of High-Energy Physics*, pp. 125–199, 8, 2000. hep-ph/0012288.

[109] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, *Phys. Rev. Lett.* **30** (1973) 1343–1346.

[110] D. J. Gross, *How to Test Scaling in Asymptotically Free Theories*, *Phys. Rev. Lett.* **32** (1974) 1071.

[111] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Determination of the strong coupling constant α_s from transverse energy–energy correlations in multijet events at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **77** (2017), no. 12 872 [1707.02562].

[112] K. G. Wilson, *Confinement of Quarks*, .

[113] **LHCb** Collaboration, R. Aaij *et. al.*, *Observation of $J/\psi\phi$ structures consistent with exotic states from amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays*, *Phys. Rev. Lett.* **118** (2017), no. 2 022003 [1606.07895].

[114] **LHCb** Collaboration, R. Aaij *et. al.*, *Amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays*, *Phys. Rev. D* **95** (2017), no. 1 012002 [1606.07898].

[115] R. Feynman, *THE BEHAVIOR OF HADRON COLLISIONS AT EXTREME ENERGIES*, pp. 497–518, 9, 1989.

[116] S. Amoroso *et. al.*, *Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report*, in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*, 3, 2020. 2003.01700.

[117] L. Del Debbio, *Parton distributions in the LHC era*, *EPJ Web Conf.* **175** (2018) 01006.

[118] **NNPDF** Collaboration, R. D. Ball *et. al.*, *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017), no. 10 663 [1706.00428].

[119] Wikipedia contributors, "Standard model of particle physics — Wikipedia, the free encyclopedia." https://en.wikipedia.org/wiki/Standard_Model, 2020.

[120] R. Gupta, *Introduction to lattice QCD: Course*, in *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*, pp. 83–219, 7, 1997. hep-lat/9807028.

[121] S. Durr *et. al.*, *Ab-Initio Determination of Light Hadron Masses*, *Science* **322** (2008) 1224–1227 [0906.3599].

[122] T. Blum, A. Denig, I. Logashenko, E. de Rafael, B. Roberts, T. Teubner and G. Venanzoni, *The Muon (g-2) Theory Value: Present and Future*, 1311.2198.

[123] T. Aoyama, M. Hayakawa, T. Kinoshita and M. Nio, *Complete Tenth-Order QED Contribution to the Muon g-2*, *Phys. Rev. Lett.* **109** (2012) 111808 [1205.5370].

[124] **Muon g-2** Collaboration, G. Bennett *et. al.*, *Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL*, *Phys. Rev. D* **73** (2006) 072003 [hep-ex/0602035].

[125] **Muon g-2** Collaboration, A. Chapelain, *The Muon g-2 experiment at Fermilab*, *EPJ Web Conf.* **137** (2017) 08001 [1701.02807].

[126] **LHCb** Collaboration, J. Alves, A.Augusto *et. al.*, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.

[127] B. Capdevila, A. Crivellin, S. Descotes-Genon, J. Matias and J. Virto, *Patterns of New Physics in* $\mathfrak{b} \to \mathfrak{s}\ell^+\ell^-$ *transitions in the light of recent data*, *JHEP* **01** (2018) 093 [1704.05340].

[128] Z.-z. Xing, *Quark Mass Hierarchy and Flavor Mixing Puzzles*, *Int. J. Mod. Phys. A* **29** (2014) 1430067 [1411.2713].

[129] **Planck** Collaboration, P. Ade *et. al.*, *Planck 2013 results. XVI. Cosmological parameters*, *Astron. Astrophys.* **571** (2014) A16 [1303.5076].

[130] **Super-Kamiokande** Collaboration, Y. Fukuda *et. al.*, *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562–1567 [hep-ex/9807003].

[131] R. A. Battye and A. Moss, *Evidence for Massive Neutrinos from Cosmic Microwave Background and Lensing Observations*, *Phys. Rev. Lett.* **112** (2014), no. 5 051303 [1308.5870].

[132] **KamLAND** Collaboration, T. Araki *et. al.*, *Measurement of neutrino oscillation with KamLAND: Evidence of spectral distortion*, *Phys. Rev. Lett.* **94** (2005) 081801 [hep-ex/0406035].

[133] **MINOS** Collaboration, J. Evans, *The MINOS Experiment: Results and Prospects*, *Adv. High Energy Phys.* **2013** (2013) 182537 [1307.0721].

[134] S. King, *Neutrino mass models*, *Rept. Prog. Phys.* **67** (2004) 107–158 [hep-ph/0310204].

[135] B. Schwingenheuer, *Status and prospects of searches for neutrinoless double beta decay*, *Annalen Phys.* **525** (2013) 269–280 [1210.7432].

[136] B. A. Dobrescu and K. T. Matchev, *Light axion within the next-to-minimal supersymmetric standard model*, *JHEP* **09** (2000) 031 [hep-ph/0008192].

[137] U. Ellwanger, J. F. Gunion, C. Hugonie and S. Moretti, *Towards a no lose theorem for NMSSM Higgs discovery at the LHC*, hep-ph/0305109.

[138] R. Dermisek and J. F. Gunion, *Escaping the large fine tuning and little hierarchy problems in the next to minimal supersymmetric model and h —> aa decays*, *Phys. Rev. Lett.* **95** (2005) 041801 [hep-ph/0502105].

[139] S. Chang, R. Dermisek, J. F. Gunion and N. Weiner, *Nonstandard Higgs Boson Decays*, *Ann. Rev. Nucl. Part. Sci.* **58** (2008) 75–98 [0801.4554].

[140] D. E. Morrissey and A. Pierce, *Modified Higgs Boson Phenomenology from Gauge or Gaugino Mediation in the NMSSM*, *Phys. Rev.* **D78** (2008) 075029 [0807.2259].

[141] T. Kaluza, *Zum Unitätsproblem der Physik*, *Sitzungsber.Preuss.Akad.Wiss.Berlin(Math.Phys.)* (1921) 966–972 [1803.08616].

[142] O. Klein, *Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English)*, *Z. Phys.* **37** (1926) 895–906.

[143] A. Einstein and P. Bergmann, *On a Generalization of Kaluza's Theory of Electricity*, *Annals Math.* **39** (1938) 683–701.

[144] C. Csaki, *TASI lectures on extra dimensions and branes*, in *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2002): Particle Physics and Cosmology: The Quest for Physics Beyond the Standard Model(s)*, pp. 605–698, 4, 2004. hep-ph/0404096.

[145] C. Csaki, J. Hubisz and P. Meade, *TASI lectures on electroweak symmetry breaking from extra dimensions*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in $D \geqq 4$*, pp. 703–776, 10, 2005. hep-ph/0510275.

[146] R. Rattazzi, *Cargese lectures on extra-dimensions*, in *Cargese School of Particle Physics and Cosmology: the Interface*, pp. 461–517, 8, 2003. hep-ph/0607055.

[147] E. Ponton, *TASI 2011: Four Lectures on TeV Scale Extra Dimensions*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: The Dark Secrets of the Terascale*, pp. 283–374, 2013. 1207.3827.

[148] H.-C. Cheng, *Introduction to Extra Dimensions*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small*, pp. 125–162, 2011. 1003.1162.

[149] L. Canetti, M. Drewes and M. Shaposhnikov, *Matter and Antimatter in the Universe*, *New J. Phys.* **14** (2012) 095012 [1204.4186].

[150] A. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, *Sov. Phys. Usp.* **34** (1991), no. 5 392–393.

[151] C. Baker *et. al.*, *An Improved experimental limit on the electric dipole moment of the neutron*, *Phys. Rev. Lett.* **97** (2006) 131801 [hep-ex/0602020].

[152] F. Zwicky, *On the Masses of Nebulae and of Clusters of Nebulae*, *Astrophysical Journal* **86** (Oct, 1937) 217.

[153] V. C. Rubin and J. Ford, W. Kent, *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, *Astrophysical Journal* **159** (Feb, 1970) 379.

[154] K. C. Freeman, *On the Disks of Spiral and S0 Galaxies*, *Astrophysical Journal* **160** (Jun, 1970) 811.

[155] T. van Albada, J. N. Bahcall, K. Begeman and R. Sancisi, *The Distribution of Dark Matter in the Spiral Galaxy NGC-3198*, *Astrophys. J.* **295** (1985) 305–313.

[156] D. Clowe, A. Gonzalez and M. Markevitch, *Weak lensing mass reconstruction of the interacting cluster 1E0657-558: Direct evidence for the existence of dark matter*, *Astrophys. J.* **604** (2004) 596–603 [astro-ph/0312273].

[157] M. Markevitch, A. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman, C. Jones, S. Murray and W. Tucker, *Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56*, *Astrophys. J.* **606** (2004) 819–824 [astro-ph/0309303].

[158] **Planck** Collaboration, N. Aghanim *et. al.*, *Planck 2018 results. VI. Cosmological parameters*, 1807.06209.

[159] S. D. White, C. Frenk and M. Davis, *Clustering in a Neutrino Dominated Universe*, *Astrophys. J.* **274** (1983) L1–L5.

[160] G. G. Raffelt, *Axions in astrophysics and cosmology*, in *30th Rencontres de Moriond: Euroconferences: Dark Matter in Cosmology, Clocks and Tests of Fundamental Laws*, pp. 159–168, 1995. hep-ph/9502358.

[161] G. Jungman, M. Kamionkowski and K. Griest, *Supersymmetric dark matter*, *Phys. Rept.* **267** (1996) 195–373 [hep-ph/9506380].

[162] C. Rovelli, *Notes for a brief history of quantum gravity*, in *9th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories (MG 9)*, pp. 742–768, 6, 2000. gr-qc/0006061.

[163] A. Ashtekar and R. Geroch, *Quantum theory of gravitation*, *Reports on Progress in Physics* **37** (Oct, 1974) 1211–1256.

[164] A. Ashtekar, M. Reuter and C. Rovelli, *From General Relativity to Quantum Gravity*, 1408.4336.

[165] R. H. Kraichnan, *Special-relativistic derivation of generally covariant gravitation theory*, *Phys. Rev.* **98** (May, 1955) 1118–1122.

[166] S. Mukhi, *String theory: a perspective over the last 25 years*, *Class. Quant. Grav.* **28** (2011) 153001 [1110.2569].

[167] J. H. Schwarz and N. Seiberg, *String theory, supersymmetry, unification, and all that*, *Rev. Mod. Phys.* **71** (1999) S112–S120 [hep-th/9803179].

[168] J. Ellis, *Prospects for Discovering Supersymmetry at the LHC*, *Eur. Phys. J. C* **59** (2009) 335–343 [0810.1178].

[169] H. Baer, D. Sengupta, S. Salam, K. Sinha and V. Barger, *Midi-review: Status of weak scale supersymmetry after LHC Run 2 and ton-scale noble liquid WIMP searches*, 2002.03013.

[170] G. P. Salam, *Towards Jetography*, *Eur. Phys. J.* **C67** (2010) 637–686 [0906.1833].

[171] M. Cacciari, G. P. Salam and G. Soyez, *The anti-*$\mathrm{k_t}$ *jet clustering algorithm*, *JHEP* **04** (2008) 063 [0802.1189].

[172] "ATLAS Stand-Alone Event Displays." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayStandAlone, Aug, 2012.

[173] **ATLAS** Collaboration, G. Aad *et. al.*, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J.* **C77** (2017) 490 [1603.02934].

[174] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J.* **C77** (2017), no. 7 466 [1703.10485].

[175] J. E. Huth *et. al.*, *Toward a standardization of jet definitions*, in *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90) Snowmass, Colorado, June 25-July 13, 1990*, pp. 0134–136, 1990.

[176] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D48** (1993) 3160–3166 [hep-ph/9305266].

[177] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001 [hep-ph/9707323].

[178] "Jet 2012 Monte Carlo Event Displays." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApprovedBOOST2014EventDisplays, Aug, 2014.

[179] M. Cacciari and G. P. Salam, *Dispelling the* N^3 *myth for the* k_t *jet-finder*, *Phys. Lett. B* **641** (2006) 57–61 [hep-ph/0512210].

[180] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [1111.6097].

[181] **ATLAS** Collaboration, G. Aad *et. al.*, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, *JINST* **14** (2019), no. 12 P12006 [1908.00005].

[182] **ATLAS** Collaboration, G. Aad *et. al.*, *Performance of* b*-Jet Identification in the ATLAS Experiment*, *JINST* **11** (2016), no. 04 P04008 [1512.01094].

[183] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Measurements of b-jet tagging efficiency with the ATLAS detector using* $t\bar{t}$ *events at* $\sqrt{s} = 13$ *TeV*, *JHEP* **08** (2018) 089 [1805.01845].

[184] **ATLAS** Collaboration, G. Aad *et. al.*, *ATLAS b-jet identification performance and efficiency measurement with* $t\bar{t}$ *events in pp collisions at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J. C* **79** (2019), no. 11 970 [1907.05120].

[185] J. Shelton, *TASI Lectures on Jet Substructure*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales (TASI 2012): Boulder, Colorado, June 4-29, 2012*, pp. 303–340, 2013. 1302.0260.

[186] **ATLAS** Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using* pp *collisions at* $\sqrt{s} = 13$ TeV, Tech. Rep. ATLAS-CONF-2017-029, CERN, Geneva, May, 2017.

[187] **ATLAS** Collaboration, *Constituent-level pile-up mitigation techniques in ATLAS*, Tech. Rep. ATLAS-CONF-2017-065, CERN, Geneva, Aug, 2017.

[188] **OPAL** Collaboration, G. Alexander *et. al.*, *A Direct observation of quark - gluon jet differences at LEP*, *Phys. Lett.* **B265** (1991) 462–474.

[189] **ALEPH** Collaboration, R. Barate *et. al.*, *Measurements of the structure of quark and gluon jets in hadronic Z decays*, *Eur. Phys. J.* **C17** (2000) 1–18.

[190] **DELPHI** Collaboration, P. Abreu *et. al.*, *Energy dependence of the differences between the quark and gluon jet fragmentation*, *Z. Phys.* **C70** (1996) 179–196.

[191] **L3** Collaboration, M. Acciarri *et. al.*, *K(s)0 and Lambda production in quark and gluon jets at LEP*, *Phys. Lett.* **B407** (1997) 389–401. [Erratum: Phys. Lett.B427,409(1998)].

[192] **CDF** Collaboration, F. Abe *et. al.*, *A Measurement of jet shapes in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.8$ *TeV*, *Phys. Rev. Lett.* **70** (1993) 713–717.

[193] **D0** Collaboration, S. Abachi *et. al.*, *Transverse energy distributions within jets in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.8$ *TeV*, *Phys. Lett.* **B357** (1995) 500–508.

[194] **ZEUS** Collaboration, J. Breitweg *et. al.*, *Measurement of jet shapes in photoproduction at HERA*, *Eur. Phys. J.* **C2** (1998) 61–75 [hep-ex/9710002].

[195] **ZEUS** Collaboration, J. Breitweg *et. al.*, *Measurement of jet shapes in high* Q^2 *deep inelastic scattering at HERA*, *Eur. Phys. J.* **C8** (1999) 367–380 [hep-ex/9804001].

[196] **H1** Collaboration, C. Adloff *et. al.*, *Measurement of internal jet structure in dijet production in deep inelastic scattering at HERA*, *Nucl. Phys.* **B545** (1999) 3–20 [hep-ex/9901010].

[197] **CMS** Collaboration, *Jet algorithms performance in 13 TeV data*, Tech. Rep. CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[198] **CMS** Collaboration, *Performance of quark/gluon discrimination in 8 TeV pp data*, Tech. Rep. CMS-PAS-JME-13-002, CERN, Geneva, 2013.

[199] **CMS** Collaboration, *Performance of quark/gluon discrimination in 13 TeV data*, Tech. Rep. CMS-DP-2016-070, CERN, Geneva, Nov, 2016.

[200] **CMS** Collaboration, *New Developments for Jet Substructure Reconstruction in CMS*, Tech. Rep. CMS-DP-2017-027, CERN, Geneva, Jul, 2017.

[201] **ATLAS Collaboration** Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, Tech. Rep. ATLAS-CONF-2016-035, CERN, Geneva, Jul, 2016.

[202] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015 [1011.2268].

[203] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[204] **ALICE** Collaboration, K. Aamodt *et. al.*, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.

[205] R. E. Allen, *The Higgs Bridge*, *Phys. Scripta* **89** (2013) 018001 [1311.2647].

[206] "ATLAS Detector & Technology." https://atlas.cern/discover/detector.

[207] "CMS Detector." https://cms.cern/detector.

[208] "CERN Accelerator Complex." https://public-archive.web.cern.ch/en/research/AccelComplex-en.html.

[209] "CERN Accelerators." https://home.cern/science/accelerators.

[210] S. Humphries, *Principles of charged particle acceleration*. J. Wiley, 1986.

[211] J. D. Jackson, *Classical Electrodynamics*. Wiley, 3 ed., 1999.

[212] *LEP design report*. CERN, Geneva, 1984. Copies shelved as reports in LEP, PS and SPS libraries.

[213] S. Myers, *The LEP Collider, from design to approval and commissioning*. John Adams' Lecture. CERN, Geneva, 1991. Delivered at CERN, 26 Nov 1990.

[214] "The Large Electron-Positron Collider." https://home.cern/science/accelerators/large-electron-positron-collider.

[215] "CERN Facts and Figures." https://home.cern/resources/faqs/facts-and-figures-about-lhc.

[216] "Powering CERN." https://home.cern/science/engineering/powering-cern.

[217] BIPM, *Le Système international d'unités / The International System of Units ('The SI Brochure')*. Bureau international des poids et mesures, 9th ed., 2019.

[218] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Measurement of the total cross section from elastic scattering in* pp *collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, *Phys. Lett. B* **761** (2016) 158–178 [1607.06605].

[219] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* **117** (2016), no. 18 182002 [1606.02625].

[220] "Parton luminosity and cross section plots." http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html, 2012.

[221] E. Wilson, *An Introduction to Particle Accelerators*. Oxford University Press, 2001.

[222] B. Holzer, *Introduction to Particle Accelerators and their Limitations*, 1705.09601.

[223] J. Liouville, *Note sur la théorie de la variation des constantes arbitraires.*, *Journal de mathématiques pures et appliquées* (1838) 342–349.

[224] J. W. Gibbs, *On the fundamental formula of statistical mechanics, with applications to astronomy and thermodynamics*, in *Proceedings of the American Association for the Advancement of Science*, pp. 57–58, 1884.

[225] "Transverse Beam Dynamics." https://indico.cern.ch/event/528094/contributions/2213316/.

[226] G. Avoni *et. al.*, *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, *JINST* **13** (2018), no. 07 P07017.

[227] "The Large Hadron Collider." https://home.cern/science/accelerators/large-hadron-collider.

[228] M. Bajko, F. Bertinelli, N. Catalan-Lasheras, S. Claudet, P. Cruikshank, K. Dahlerup-Petersen, R. Denz, P. Fessia, C. Garion, J. Jimenez, G. Kirby, P. Lebrun, S. Le Naour, K.-H. Mess, M. Modena, V. Montabonnet, R. Nunes, V. Parma, A. Perin, G. de Rijk, A. Rijllart, L. Rossi, R. Schmidt, A. Siemko, P. Strubin, L. Tavian, H. Thiesen, J. Tock, E. Todesco, R. Veness, A. Verweij, L. Walckiers, R. Van Weelderen, R. Wolf, S. FehÃľr, R. Flora, M. Koratzinos,

P. Limon and J. Strait, *Report of the Task Force on the Incident of 19th September 2008 at the LHC*, Tech. Rep. LHC-PROJECT-Report-1168. CERN-LHC-PROJECT-Report-1168, CERN, Geneva, Mar, 2009.

[229] "Luminosity Public Results." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[230] "Luminosity Public Results Run 2." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[231] "LHC Commissioning - Longer Term LHC Schedule." http://lhc-commissioning.web.cern.ch/lhc-commissioning/schedule/LHC-long-term.htm.

[232] J. Boyd, *LHC Run-2 and Future Prospects*, in *2019 European School of High-Energy Physics*, 1, 2020. 2001.04370.

[233] "First report to the LMC on Run III machine configurations (p-p runs)." https://lhc-commissioning.web.cern.ch/2020/LCR3report_2019_1.pdf.

[234] "High-Luminosity LHC." https://home.cern/science/accelerators/high-luminosity-lhc.

[235] M. J. Swiatlowski, *Measuring the Standard Model and Searching for New Physics Using the ATLAS Detector*. PhD thesis, Stanford U., Phys. Dept., 2015.

[236] B. Nachman, *Investigating the Quantum Properties of Jets and the Search for a Supersymmetric Top Quark Partner with the ATLAS Detector*. PhD thesis, Stanford U., Phys. Dept., 2016. 1609.03242.

[237] J. Pequenao, "Event Cross Section in a computer generated image of the ATLAS detector.." Mar, 2008.

[238] *Track Reconstruction Performance of the ATLAS Inner Detector at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. ATL-PHYS-PUB-2015-018, CERN, Geneva, Jul, 2015.

[239] A. Lechner, *Particle interactions with matter*, *CERN Yellow Rep. School Proc.* **5** (2018) 47. 22 p.

[240] E. Halkiadakis, "Particle interactions with matter (part ii) and particle id." http://www.physics.rutgers.edu/~evahal/talks/tasi09/TASI_day3_school.pdf, Aug, 2009.

[241] H. Abreu *et. al.*, *Performance of the electronic readout of the ATLAS liquid argon calorimeters*, *JINST* **5** (2010) P09003.

[242] **ATLAS** Collaboration, A. Solodkov, *Performance of the ATLAS Tile Calorimeter*, Tech. Rep. ATL-TILECAL-PROC-2015-024, CERN, Geneva, Dec, 2015.

[243] G. Luisoni, "An introduction to powheg." https://indico.cern.ch/event/656211/contributions/2673382/attachments/1498653/2333153/luisoni_powheg.pdf, Mar, 2017.

[244] T. Sjöstrand, S. Mrenna and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867 [0710.3820].

[245] T. SjÃűstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen and P. Z. Skands, *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177 [1410.3012].

[246] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert and J. Winter, *Event generation with sherpa 1.1*, *JHEP* **02** (2009) 007 [0811.4622].

[247] M. Bahr *et. al.*, *Herwig++ Physics and Manual*, *Eur. Phys. J.* **C58** (2008) 639–707 [0803.0883].

[248] J. Bellm *et. al.*, *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J.* **C76** (2016), no. 4 196 [1512.01178].

[249] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [0709.2092].

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

[251] **ATLAS** Collaboration, *ATLAS Pythia 8 tunes to 7 TeV datas*, Tech. Rep. ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov, 2014.

[252] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin and C. P. Yuan, *New parton distributions for collider physics*, *Phys. Rev. D* **82** (2010) 074024 [1007.2241].

[253] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152–155.

[254] **ATLAS** Collaboration, Z. Marshall, *Simulation of Pile-up in the ATLAS Experiment*, *J. Phys. Conf. Ser.* **513** (2014) 022024.

[255] **GEANT4** Collaboration, S. Agostinelli *et. al.*, *GEANT4: a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[256] "Computing and Software - Public Results." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ComputingandSoftwarePublicResults.

[257] "HEPiX Benchmarking Working Group." https://w3.hepix.org/benchmarking.html.

[258] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Performance of the ATLAS Trigger System in 2015*, *Eur. Phys. J. C* **77** (2017), no. 5 317 [1611.09661].

[259] **ATLAS** Collaboration, A. Martinez, *The Run-2 ATLAS Trigger System*, *J. Phys. Conf. Ser.* **762** (2016), no. 1 012003.

[260] **ATLAS** Collaboration, *Operation of the ATLAS trigger system in Run 2*, Tech. Rep. ATLAS-TRIG-2019-04-001, CERN, Geneva, May, 2020. [Publication Draft, Not Public].

[261] **ATLAS** Collaboration, A. Buckley, T. Eifert, M. Elsing, D. I. Gillberg, K. Koeneke, A. Krasznahorkay, E. Moyse, M. Nowak, S. Snyder and P. van Gemmeren, *Implementation of the ATLAS Run 2 event data model*, Tech. Rep. ATL-SOFT-PROC-2015-003. 7, CERN, Geneva, May, 2015.

[262] "Trigger Operation Public Results." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults.

[263] **ATLAS** Collaboration, *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, Tech. Rep. ATL-DAQ-PUB-2016-001, CERN, Geneva, Mar, 2016.

[264] **ATLAS** Collaboration, *Trigger Menu in 2016*, Tech. Rep. ATL-DAQ-PUB-2017-001, CERN, Geneva, Jan, 2017.

[265] **ATLAS** Collaboration, *Trigger Menu in 2017*, Tech. Rep. ATL-DAQ-PUB-2018-002, CERN, Geneva, Jun, 2018.

[266] **ATLAS** Collaboration, *Trigger menu in 2018*, Tech. Rep. ATL-DAQ-PUB-2019-001, CERN, Geneva, Oct, 2019.

[267] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for low-mass dijet resonances using trigger-level jets with the ATLAS detector in* pp *collisions at* $\sqrt{s} = 13$ *TeV*, *Phys. Rev. Lett.* **121** (2018), no. 8 081801 [1804.03496].

[268] M. Shochet, L. Tompkins, V. Cavaliere, P. Giannetti, A. Annovi and G. Volpi, *Fast TracKer (FTK) Technical Design Report*, Tech. Rep. CERN-LHCC-2013-007. ATLAS-TDR-021, Jun, 2013. ATLAS Fast Tracker Technical Design Report.

[269] **ATLAS Collaboration** Collaboration, *Characterization of Interaction-Point Beam Parameters Using the pp Event-Vertex Distribution Reconstructed in the*

ATLAS Detector at the LHC, Tech. Rep. ATLAS-CONF-2010-027, CERN, Geneva, May, 2010.

[270] A. Cukierman and R. Bartoldus, "Bayesian Inference for Beamspot Width Measurement." https://indico.cern.ch/event/802607/#3-weekly-presentation, Mar, 2019.

[271] **ATLAS** Collaboration, G. Aad *et. al.*, *Triggers for displaced decays of long-lived neutral particles in the ATLAS detector*, *JINST* **8** (2013) P07015 [1305.2284].

[272] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017), no. 10 673 [1704.07983].

[273] A. Collaboration, "ATLAS event at 900 GeV - 6 May 2015 - Run 264034 Evt 11475271." General Photo, May, 2015.

[274] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth. A* **262** (1987) 444–450.

[275] **ATLAS** Collaboration, G. Aad *et. al.*, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, *JINST* **9** (2014) P09009 [1406.7690].

[276] **ATLAS** Collaboration, G. Aad *et. al.*, *The ATLAS Inner Detector commissioning and calibration*, *Eur. Phys. J. C* **70** (2010) 787–821 [1004.5293].

[277] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017), no. 5 332 [1611.10235].

[278] **ATLAS** Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. ATL-PHYS-PUB-2015-026, CERN, Geneva, Jul, 2015.

[279] **ATLAS** Collaboration, F. Meloni, *Primary vertex reconstruction with the ATLAS detector*, *JINST* **11** (2016), no. 12 C12060.

[280] "Public Liquid-Argon Calorimeter Plots on Upgrade." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LArCaloPublicResultsUpgrade.

[281] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at* $\sqrt{s}$ *= 13 TeV*, *Eur. Phys. J. C* **79** (2019), no. 8 639 [1902.04655].

[282] **ATLAS** Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, Tech. Rep. ATLAS-CONF-2016-024, CERN, Geneva, Jun, 2016.

[283] **ATLAS** Collaboration, ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, *Eur. Phys. J. C* **76** (2016), no. 12 666 [1606.01813].

[284] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, *Eur. Phys. J. C* **79** (2019), no. 3 205 [1810.05087].

[285] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data*, *JINST* **14** (2019), no. 03 P03017 [1812.03848].

[286] **ATLAS** Collaboration, G. Aad *et. al.*, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, *Eur. Phys. J. C* **74** (2014), no. 10 3071 [1407.5063].

[287] **ATLAS** Collaboration, *Improving jet substructure performance in ATLAS using Track-CaloClusters*, Tech. Rep. ATL-PHYS-PUB-2017-015, CERN, Geneva, Jul, 2017.

[288] G. Lejeune Dirichlet, *Über die reduction der positiven quadratischen formen mit drei unbestimmten ganzen zahlen*, *Journal für die reine und angewandte Mathematik* **40** (1850) 209–227.

[289] G. Voronoi, *Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. sur quelques propriétés des formes quadratiques positives parfaites*, *Journal für die reine und angewandte Mathematik* **133** (1908) 97–178.

[290] S. Fortune, *A sweepline algorithm for voronoi diagrams*, in *Proceedings of the Second Annual Symposium on Computational Geometry*, SCG âĂŹ86, (New York, NY, USA), p. 313âĂŞ322, Association for Computing Machinery, 1986.

[291] **ATLAS** Collaboration, *Jet energy scale and uncertainties in 2015-2017 data and simulation*, Tech. Rep. JETM-2018-006, CERN, Geneva, Nov, 2018.

[292] **ATLAS** Collaboration, *Simultaneous Jet Energy and Mass Calibrations with Neural Networks*, Tech. Rep. ATL-PHYS-PUB-2020-001, CERN, Geneva, Jan, 2020.

[293] B. T. Elder and J. Thaler, *Aspects of Track-Assisted Mass*, *JHEP* **03** (2019) 104 [1805.11109].

[294] *Measurement of large radius jet mass reconstruction performance at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Tech. Rep. ATLAS-CONF-2016-008, CERN, Geneva, Mar, 2016.

[295] "Plots of Njets vs Pile-Up in 2017 data." https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2017-009/.

[296] **ATLAS** Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2014-018, CERN, Geneva, May, 2014.

[297] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski,

S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. K. Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter and A. Zemla, *Tmva - toolkit for multivariate data analysis*, 2007.

[298] **ATLAS** Collaboration, *Optimisation and performance studies of the ATLAS* b*-tagging algorithms for the 2017-18 LHC run*, Tech. Rep. ATL-PHYS-PUB-2017-013, CERN, Geneva, Jul, 2017.

[299] **ATLAS** Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, Tech. Rep. ATL-PHYS-PUB-2017-011, CERN, Geneva, Jun, 2017.

[300] **ATLAS** Collaboration, *Topological* b*-hadron decay reconstruction and identification of* b*-jets with the JetFitter package in the ATLAS experiment at the LHC*, Tech. Rep. ATL-PHYS-PUB-2018-025, CERN, Geneva, Oct, 2018.

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

[302] **ATLAS** Collaboration, G. Aad *et. al.*, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in* pp *collisions at* $\sqrt{s}$*=8 TeV*, *Eur. Phys. J. C* **75** (2015), no. 7 303 [1412.7086].

[303] **ATLAS** Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, Tech. Rep. ATL-PHYS-PUB-2015-045, CERN, Geneva, Nov, 2015.

[304] **ATLAS** Collaboration, G. Aad *et. al.*, *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, *Eur. Phys. J. C* **76** (2016), no. 5 295 [1512.05955].

[305] **ATLAS** Collaboration, G. Aad *et. al.*, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, *Eur. Phys. J. C* **74** (2014), no. 11 3130 [1407.3935].

[306] **ATLAS** Collaboration, G. Aad *et. al.*, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s}$ *=13 TeV*, *Eur. Phys. J. C* **76** (2016), no. 5 292 [1603.05598].

[307] "Event Displays from Run 2 physics analyses." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun2Physics, April, 2020.

[308] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for the dimuon decay of the Higgs boson in* pp *collisions at* $\sqrt{s}$ *= 13 TeV with the ATLAS detector*, *Phys. Rev. Lett.* **119** (2017), no. 5 051802 [1705.04582].

[309] J. Illingworth and J. Kittler, *A survey of the hough transform*, *Computer Vision, Graphics, and Image Processing* **44** (1988), no. 1 87 – 116.

[310] **ATLAS** Collaboration, G. Aad *et. al.*, *Performance of algorithms that reconstruct missing transverse momentum in* $\sqrt{s}$ *= 8 TeV proton-proton collisions in the ATLAS detector*, *Eur. Phys. J. C* **77** (2017), no. 4 241 [1609.09324].

[311] **ATLAS** Collaboration, M. Aaboud *et. al.*, *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), no. 11 903 [1802.08168].

[312] **ATLAS** Collaboration, E_T^{miss} *performance in the ATLAS detector using 2015-2016 LHC p-p collisions*, Tech. Rep. ATLAS-CONF-2018-023, CERN, Geneva, Jun, 2018.

[313] **ATLAS** Collaboration, G. Aad *et. al.*, *Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS*, *Eur. Phys. J. C* **72** (2012) 1844 [1108.5602].

[314] **ATLAS** Collaboration, *Object-based missing transverse momentum significance in the ATLAS detector*, Tech. Rep. ATLAS-CONF-2018-038, CERN, Geneva, Jul, 2018.

[315] K. Fukushima, *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*, *Biological cybernetics* **36** (1980), no. 4 193–202.

[316] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, *Gradient-based learning applied to document recognition*, *Proceedings of the IEEE* **86** (1998), no. 11 2278–2324.

[317] J. Schmidhuber, *Deep learning in neural networks: An overview*, *Neural networks* **61** (2015) 85–117.

[318] **ATLAS** Collaboration, *Convolutional Neural Networks with Event Images for Pileup Mitigation with the ATLAS Detector*, Tech. Rep. ATL-PHYS-PUB-2019-028, CERN, Geneva, Jul, 2019.

[319] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for Higgs boson decays into pairs of light (pseudo)scalar particles in the* $\gamma\gamma$jj *final state in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Phys. Lett.* **B782** (2018) 750–767 [1803.11145].

[320] **ATLAS, CMS** Collaboration, ATLAS Collaboration, CMS Collaboration, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at* $\sqrt{s} = 7$ *and 8 TeV*, *JHEP* **08** (2016) 045 [1606.02266].

[321] V. Silveira and A. Zee, *Scalar phantoms*, *Physics Letters B* **161** (1985), no. 1 136 – 140.

[322] M. Pospelov, A. Ritz and M. B. Voloshin, *Secluded WIMP Dark Matter*, *Phys. Lett.* **B662** (2008) 53–61 [0711.4866].

[323] P. Draper, T. Liu, C. E. M. Wagner, L.-T. Wang and H. Zhang, *Dark Light Higgs*, *Phys. Rev. Lett.* **106** (2011) 121805 [1009.3963].

[324] S. Ipek, D. McKeen and A. E. Nelson, *A Renormalizable Model for the Galactic Center Gamma Ray Excess from Dark Matter Annihilation*, *Phys. Rev.* **D90** (2014), no. 5 055021 [1404.3716].

[325] A. Martin, J. Shelton and J. Unwin, *Fitting the Galactic Center Gamma-Ray Excess with Cascade Annihilations*, *Phys. Rev.* **D90** (2014), no. 10 103513 [1405.0272].

[326] S. Profumo, M. J. Ramsey-Musolf and G. Shaughnessy, *Singlet Higgs phenomenology and the electroweak phase transition*, *JHEP* **08** (2007) 010 [0705.2425].

[327] N. Blinov, J. Kozaczuk, D. E. Morrissey and C. Tamarit, *Electroweak Baryogenesis from Exotic Electroweak Symmetry Breaking*, *Phys. Rev.* **D92** (2015), no. 3 035012 [1504.05195].

[328] G. Burdman, Z. Chacko, H.-S. Goh and R. Harnik, *Folded supersymmetry and the LEP paradox*, *JHEP* **02** (2007) 009 [hep-ph/0609152].

[329] N. Craig, A. Katz, M. Strassler and R. Sundrum, *Naturalness in the Dark at the LHC*, *JHEP* **07** (2015) 105 [1501.05310].

[330] D. Curtin and C. B. Verhaaren, *Discovering Uncolored Naturalness in Exotic Higgs Decays*, *JHEP* **12** (2015) 072 [1506.06141].

[331] **CMS** Collaboration, CMS Collaboration, *Search for light bosons in decays of the 125 GeV Higgs boson in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*, 1701.02032.

[332] **ATLAS** Collaboration, ATLAS Collaboration, *Search for new light gauge bosons in Higgs boson decays to four-lepton final states in* pp *collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector at the LHC*, *Phys. Rev.* **D92** (2015), no. 9 092001 [1505.07645].

[333] **ATLAS** Collaboration, ATLAS Collaboration, *Search for Higgs bosons decaying to aa in the $\mu\mu\tau\tau$ final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS experiment*, *Phys. Rev.* **D92** (2015), no. 5 052002 [1505.01609].

[334] A. Martin, *Higgs Cascade Decays to gamma gamma + jet jet at the LHC*, hep-ph/0703247.

[335] **ATLAS** Collaboration, ATLAS Collaboration, *Search for new phenomena in events with at least three photons collected in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C76** (2016), no. 4 210 [1509.05051].

[336] **LHC Higgs Cross Section Working Group** Collaboration, S. Dittmaier *et. al.*, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, 1101.0593.

[337] **LHC Higgs Cross Section Working Group** Collaboration, D. de Florian *et. al.*, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 2016.

[338] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [hep-ph/0409146].

[339] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09** (2014) [1406.3660].

[340] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823 [1005.4568].

[341] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J. C* **63** (2009) 189–285 [0901.0002].

[342] ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, *Eur. Phys. J. C* **73** (2013) [1210.6210].

[343] **LHC Higgs Cross Section Working Group** Collaboration, J. R. Andersen *et. al.*, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, 2013.

[344] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, *Comput. Phys. Commun.* **10** (1975) 343–367.

[345] A. L. Read, *Presentation of search results: The* $\mathsf{CL_s}$ *technique*, *J. Phys. G* **28** (2002) 2693–2704.

[346] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J.* **C71** (2011) 1554 [1007.1727]. [Erratum: Eur. Phys. J.C73,2501(2013)].

[347] **ATLAS** Collaboration, A. Collaboration, *A search for pairs of highly collimated photon-jets in* pp *collisions at* $\sqrt{\mathsf{s}}$ *= 13 TeV with the ATLAS detector*, *Phys. Rev.* **D99** (2019), no. 1 012008 [1808.10515].

[348] **ATLAS** Collaboration, *Search for a Higgs boson decaying to four photons through light CP-odd scalar coupling using 4.9fb*$^{-1}$ *of 7 TeV* pp *collision data taken with ATLAS detector at the LHC*, Tech. Rep. ATLAS-CONF-2012-079, CERN, Geneva, Jul, 2012.

[349] A. Cukierman and B. Nachman, *Mathematical Properties of Numerical Inversion for Jet Calibrations*, *Nucl. Instrum. Meth.* **A858** (2017) 1–11 [1609.05195].

[350] **ATLAS** Collaboration, *Generalized Numerical Inversion: A Neural Network Approach to Jet Calibration*, Tech. Rep. ATL-PHYS-PUB-2018-013, CERN, Geneva, Jul, 2018.

[351] **ATLAS** Collaboration, G. Aad *et. al.*, *Dijet resonance search with weak supervision using* $\sqrt{\mathsf{s}} = 13$ *TeV* pp *collisions in the ATLAS detector*, 2005.02983.

[352] **ATLAS** Collaboration, G. Aad *et. al.*, *Search for New Particles in Two-Jet Final States in 7 TeV Proton-Proton Collisions with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* **105** (2010) 161801 [1008.2461].

[353] **ATLAS** Collaboration, G. Aad *et. al.*, *Search for New Physics in Dijet Mass and Angular Distributions in pp Collisions at $\sqrt{s} = 7$ TeV Measured with the ATLAS Detector*, *New J. Phys.* **13** (2011) 053044 [1103.3864].

[354] **ATLAS** Collaboration, G. Aad *et. al.*, *Search for New Physics in the Dijet Mass Distribution using 1 fb^{-1} of pp Collision Data at $\sqrt{s}$ =7 TeV collected by the ATLAS Detector*, *Phys. Lett.* **B708** (2012) 37–54 [1108.6311].

[355] **ATLAS** Collaboration, G. Aad *et. al.*, *Search for new phenomena in the dijet mass distribution using $p - p$ collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Phys. Rev.* **D91** (2015), no. 5 052007 [1407.1376].

[356] **ATLAS** Collaboration, G. Aad *et. al.*, *Search for new phenomena in dijet mass and angular distributions from pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, *Phys. Lett.* **B754** (2016) 302–322 [1512.01530].

[357] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector*, *Phys. Rev.* **D96** (2017), no. 5 052004 [1703.09127].

[358] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for Minimal Supersymmetric Standard Model Higgs bosons H/A and for a Z′ boson in the $\tau\tau$ final state produced in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, *Eur. Phys. J.* **C76** (2016), no. 11 585 [1608.00890].

[359] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **01** (2018) 055 [1709.07242].

[360] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for resonances in the mass distribution of jet pairs with one or two jets identified as* b*-jets in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Phys. Lett.* **B759** (2016) 229–246 [1603.08791].

[361] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for resonances in the mass distribution of jet pairs with one or two jets identified as* b*-jets in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Phys. Rev.* **D98** (2018) 032016 [1805.09299].

[362] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Eur. Phys. J.* **C78** (2018), no. 7 565 [1804.10823].

[363] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for* $W' \to tb$ *decays in the hadronic final state using pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Phys. Lett.* **B781** (2018) 327–348 [1801.07893].

[364] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Searches for heavy diboson resonances in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *JHEP* **09** (2016) 173 [1606.04833].

[365] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for* WW/WZ *resonance production in* $\ell\nu qq$ *final states in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *JHEP* **03** (2018) 042 [1710.07235].

[366] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Searches for heavy* ZZ *and* ZW *resonances in the* $\ell\ell qq$ *and* $\nu\nu qq$ *final states in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *JHEP* **03** (2018) 009 [1708.09638].

[367] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for diboson resonances with boson-tagged jets in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Phys. Lett.* **B777** (2018) 91–113 [1708.04445].

[368] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for a heavy Higgs boson decaying into a* Z *boson and another heavy Higgs boson in the* $\ell\ell\mathsf{bb}$ *final state in* pp *collisions at* $\sqrt{\mathsf{s}} = 13$ *TeV with the ATLAS detector*, *Phys. Lett.* **B783** (2018) 392–414 [1804.01126].

[369] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for heavy resonances decaying into a* W *or* Z *boson and a Higgs boson in final states with leptons and* b*-jets in 36* fb^{-1} *of* $\sqrt{\mathsf{s}} = 13$ *TeV* pp *collisions with the ATLAS detector*, *JHEP* **03** (2018) 174 [1712.06518]. [Erratum: JHEP11,051(2018)].

[370] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for heavy resonances decaying to a* W *or* Z *boson and a Higgs boson in the* $\mathsf{q}\bar{\mathsf{q}}^{(\prime)}\mathsf{b}\bar{\mathsf{b}}$ *final state in* pp *collisions at* $\sqrt{\mathsf{s}} = 13$ *TeV with the ATLAS detector*, *Phys. Lett.* **B774** (2017) 494–515 [1707.06958].

[371] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for pair production of Higgs bosons in the* $\mathsf{b}\bar{\mathsf{b}}\mathsf{b}\bar{\mathsf{b}}$ *final state using proton-proton collisions at* $\sqrt{\mathsf{s}} = 13$ *TeV with the ATLAS detector*, *JHEP* **01** (2019) 030 [1804.06174].

[372] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton-proton collisions at* $\sqrt{\mathsf{s}} = 13$ *TeV with the ATLAS detector*, *Phys. Lett.* **B788** (2019) 316–335 [1801.08769].

[373] ATLAS Collaboration, *Search for low-mass resonances decaying into two jets and produced in association with a photon using* pp *collisions at* $\sqrt{\mathsf{s}} = 13$ *TeV with the ATLAS detector*, *Phys. Lett.* **B795** (2019) 56–75 [1901.10917].

[374] N. Craig, P. Draper, K. Kong, Y. Ng and D. Whiteson, *The unexplored landscape of two-body resonances*, *Acta Phys. Polon.* **B50** (2019) 837 [1610.09392].

[375] J. H. Kim, K. Kong, B. Nachman and D. Whiteson, *The motivation and status of two-body resonance decays after the LHC Run 2 and beyond*, 1907.06659.

[376] **ATLAS** Collaboration, M. Aaboud *et. al.*, *A search for resonances decaying into a Higgs boson and a new particle* X *in the* XH → qqbb *final state with the ATLAS detector*, *Phys. Lett.* **B779** (2018) 24–45 [1709.06783].

[377] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Search for long-lived particles produced in* pp *collisions at* $\sqrt{s} = 13$ *TeV that decay into displaced hadronic jets in the ATLAS muon spectrometer*, *Submitted to: Phys. Rev.* (2018) [1811.07370].

[378] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175].

[379] R. D. Ball *et. al.*, *Parton distributions with LHC data*, *Nucl. Phys.* **B867** (2013) 244–289 [1207.1303].

[380] L. M. Dery, B. Nachman, F. Rubbo and A. Schwartzman, *Weakly Supervised Classification in High Energy Physics*, *JHEP* **05** (2017) 145 [1702.00414].

[381] T. Cohen, M. Freytsis and B. Ostdiek, *(Machine) Learning to Do More with Less*, *JHEP* **02** (2018) 034 [1706.09451].

[382] P. T. Komiske, E. M. Metodiev, B. Nachman and M. D. Schwartz, *Learning to classify from impure samples with high-dimensional data*, *Phys. Rev.* **D98** (2018), no. 1 011502 [1801.10158].

[383] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* **784** (2018) 173–191 [1806.00425].

[384] R. T. D'Agnolo and A. Wulzer, *Learning New Physics from a Machine*, *Phys. Rev.* **D99** (2019), no. 1 015014 [1806.02350].

[385] R. T. D'Agnolo, G. Grosso, M. Pierini, A. Wulzer and M. Zanetti, *Learning Multivariate New Physics*, 1912.12155.

[386] G. M. Alessandro Casa, *Nonparametric semisupervised classification for signal detection in high energy physics*, 1809.02977.

[387] A. De Simone and T. Jacques, *Guiding New Physics Searches with Unsupervised Learning*, *Eur. Phys. J.* **C79** (2019), no. 4 289 [1807.06038].

[388] A. Mullin, H. Pacey, M. Parker, M. White and S. Williams, *Does SUSY have friends? A new approach for LHC event analysis*, 1912.10625.

[389] J. Aguilar-Saavedra, J. H. Collins and R. K. Mishra, *A generic anti-QCD jet tagger*, *JHEP* **11** (2017) 163 [1709.01087].

[390] S. Adorni, N. Calace, K. Iordanidou, R. Jansky, E. Kajomovitz, B. Murray, H. R. Nindhito, T. Nobe, Y. Okazaki and S. Schramm, *Search for diboson resonances with jets in 140 fb^{-1} of* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Tech. Rep. ATL-COM-PHYS-2018-1580, CERN, Geneva, Nov, 2018.

[391] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

[392] D. Kriesel, *A Brief Introduction to Neural Networks*. 2007.

[393] C. M. Bishop, *Neural networks for pattern recognition*. Oxford University Press, 2013.

[394] R. D. Reed and R. J. Marks, *Neural smithing: supervised learning in feedforward artificial neural networks*. CogNet, 1999.

[395] I. Goodfellow, Y. Bengio and A. Courville, *Deep learning*. The MIT Press, 2017.

[396] A. Ng, "CS229 Lecture notes." http://cs229.stanford.edu/notes/.

[397] J. H. Friedman, *Greedy function approximation: a gradient boosting machine*, *Annals of statistics* (2001) 1189–1232.

[398] J. H. Friedman, *Stochastic gradient boosting*, *Computational statistics & data analysis* **38** (2002), no. 4 367–378.

[399] J. Friedman, T. Hastie, R. Tibshirani *et. al.*, *Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)*, *The annals of statistics* **28** (2000), no. 2 337–407.

[400] J. Friedman, T. Hastie and R. Tibshirani, *The elements of statistical learning*, vol. 1. Springer series in statistics New York, 2001.

[401] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014.

[402] "Jet Uncertainties 2019." https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetUncertaintiesRel21Summer2019LargeR.

[403] "Luminosity for Physics." https://twiki.cern.ch/twiki/bin/viewauth/Atlas/LuminosityForPhysics#2018_13_TeV_proton_proton_Morion.

[404] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Luminosity determination in pp collisions at* $\sqrt{s}$ *= 8 TeV using the ATLAS detector at the LHC*, *Eur. Phys. J.* **C76** (2016), no. 12 653 [1608.03953].

[405] **UA2** Collaboration, J. Alitti *et. al.*, *A Measurement of two jet decays of the* W *and* Z *bosons at the CERN* $\bar{p}p$ *collider*, *Z. Phys.* **C49** (1991) 17–28.

[406] D. Pappadopulo, A. Thamm, R. Torre and A. Wulzer, *Heavy Vector Triplets: Bridging Theory and Data*, *JHEP* **09** (2014) 060 [1402.4431].

[407] J. Cogan, M. Kagan, E. Strauss and A. Schwartzman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **02** (2015) 118 [1407.5675].

[408] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069 [1511.05190].

[409] A. Schwartzman, *Jet energy calibration at the LHC*, *Int. J. Mod. Phys. A* **30** (2015), no. 31 1546002 [1509.05459].

[410] CMS Collaboration, *Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS*, *JINST* **6** (2011) [1107.4277].

[411] CMS Collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, *JINST* **12** (2017) [1607.03663].

[412] D. Lopez Mateos, E. W. Hughes and A. Schwartzman, *A Simple* p_T*- and* η*-Dependent Monte Carlo-Based Jet Calibration*, Tech. Rep. ATL-COM-PHYS-2009-076, CERN, Geneva, Feb, 2009.

[413] **ATLAS** Collaboration, M. Aaboud *et. al.*, *Determination of jet calibration and energy resolution in proton-proton collisions at* $\sqrt{s}$ *= 8 TeV using the ATLAS detector*, 1910.04482.

[414] **ATLAS** Collaboration, *Jet global sequential corrections with the ATLAS detector in proton-proton collisions at sqrt(s) = 8 TeV*, Tech. Rep. ATLAS-CONF-2015-002, CERN, Geneva, Mar, 2015.

[415] K. Hornik, *Approximation capabilities of multilayer feedforward networks*, *Neural Networks* **4** (1991), no. 2 251 – 257.

[416] **ATLAS** Collaboration, *Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high pT jets*, Tech. Rep. ATL-PHYS-PUB-2014-008, CERN, Geneva, Jun, 2014.

[417] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, *Psychological Review* (1958) 65–386.

[418] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, *Mathematics of Control, Signals and Systems* **2** (Dec, 1989) 303–314.

[419] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *Scikit-learn:*

Machine learning in Python, *Journal of Machine Learning Research* **12** (2011) 2825–2830.

[420] S. Cheong, A. Cukierman, B. Nachman, M. Safdari and A. Schwartzman, *Parametrizing the Detector Response with Neural Networks*, *JINST* **15** (2020), no. 01 P01030 [1910.03773].

[421] A. Collaboration, *The ATLAS level-1 calorimeter trigger*, *JINST* **3** (2008) P03001.

[422] **ATLAS** Collaboration, A. Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, *Eur. Phys. J.* **C80** (2020), no. 1 47 [1909.00761].

[423] **ATLAS** Collaboration, A. Collaboration, *Search for Higgs bosons produced via vector-boson fusion and decaying into bottom quark pairs in* $\sqrt{s} = 13$ TeV pp *collisions with the ATLAS detector*, *Phys. Rev.* **D98** (2018), no. 5 052003 [1807.08639].

[424] "Lowest un-prescaled triggers per data-taking period." https://twiki.cern.ch/twiki/bin/view/Atlas/LowestUnprescaled, Sep, 2019.

[425] **CMS** Collaboration, V. Khachatryan *et. al.*, *Identification techniques for highly boosted W bosons that decay into hadrons*, *JHEP* **12** (2014) 017 [1410.4227].

[426] S. Cheong and J. Pearkes, "Merged Photon ID." https://indico.cern.ch/event/747068/contributions/3089187/, Jul, 2018.

[427] R. G. Miller, *Simultaneous statistical inference*. McGraw-Hill, 1966.

[428] R. Mittelhammer, G. G. Judge and D. Miller, *Econometric foundations*. Cambridge University Press, 2000.

[429] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, *Eur. Phys. J. C* **70** (2010) 525–530 [1005.1891].

[430] C. E. Bonferroni, *Teoria statistica delle classi e calcolo delle probabilità.*, .

[431] O. J. Dunn, *Estimation of the medians for dependent variables*, *Ann. Math. Statist.* **30** (03, 1959) 192–197.

[432] O. J. Dunn, *Multiple comparisons among means*, *Journal of the American Statistical Association* **56** (1961), no. 293 52–64 [https://www.tandfonline.com/doi/pdf/10.1080/01621459.1961.10482090].

[433] L. Demortier, *P values: What they are and how to use them*, .

[434] L. Lyons, *Open statistical issues in Particle Physics*, *Ann. Appl. Stat.* **2** (2008) 887–915.

[435] L. Lyons and N. Wardle, *Statistical issues in searches for new phenomena in High Energy Physics*, *J. Phys. G* **45** (2018), no. 3 033001.

[436] G. Cowan, *Statistical data analysis*. Clarendon Press, 2004.

[437] R. J. Barlow, *Statistics a guide to the use of statistical methods in the physical sciences*. Wiley, 2002.

[438] G. Bohm and G. Zech, *Introduction to statistics and data analysis for physicists*, vol. 1. Desy Hamburg, 2010.

[439] L. Lyons, *Discovering the Significance of 5 sigma*, 1310.1284.

[440] P. K. Sinervo, *Signal significance in particle physics*, in *Conference on Advanced Statistical Techniques in Particle Physics*, pp. 64–76, 6, 2002. hep-ex/0208005.

[441] J. Neyman and E. S. Pearson, *Ix. on the problem of the most efficient tests of statistical hypotheses*, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **231** (1933), no. 694-706 289–337.

[442] J. P. Egan, *Signal detection theory and ROC-analysis*. Academic press, 1975.

[443] W. Verkerke and D. Kirkby, *The roofit toolkit for data modeling*, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, pp. 186–189. World Scientific, 2006.

[444] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke and M. Wolf, *The RooStats Project*, *PoS* **ACAT2010** (2010) 057 [1009.1003].

[445] **ROOT Collaboration** Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Tech. Rep. CERN-OPEN-2012-016, New York U., New York, Jan, 2012.

[446] B. Nachman and T. Rudelius, *Evidence for conservatism in LHC SUSY searches*, *Eur. Phys. J. Plus* **127** (2012) 157 [1209.3522].