ATLAS NOTE

May 27, 2015

Run II Analysis Framework and Intial Validation Studies for $H \rightarrow ZZ^* \rightarrow 4\ell$ Analysis

Syed Haider Abidi[a] and Richard Teuscher[b]

[a]998156819, *University of Toronto*
[b]*Supervisor, University of Toronto*

Abstract

This undergraduate thesis focuses on the development of a user analysis framework for the ATLAS Run 2 $H \rightarrow ZZ^* \rightarrow 4\ell$ analysis. The Run 1 analysis model is investigated and requirements and constraints for a new model are derived. Based on these and the new ATLAS software upgrades, the design of a new code base is outlined and implemented. Initial validation studies using this framework are also presented.

Acknowledgments

First and foremost, I would like to thank my supervisor, Richard Teuscher, who has been tirelessly mentoring me since the day I started working in ATLAS. His perfect balance between a gentle prod and letting me figure things out independently have been instrumental in teaching me how to do ‘physics’.

Also, I would like to thank Fabien Tarrade, whose guidance throughout the analysis has been more valuable than imaginable. Without his unwavering support and belief in my abilities, this thesis and all the work before it would have not been possible.

I owe a huge debt of gratitude to RS Orr, Sarah Torrie and Robert Schweitzer for setting me on a path of research and physics. Without them, I would not be standing here.

Also, in no particular order, I would like to thank Ludovica, Eleni, RD, Roberto and Rosy for helping me find my sanity through the crazy times.

And thank you to Yubo, Daniel, Faizan and my family for knowing when to be simply there.

Contents

List of Figures

List of Tables

1 Introduction

The Standard Model (SM) is a theoretical framework, developed by physicists in the 20^{th} century, that describes the elementary particles of Nature and their interaction with each other. The SM contains three of the four known fundamental forces, namely: electromagnetism, the weak force and the strong force, with gravity being too weak to have a discernible effect on any phenomenon related to elementary particles. Concurrently with the theoretical developments, several experiments have been carried out which tested the validity of this model across many energy regimes which the SM has confirmed. It is regarded as one of the most precisely tested theories as many of its predictions have been tested to multiple decimal places.

An important aspect of the SM is the Higgs mechanism, introduced by Peter Higgs and et al (see Section 2.1.3). Without this mechanism, the introduction of mass for the bosons breaks gauge invariance, the property from which the interactions between particles are derived. On the other hand, the Higgs mechanism provides a consistent method for the gauge bosons to gain mass while preserving the necessary symmetries. Furthermore, it also mitigates other problems in the SM, such as WW scattering violating unitarity at high energy.

An experimental prediction of this mechanism is the existence of a fundamental scalar boson, known as the Higgs boson. For the last few decades, all experiments have failed to find any evidence for this particle except until 2012. In Run 1 of the Large Hadron Collider (LHC), ATLAS and CMS, the two general purpose detectors, collected sufficient data at $\sqrt{s}$ = 7 & 8 TeV to claim the discovery of a Higgs boson with a mass of approximately 125 GeV. These results were mainly driven by the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channels. Despite the low statistics, $H \rightarrow ZZ^* \rightarrow 4\ell$ is typically referred to as 'the Golden Channel' due to its high signal to noise ratio and fully reconstructible final state.

Currently the LHC and ATLAS are undergoing upgrades to operate at $\sqrt{s}$ = 13 TeV and at a higher instantaneous luminosity for Run 2. These changes will allow properties of the Higgs boson to be measured at a higher accuracy. Furthermore with these upgrades, searches for other Higgs bosons will probe previously unexplored parts of the parameter space. These measurements are needed to rule out or constrain Beyond SM (BSM) theories and pin down the exact nature of the newly found particle.

This thesis documents the work performed for the Run 2 preparation for ATLAS $H \rightarrow ZZ^* \rightarrow 4\ell$ analysis. It specifically focuses on the design and development of a central analysis framework to accommodate the new ATLAS event data model and the needs of the analysis. Moreover, studies performed using the new analysis framework for validating the new ATLAS software are presented.

2 Standard Model

According to the Standard Model (SM), there are two types of fundamental particles: fermions and bosons. All the matter particles have half integer spin and are subsequently classified as fermions. These obey Fermi-Dirac statistics and, due to the Pauli exclusion principle, cannot occupy the same quantum state. On the other hand, force meditators, which are bosons, have integer spins and are governed by Bose-Einstein statistics. Unlike fermions, multiple bosons can occupy the same state.

Fermions are further sub-classified into leptons and quarks and they exist in three generations or flavours, with each being a heavier copy of the previous one. In each generation, the lepton family consists of a charged lepton and an uncharged neutrino. Table 1 lists the known leptons according to their generation. Due to the presence of electric charge and weak hypercharge, leptons interacts both electromagnetically and weakly, while neutrinos only couple to the weak force.

Table 1: List of leptons and their properties [1]. Note that weak hypercharge depends on the handedness of the particle. For charged leptons, both right-handed (RH) and left-handed (LH) states carry hypercharge, while for neutrinos only LH particles have a hypercharge.

Generation	Lepton	Mass [MeV]	Electric Charge	Weak hypercharge
First	Electron (e)	0.511	-1	-1 (LH) & -2 (RH)
	Electron-neutrino (ν_e)	$< 2 \times 10^{-3}$	0	-1 (LH)
Second	Muon (μ)	105.7	-1	-1 (LH) & -2 (RH)
	Muon-neutrino (ν_μ)	< 0.19	0	-1 (LH)
Third	Tau (τ)	1776.8	-1	-1 (LH) & -2 (RH)
	Tau-neutrino (ν_τ)	< 18.2	0	-1 (LH)

Quarks are distinguishable from leptons as they have fractional electric charge and carry a colour charge[1]. Table 2 lists the experimentally observed quarks and their properties. Due to the presence of colour charge and confinement, quarks interact via the strong force and cannot exist as free particles. Rather they are required to form colour singlets, such as mesons and baryons, to create observable states. Lastly, the weak force mixes the quark mass eigenstates, and thereby allows quarks to decay to a different generation.

Table 2: List of quarks and their properties [1]. Note that weak hypercharge depends on the handedness of the particle.

Generation	Quarks	Mass [MeV]	Electric Charge	Weak Hypercharge
First	Up (u)	2.3	$^2/_3$	$^1/_3$ (LH) & $^4/_3$ (RH)
	Down (d)	4.8	$-^1/_3$	$^1/_3$ (LH) & $-^2/_3$ (RH)
Second	Charm (c)	1.3×10^3	$^2/_3$	$^1/_3$ (LH) & $^4/_3$ (RH)
	Strange (s)	95	$-^1/_3$	$^1/_3$ (LH) & $-^2/_3$ (RH)
Third	Top (t)	173.2×10^3	$^2/_3$	$^1/_3$ (LH) & $^4/_3$ (RH)
	Bottom (b)	4.18×10^3	$-^1/_3$	$^1/_3$ (LH) & $-^2/_3$ (RH)

[1]The three colour charge states are referred to as ‘Red’, ‘Green’ and ‘Blue’.

In the SM, each force is mediated by a gauge boson. Table 3 lists the forces with their corresponding bosons. Despite being a negligible force at the fundamental particle length scale and there being no consistent quantum model, it is hypothesized that the exchange of a Graviton (G) mediates the gravitational interaction between two particles. The massless photon (γ) is a quanta of the electromagnetic (EM) field and facilitates EM interactions between all charged particles. For the weak force, the charged-current interactions are mediated by massive $W^{\pm}$ bosons and the neutral-current interactions by the Z boson. The force carriers for the strong interaction are gluons (g). Unlike other bosons, gluons carry colour charge and occur in 8 colour states. Lastly, the Higgs Boson (H) is an excitation of the Higgs field, which couples to other particles and allows them to gain mass while preserving gauge symmetry.

Table 3: List of gauge bosons and the associated force [1].

Force	Boson	Mass [GeV]	Spin
Strong	Gluon (g)	0	1
Electromagnetic	Photon (γ)	0	1
Weak	$W^{\pm}$	80.4	1
	Z	91.2	1
Gravity	Graviton (G) ?	0	2
Higgs Mechanism	Higgs (H)	≈ 125	0

The SM is the theoretical framework that describes these particles and their interactions. It is developed within the context of Quantum Field Theory (QFT), a many-body Lorentz invariant quantum theory. Section 2.1 describes the theoretical basics of the SM which are relevant to this thesis.

2.1 Theory Elements

2.1.1 Quantum Field Theory

In quantum mechanics, a particle is described by a wavefunction, ψ, and its dynamics are governed by the Schrödinger equation. However in QFT, to combine special relativity and quantum mechanics, the particles are described as an excitation of the field $\phi(x^{\mu})$, where x^{μ} are the space-time coordinates [2, 3, 4]. The dynamics of this are calculated by using a Lagrangian density, $\mathcal{L}$, which only depends on ϕ and its space-time derivatives, $\partial_{\mu}\phi$. The action, S, is defined as:

$$S = \int \mathrm{d}^4\mathrm{x}\ \mathcal{L}(\phi, \partial_{\mu}\phi). \tag{1}$$

By requiring the variation in action, δS, to be zero, the Euler-Lagrange equation describing the dynamics is derived to be [3]:

$$\partial_{\mu}\left(\frac{\partial \mathcal{L}}{\partial(\partial_{\mu}\phi)}\right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \tag{2}$$

In QFT, symmetries play an important role. They are defined as any variation to $\mathcal{L}$ that leaves the final equation of motion invariant. For a specific subset, Noether's theorem states that for any continuous symmetry of the Lagrangian, there exits a conserved current and, hence, a conserved charge [2]. Section 2.1.2 outlines a specific example where symmetries lead to observable conserved charges and interactions.

2.1.2 Quantum Electrodynamics

The Dirac Lagrangian [2, 4], which represents a free massive spin $\frac{1}{2}$ particle, is given by:

$$\mathcal{L}_D = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi \tag{3}$$

where ψ is the spinor field, m is the mass of the particle and γ^{μ} are the 4 gamma matrices, which obey the anticommutation relationship, $\{\gamma^{\mu}, \gamma^{\nu}\} = 2\mathrm{g}^{\mu\nu}$. $\mathcal{L}_D$ is invariant under a global U(1) phase transformation:

$$\psi \rightarrow \psi' = e^{iq\theta}\psi. \tag{4}$$

This, according to Noether's theorem, creates a conserved current:

$$j^{\mu} = \bar{\psi}\gamma^{\mu}\psi. \tag{5}$$

Equation 5, when interpreted under the context of classical or QFT framework, leads to the well-known EM continuity equation and conservation of electric charge.

A more interesting phenomenon emerges when local U(1) transformations are considered. This transformation is given by:

$$\psi \rightarrow \psi' = e^{iq\theta(x)}\psi. \tag{6}$$

Due to the presence of a derivative, $\mathcal{L}_D$ is not invariant under this but transforms according to:

$$\begin{aligned}\mathcal{L}_D \rightarrow \mathcal{L}'_D &= ie^{-iq\theta}\bar{\psi}\gamma^{\mu}[e^{iq\theta}\partial_{\mu}\psi + iq(\partial_{\mu}\theta)e^{iq\theta}\psi] - me^{-iq\theta}\bar{\psi}e^{iq\theta}\psi \\ &= \mathcal{L}_D - q\bar{\psi}\gamma^{\mu}(\partial_{\mu}\theta)\psi.\end{aligned} \tag{7}$$

However if this symmetry is explicitly required from the Lagrangian, it forces a substitution of ∂_{μ} with its covariant derivative D_{μ}:

$$D_{\mu} = \partial_{\mu} + iqA_{\mu} \tag{8}$$

where A_{μ} is a vector field with its U(1) gauge transformation being defined by:

$$A_{\mu} \rightarrow A'_{\mu} = A_{\mu} - \partial_{\mu}\theta. \tag{9}$$

Since a new field was included, its kinematic term must also be added to $\mathcal{L}_D$. This leads to the final form of the Lagrangian, which describes Quantum Electrodynamics (QED):

$$\mathcal{L}_{\text{QED}} = \bar{\psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + e\bar{\psi}\gamma^{\mu}\psi A_{\mu} \tag{10}$$

where $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ and q has been replaced with the electron charge. The last term in $\mathcal{L}_{\text{QED}}$ is the potential term which represents the coupling of the electron and photon field with a strength of e.

It should be noted that the above procedure provides a general method to describe interactions between particles in the SM. For QED, by starting with a Lagrangian for a free electron and imposing local U(1) gauge symmetry, it forces the introduction of a photon field and electron-photon coupling, which ultimately leads to EM interactions. Similarly for Electroweak and Quantum Chromodynamics (QCD), by imposing local $\mathrm{SU(2)}_L \otimes \mathrm{U(1)}_Y$ and SU(3) gauge symmetry respectively on a free particle Lagrangian, the weak and strong forces between particles naturally arise.

2.1.3 Higgs Mechanism

As shown in the Section 2.1.1, local gauge symmetry provides an elegant and consistent method to describe all known fundamental forces which have a measurable effect on fundamental particles. However, a shortcoming of this process is that the introduction of a mass term for the vector field breaks the imposed symmetry. For example, if the photon had a mass of m_γ, $\mathcal{L}_{\text{QED}}$ would be modified to:

$$\mathcal{L}'_{\text{QED}} = \mathcal{L}_{\text{QED}} + \frac{1}{2}m_\gamma^2 A^\mu A_\mu. \tag{11}$$

Clearly, the last term in Equation 11 is not invariant under local U(1), as its transformation is given by:

$$\begin{aligned}\frac{1}{2}m_\gamma^2 A^\mu A_\mu \to \frac{1}{2}m_\gamma^2 A'^\mu A'_\mu &= \frac{1}{2}m_\gamma^2(A^\mu - \partial^\mu\theta)(A_\mu - \partial_\mu\theta)\\ &\neq \frac{1}{2}m_\gamma^2 A^\mu A_\mu.\end{aligned} \tag{12}$$

This issue is particularly problematic for the $W^\pm$ and Z bosons as they are experimentally found to be massive (see Table 3).

To mitigate this problem, the concept of spontaneous breaking was borrowed from BCS theory and applied to the SM [5]. This mechanism was independently proposed in 1964 [5] by Englert and Brout [6], Higgs [7], and subsequently by Guralnik, Hagen and Kibble [8]. It consists of expanding the ground state of a scalar field around a specific non-zero point, which allows the vector field to gain mass while still preserving local symmetry.

Mathematically, if a complex scalar field, ϕ, is considered with a potential:

$$V(\phi) = \mu^2\phi^2 + \lambda\phi^4 \tag{13}$$

it leads to the following Lagrangian:

$$\mathcal{L} = (\partial_\mu\phi)^*(\partial^\mu\phi) - V(\phi). \tag{14}$$

Similar to the method described in Section 2.1.1, imposing local U(1) symmetry gives a modified $\mathcal{L}$:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^*(D^\mu\phi) - V(\phi) \tag{15}$$

where:

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{16}$$

and:

$$(D_\mu\phi)^*(D^\mu\phi) = (\partial_\mu - igB_\mu)\phi^*(\partial^\mu + igB^\mu)\phi. \tag{17}$$

It should be be noted that, as before, U(1) symmetry prevents the introduced vector field, B_μ, to become massive [4, 9, 10].

If, for $V(\phi)$, it is imposed that $\mu^2 < 0$, it leads to the potential shown in Figure 1 [11]. The minimum for this function exists on a curve defined by $\phi^2 = \frac{-\mu^2}{\lambda} = v^2$. Hence, the physical vacuum or ground state will correspond to a single point along this curve, breaking the global U(1) symmetry. Moreover,

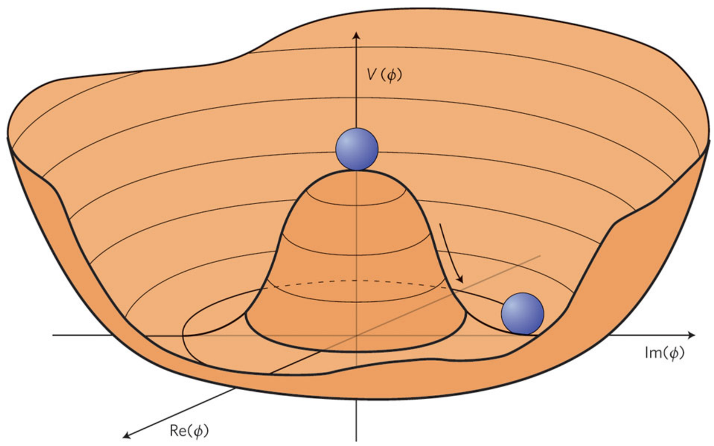


Figure 1: The Higgs potential for $\mu^2 < 0$.

the excitation of this field can be investigated by expanding ϕ around this point. In the Unitary gauge, variations around the vacuum state are defined by,

$$\phi = \frac{1}{\sqrt{2}}(v + h(x)) \tag{18}$$

where is h is the physical Higgs field. This leads Equation 15 to be,

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}(\partial_\mu h)^*(\partial^\mu h) - \lambda v^2 h - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2 v^2 B_\mu B^\mu \\ &+ g^2 v B_\mu B^\mu h + \frac{1}{2}g^2 B_\mu B^\mu h^2 - \lambda v h^3 - \frac{1}{4}\lambda^4 h^4.\end{aligned} \tag{19}$$

The first two terms in Equation 19 describe a Higgs boson with a mass of $m_H = \sqrt{2}gv$. However, the more interesting parts are the next two terms which describe a massive gauge boson associated with the B vector field. It should be noted that while this mass term breaks local U(1) symmetry, the complete Lagrangian is still invariant. The last line describe the interactions between the Higgs and B boson and the Higgs self interaction terms [3, 4, 9, 10].

Overall, the Higgs mechanism provides a novel way for massive gauge bosons to exist in the theory. In the SM, this mechanism is used in the Electroweak theory to allow $W^\pm$ and Z bosons to be massive, while still preserving $SU(2)_L \otimes U(1)_Y$ symmetry. The existence of the massive scalar Higgs boson forms the main prediction of this mechanism, with its mass being the only free parameter [3, 4, 9, 10].

2.2 Searches for the Higgs Boson

2.2.1 Theoretical Constraints

Since the Higgs field couples to the weak bosons, the mass of the Higgs boson (m_H) affects higher order corrections to many Electroweak predictions and m_H can be constrained by fitting these parameters. The

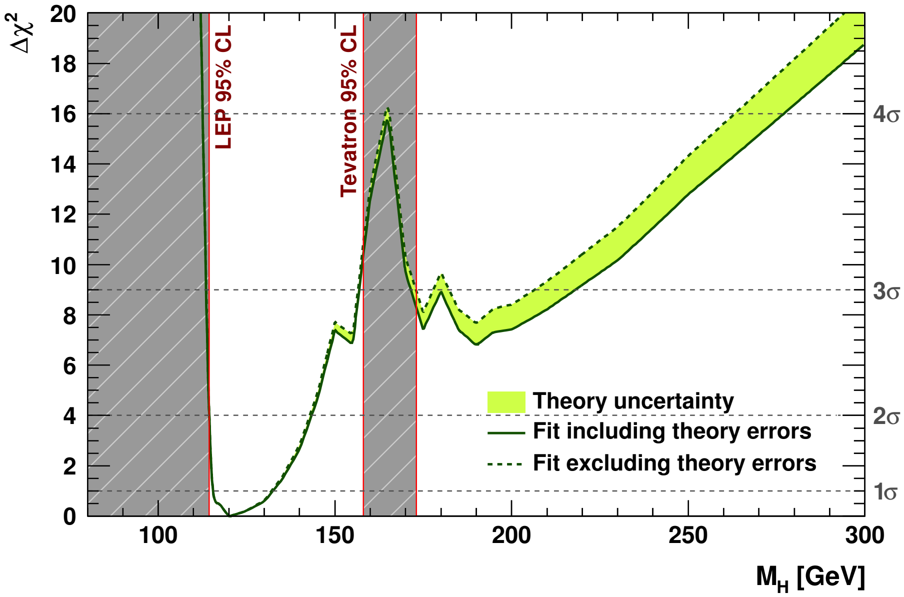


Figure 2: $\Delta\chi^2$ curve for Higgs mass fit based on Electroweak observables and Higgs searches at LEP and Tevatron [12]. The shaded regions are excluded by direct searches.

GFitter program [12] provides such a method. It uses a χ^2 minimization to fit Electroweak observables, which have been measured precisely at previous accelerator experiments. In this 'standard fit', the best fit is found to be [12],

$$m_H = 91^{+30}_{-23} \text{ GeV} \tag{20}$$

Moreover, by including the results from direct Higgs searches at the Large Electron Positron (LEP) and Tevatron colliders, this estimate is improved and yields the best fit for Higgs mass to be [12],

$$m_H = 120^{+12}_{-5} \text{ GeV} \tag{21}$$

Figure 2 shows the $\Delta\chi^2$ curve for this 'complete' fit [12].

2.2.2 Direct Searches

In the pre-LHC era, the most significant results for direct Higgs searches came from the experiments at the LEP and Tevatron colliders. The LEP collider at CERN excluded the existence of the Higgs boson with $m_H < 114.4$ GeV at 95% confidence level (CL) [13]. These limits were mainly driven by results probing $e^+e^- \to Z \to HZ$. On the other hand, Tevatron searches in the $H \to b\bar{b}$ and $H \to W^+W^-$ channels excluded the possibility of a Higgs boson at 95% CL in the mass range of $100 < m_H < 103$ GeV and $147 < m_H < 180$ GeV and saw an $\approx 3\sigma$ excess over the predicted background in the range $115 < m_H < 140$ GeV [14].

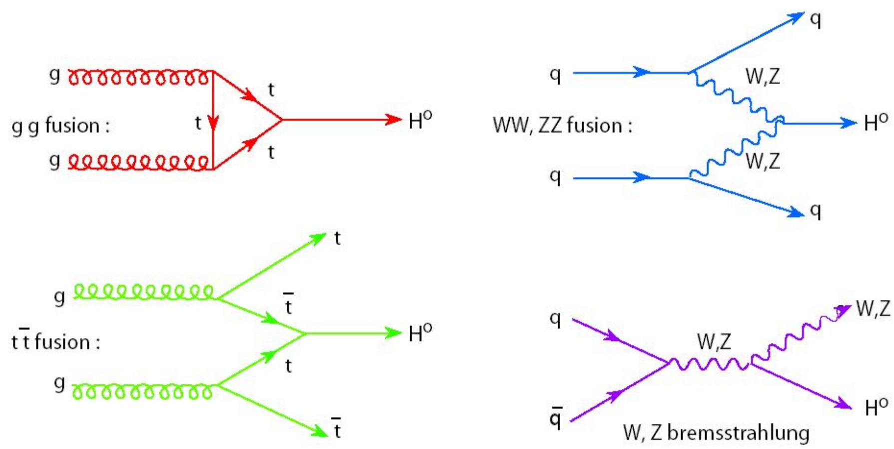


Figure 3: Leading order Feynman diagram for Higgs production modes available at LHC. ggF production dominates, followed by VBF, VH and ttH respectively.

At the LHC, there are four main modes through which a Higgs boson can be produced: gluon-gluon fusion (ggF), vector boson fusion (VBF), vector boson associated production (VH) and top quark associated production (ttH). Figure 3 shows the leading order Feynman diagrams for these modes. At m_H = 125 GeV and $\sqrt{s}$ = 8 TeV, the ggF contribution at 19.27 pb is approximately 10 times greater than VBF at 1.578 pb and 100 times greater than the VH (0.704 pb) and ttH (0.129 pb) production modes [15].

Once a Higgs boson is produced, it can decay into a variety of final states as shown in Figure 4 [15]. In the low mass regime, $H \to b\bar{b}$ and $H \to \tau\tau$ have a significant branching ratio; however, large QCD backgrounds make the analysis challenging. $H \to W^+W^-$ is available across a large range of m_H but the subsequent hadronic decays of $W^\pm$ suffer from large backgrounds and the leptonic decay channels are not fully reconstructable due to the presence of a neutrino.

In Run 1 of the LHC, the Higgs boson searches were driven by the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ decay channels. This was due to the clean photon or charged lepton final state, large signal to background ratio and fully reconstructable final state. In 2011 and 2012, ATLAS and CMS collected approximately 25 fb^{-1}of data and observed a Higgs boson [16, 17]. The final ATLAS mass measurement result for the combination of the two channels was [16]:

$$m_H = 125.36 \pm 0.37(\text{stat}) \pm 0.18(\text{syst})\ \text{GeV}. \tag{22}$$

Other properties of the particle were also tested by both collaborations and found to be consistent with a SM Higgs boson. Specifically for the $H \to ZZ^* \to 4\ell$ channel, the overall signal strength[2] was measured to be $1.44^{+0.40}_{-0.33}$ [18], the spin 0 hypothesis was preferred over others [19] and the measured couplings to different production modes were compatible with the SM predictions [18]. Figure 5 shows the final distribution of the selected candidates in this channel [16].

[2]Signal strength is defined as the ratio between the measured and predicted cross-section

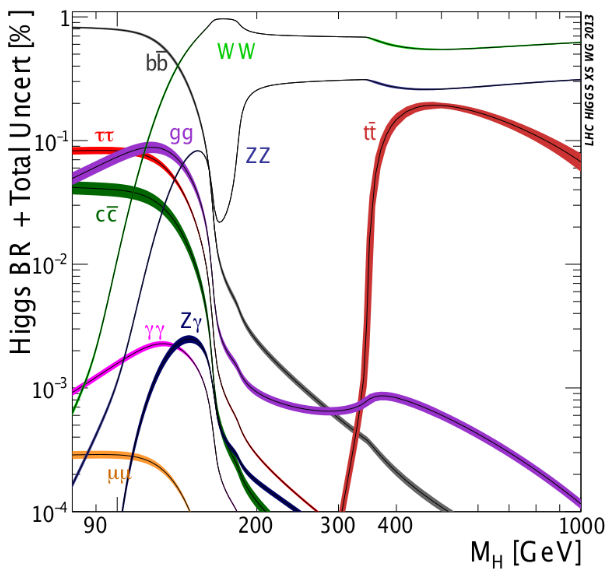
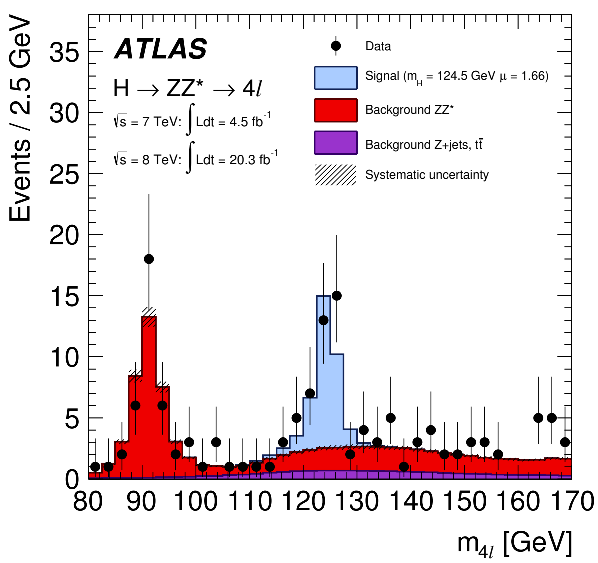
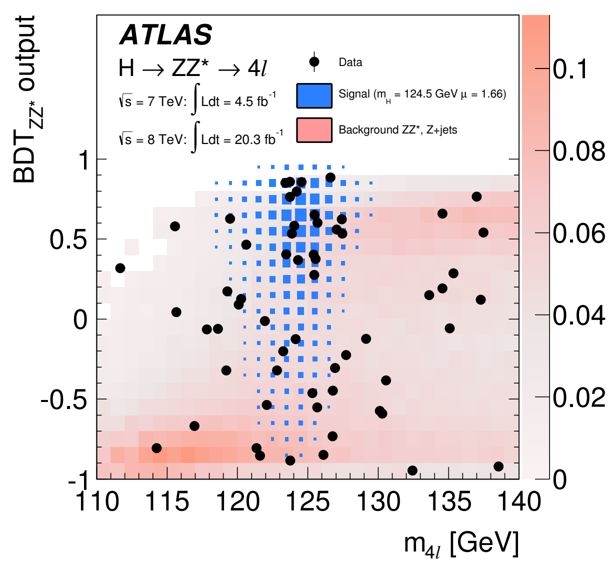


Figure 4: Predicted branching ratio for the Higgs boson as a function of Higgs mass.



(a) The four lepton invariant mass distribution [16]. ('Plot made by S.H. Abidi.')



(b) The four lepton invariant mass and BDT output distribution [16]. BDT was designed to separate the signal and background. ('Plot made by S.H. Abidi.')

Figure 5: Data and Monte Carlo distributions for selected candidates in the $H \to ZZ^* \to 4\ell$ channel.

3 The Large Hadron Collider

The LHC is a superconducting accelerator and collider installed in the 27 km underground tunnel at CERN. It is designed to accelerate bunches of protons and heavy-ions to a centre of mass energy of 14 TeV and 5.52 TeV per nucleon respectively at a peak instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ [20]. At four points along the ring, the LHC collides two oppositely rotating particle beams and each interaction point accommodates a main experiment: ATLAS, ALICE, CMS or LHCb. ATLAS and CMS are two general purpose detectors, designed to investigate a wide variety of physical phenomenon. On the other hand, ALICE focuses on investigating QCD and LHCb is optimized for physics measurements with hadrons containing a b-quark.

LHC is the final step of the accelerator chain at CERN, shown in Figure 6, that increases the energy of particles to their final values [20]. Protons start in LINAC 2 where their energy is increased to 50 MeV. They are then transported to the Proton Synchrotron (PS) Booster and then to PS which results in an energy of 25 GeV. These protons are then injected into the Super Proton Synchrotron (SPS) that accelerates them to 450 GeV and creates bunches appropriate for LHC injection.

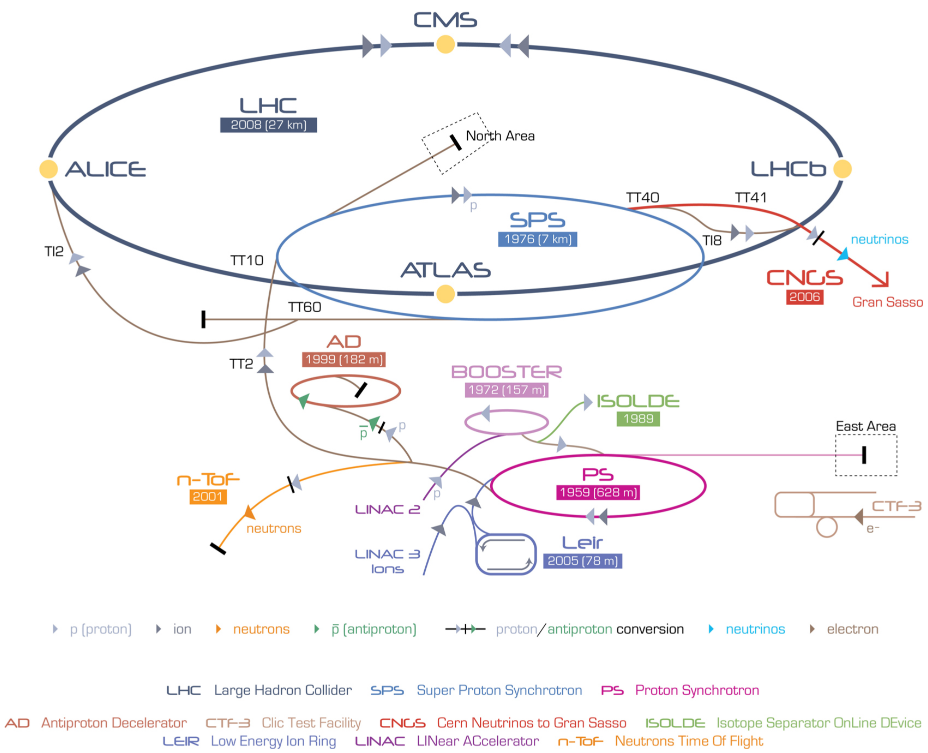


Figure 6: The accelerator complex at CERN.

Once the particle bunches are transported to the LHC, a radio-frequency (RF) system is used to increase the beam energy. This RF system consists of 16 400MHz single-cell Niobium sputtered copper cavities [20], which are designed to create a superconducting inner layer when operated at 1.9K. Four of these cavities are housed inside a cryomodule and two are used per beam. Figure 7 shows the schematic of

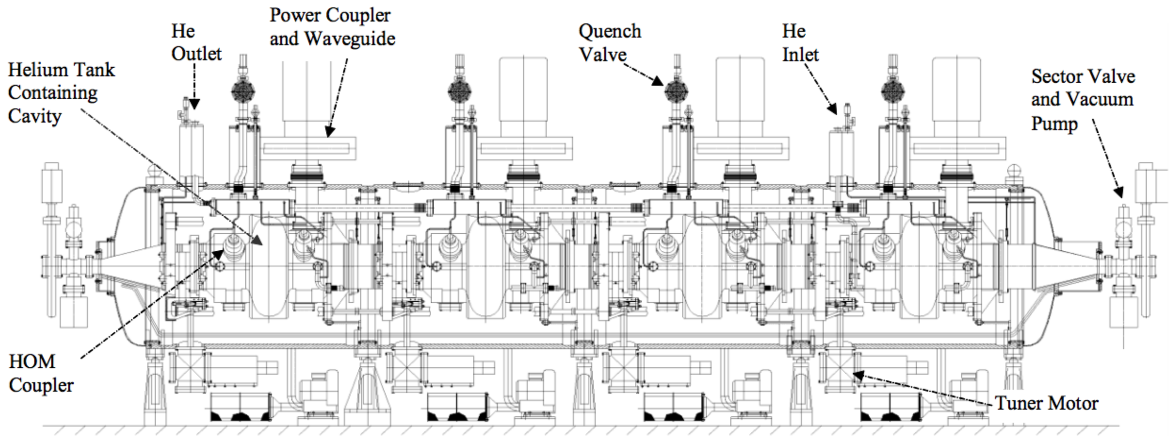


Figure 7: Schematic of the LHC accelerator cryomodule.

the LHC cryomodules. Cavities, operating at their nominal gradient of $5.5^{\mathrm{MV}}/_{\mathrm{m}}$, provide an energy increase of 0.5 MeV per beam turn [20]. The cryomodule also contains other auxiliary devices to ensure proper operation of the cavities. These include a 'tuner' to correct changes in cavity resonance frequency caused by external factors such as variations in helium pressure and a 'Higher Order Modes (HOM) coupler' to dampen the higher order resonances excited by beam wakefields [20].

4 The ATLAS Detector

ATLAS (A Toroidal LHC ApparatuS) [21] is a general purpose detector built around one of the LHC's interaction points. It is 44 m long, 25 m high, weighs over 7000 tons and covers almost the entire 4π solid angle. Figure 8 shows a picture of ATLAS while Figure 9 shows a cross-sectional view with simulated particles and their interactions within the detector. The various sub-detectors are layered cylindrically to provide uniform coverage. From the centre to the outside, these are:

- The *Inner Detector* measures tracks of charged particles and provides momentum and vertex measurements. It also aids in electron identification.
- The *Electromagnetic and Hadronic Calorimeters* absorb electrons, photons and hadrons and measure their energy.
- As muons escape the calorimeters, the *Muon Spectrometer* provides additional measurements that can be used to independently reconstruct muons.

A system of toroidal and solenoid magnets provides magnetic fields of 4 T and 2 T respectively. These bend the trajectory of charged particles and allow their momentum to be measured. Moreover, since the interaction rate at LHC design luminosity is approximately 1 GHz and data recording is limited to ~ 100 Hz, a trigger system is used to reject events by a factor of $\sim 10^7$. To maximize the physics reach of ATLAS, this system is designed to accept events with decay signatures from rare and exotic particles.

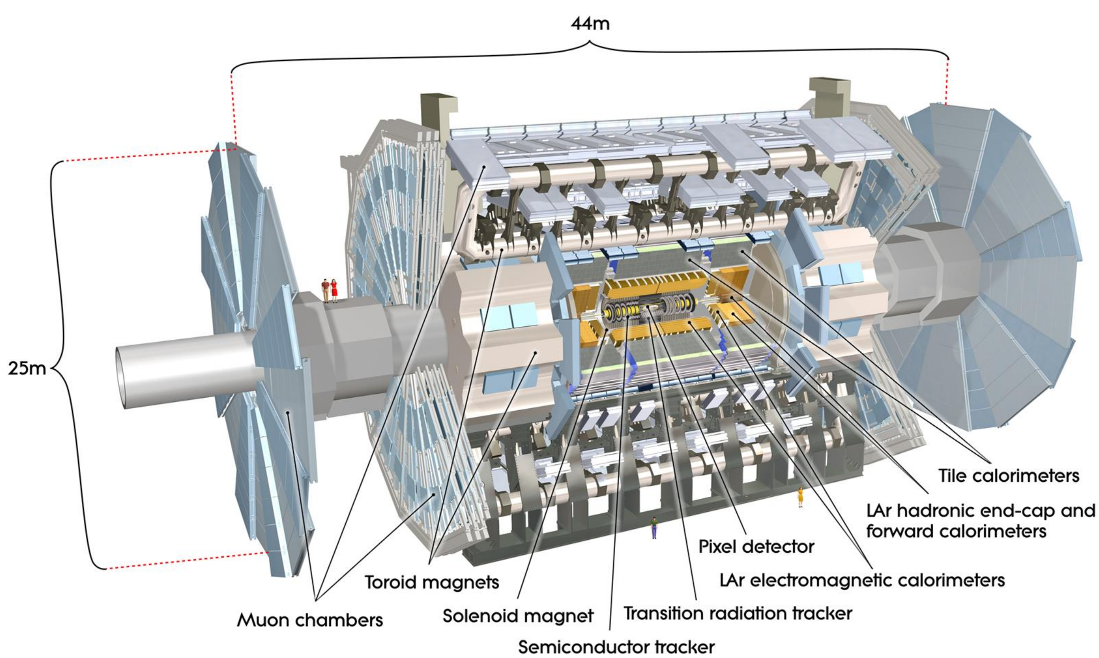


Figure 8: Cut away view of the ATLAS detector with its various subsystems labelled.

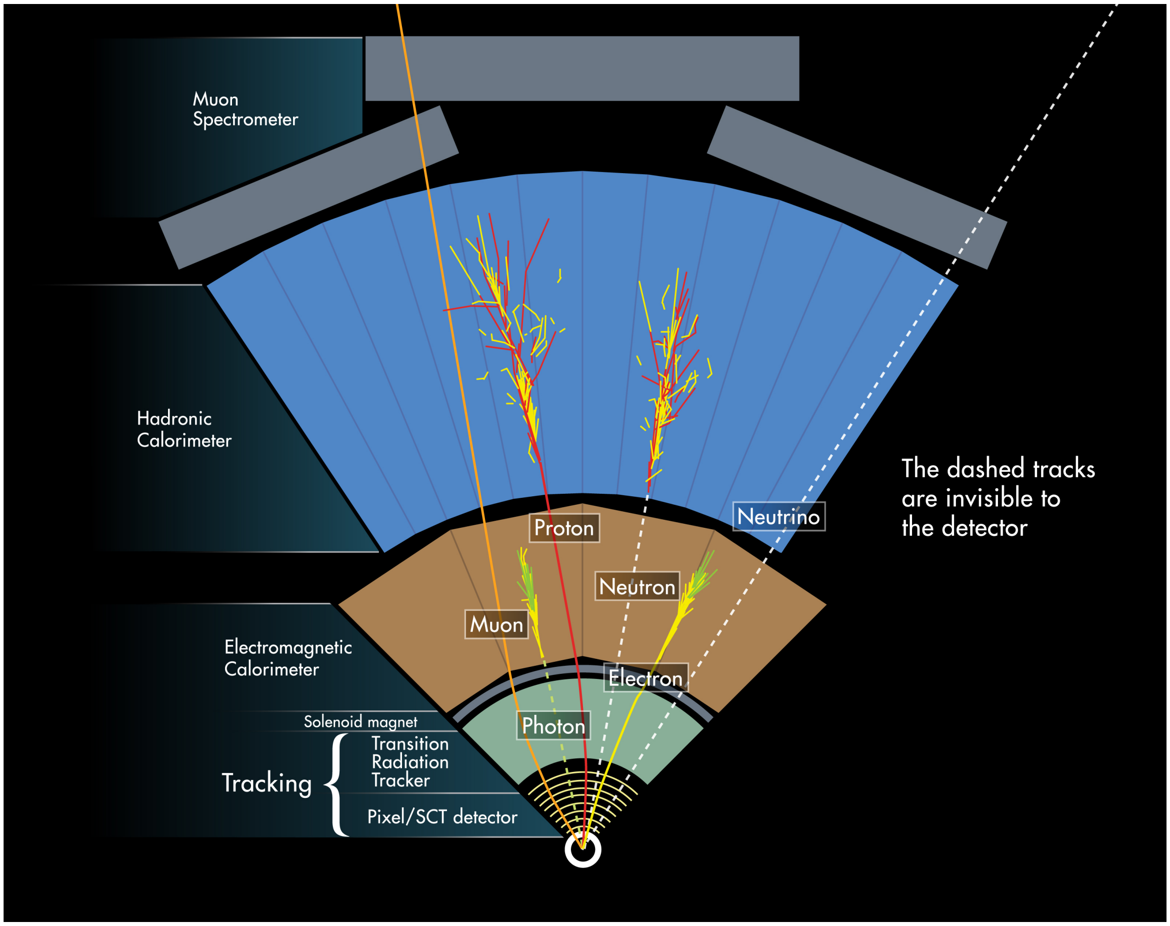


Figure 9: Cross-sectional view of ATLAS. Simulated particles and their interactions within the detector are overlaid.

ATLAS uses a right-handed coordinate system. The origin is placed at the nominal beam interaction point, which also corresponds to the centre of the detector. The z-axis is oriented along the beam pipe and the x-y plane is orthogonal to it. The x-axis points towards the centre of the LHC ring and the y-axis points upward. In polar coordinates, the azimuthal angle ϕ is defined around the beam pipe. The pseudo-rapidity η is defined as $\eta = -\ln[\tan(\theta/2)]$ where θ is the polar angle.

4.1 Inner Detector

Due to the large number of particles created in a single event and the required momentum resolution, vertexing and identification requirements, the Inner Detector (ID) employs three different technologies to measure the path of a charged particle in the $|\eta| < 2.5$ region [21, 22]. The pixel detector and the semiconductor tracker use silicon pixels and strips, while the transition radiation tracker uses straw tubes to track particles and provides identification using transition radiation. Figure 10 shows a view of the ID.

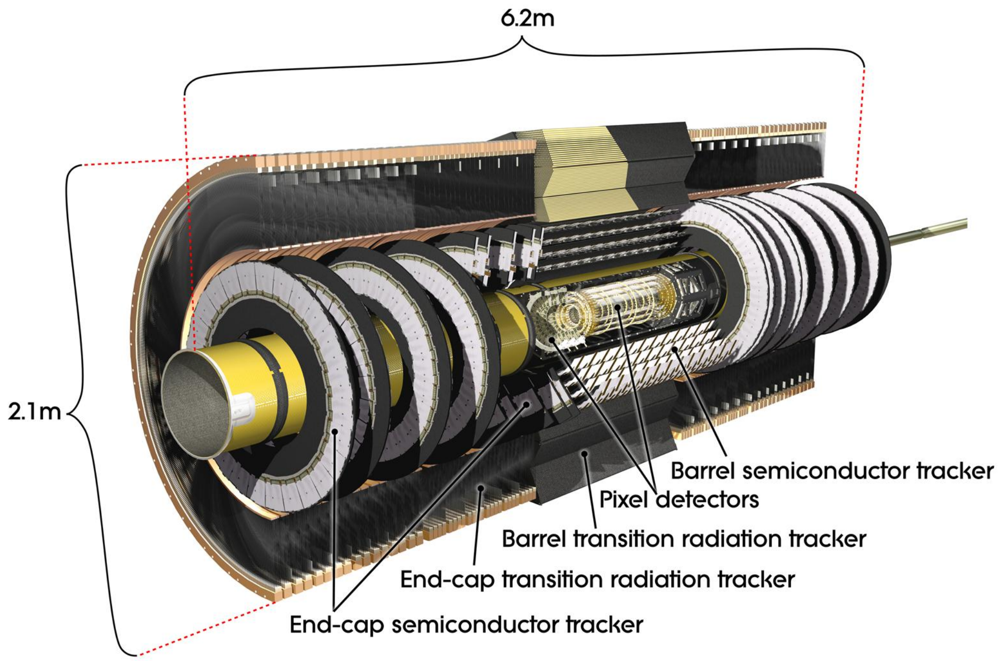


Figure 10: Perspective view of the ATLAS Inner Detector.

The pixel detector is closest to the interaction point and has the highest granularity. It uses reverse-biased doped silicon pixels to measure the passage of a charged particle: when they interact with the semiconductor, electron-hole pairs are created and a subsequent collection of these pairs creates the required electrical signal. ATLAS uses 80.4 million pixel sensors, each with a size of 50×400 μm^2 in $R - \phi$. These are arranged in 3 concentric cylinders in the central barrel region and in 3 rings for each end-cap region. The barrel (ring) region has an intrinsic accuracy of 10 μm in $R - \phi$ and 115 μm in z (R). The pixel detector provides tracking up to $|\eta| < 2.5$ and a charged particle, on average, leaves 3 hits.

The semiconductor tracker (SCT) uses small angle stereo silicon strips [23] to provide a practical method to accurately track particles. Two strip detectors are stacked together with a 40 mrad offset and the long sensor axis is placed along the beam direction. This stereo configuration provides longitudinal (radial) constraint in the barrel (end-cap) regions and lead to an intrinsic accuracy of 17×580 μm in $R - \phi \times z$ in the barrel (R in the end-cap). The 6.3 million SCT sensors are arranged in 4 layers in the barrel region and in 9 rings in each end-cap region. A particle typically leaves 8 hits in the SCT region.

The transition radiation tracker (TRT) provides approximately 36 hits per track in the $|\eta| < 2.0$ region. It uses 2 mm straw with a 32 μm tungsten wire strung at the centre. The straw is filled with a xenon gas mixture and the wire is held at -1500 V with respect to the straw. When a charged particle traverses the straw, it ionizes the gas. The ejected electrons cause a further avalanche of ionization and drift toward the wire where they are collected. Due to the long lever-arm, TRT measurements contribute significantly to the momentum measurement.

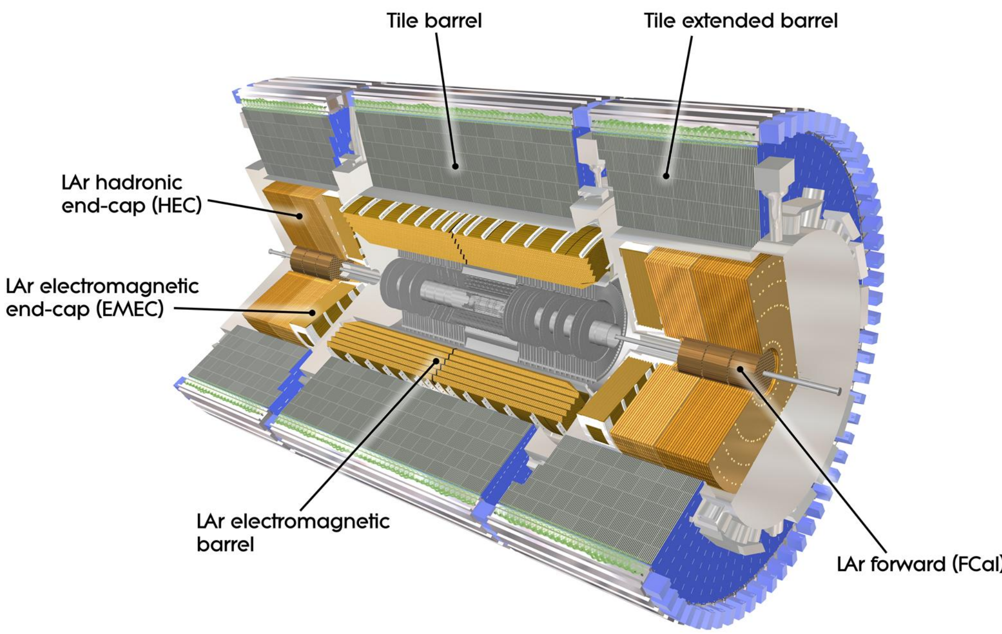


Figure 11: Cut away view of the ATLAS calorimeter system.

4.2 Calorimeter

The ATLAS calorimetry system consists of 2 major components, covering the $|\eta| < 4.9$ region: the electromagnetic calorimeter (ECal) and the hadronic calorimeter (HCal). The ECal is designed to measure the energy of particles that predominately interact electromagnetically, such as electrons and photons. While jets leave a significant energy deposit in the ECal, they escape it but are contained and measured within the HCal. Figure 11 shows a cut away view of the calorimeters.

The ECal is a liquid argon-lead sampling calorimeter. When an electron or photon passes through the lead absorber material, it creates a shower of particles through $\gamma \to e^+e^-$ conversions and the $e^\pm \to \gamma e^\pm$ bremsstrahlung process. The active liquid argon layer 'samples' these showers through ionization and provides a measurement for the energy of the incoming particle. The ECal is divided into a barrel region ($|\eta| < 1.475$) and two end-cap components ($1.375 < |\eta| < 3.2$). These are approximately 22 radiation lengths (X_o) in depth to contain most of the EM shower. X_o is defined as the distance needed for an electron to lose $^1/_e$ of its energy to bremsstrahlung or 54% of photons to undergo conversion.

An important aspect of the ECAL is its accordion shaped geometry, as shown in Figure 12. This provides a constant coverage in azimuthal angle without any cracks. The first layer of this geometry, with the finest granularity, aids in photon identification. The second layer, with $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ cells, are designed to contain most of the deposited energy. The third layer provides an extra measurement before the particles enter the HCal.

The HCal uses multiple different types of calorimeters, depending on the expected radiation levels. In the barrel region ($|\eta| < 1$) and the extended barrel region ($0.8 < |\eta| < 1.7$), a tile calorimeter is

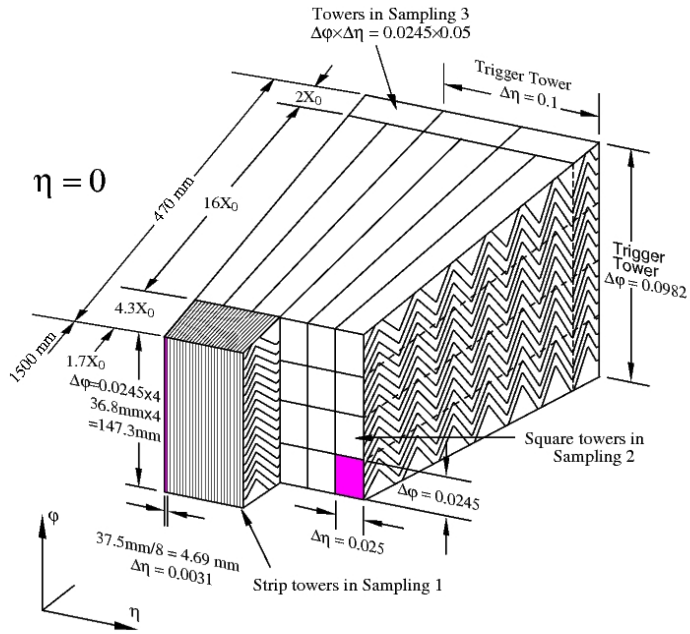


Figure 12: Accordion shaped geometry of the ECal.

used. Using a similar principle as the ECal, it employs steel absorbers and scintillating tiles as the active material. It is approximately 11 interaction lengths thick, which reduces the hadronic 'punch-through' to the muon spectrometer. In the end-cap region ($1.7 < |\eta| < 3.2$), a liquid argon with copper absorber calorimeter is used. Finally, in the forward region ($3.2 < |\eta| < 4.9$), another liquid argon calorimeter is used. It uses three layers of absorbers: 1 layer of copper absorber and 2 layers of tungsten absorber. This is due to the high levels of radiation in the forward region and the fact that tungsten is a more radiation hard material.

4.3 Muon Spectrometer

The muon spectrometer (MS) is designed to provide an independent measurement for muons, which are minimum ionizing and escape the ID and the calorimeters. It can measure muon momenta accurately down to 3 GeV. For high p_{T} ($\approx$ 3 TeV) muons, the MS has an adequate resolution of $\approx 10\%$ and provides excellent charge identification. Moreover, the MS also contains a series of dedicated chambers for fast triggering.

The monitored drift tubes (MDT) and cathode strip chambers (CSC) are two technologies used in the MS for precision tracking. Similar to TRT, the MDTs use aluminium drift tubes with a tungsten wire at the centre and are filled with an argon-carbon dioxide gas mixture. Three layers of MDTs are used to provide tracking in the $|\eta| < 2$ region. The MDTs are consistently monitored through a laser system to ensure a combined resolution of 50 μm for the tracks. The CSCs are multi-wire proportional chambers designed to cope with the high interaction rates in the forward region ($2 < |\eta| < 2.7$). The ionized gas

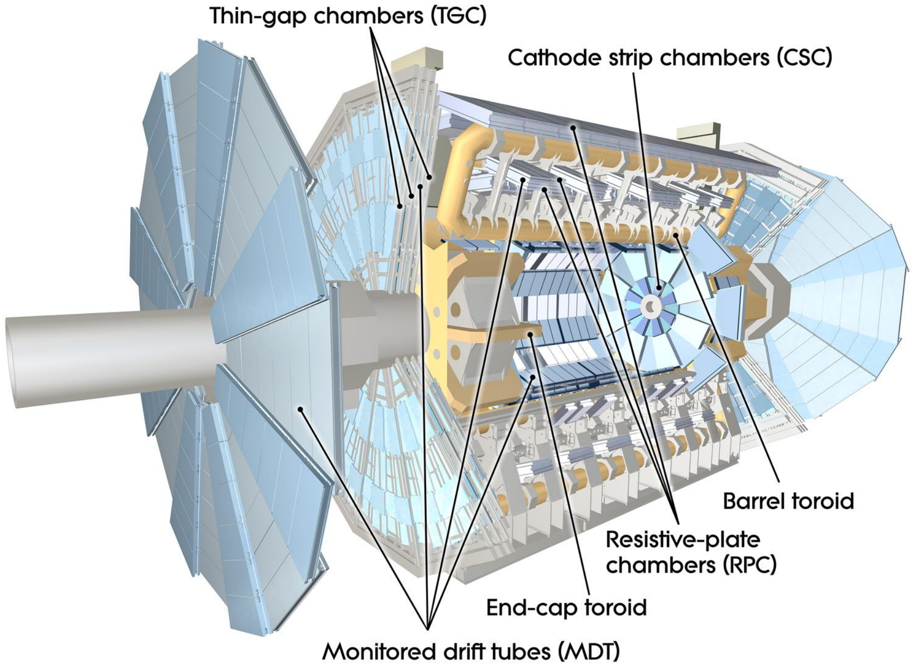


Figure 13: The ATLAS muons spectrometer.

created by the passage of a muon is collected using the anode wire and also the particle induces induces a charge in the cathode strips. This combined measurement in orthogonal directions allow for a 60 μm resolution in the track position.

To allow for triggering in the event crossing time space, the muon triggering chambers are designed for fast response. Resistive Plate Chambers (RPCs) are used in the barrel region and are located on both sides of a MDT station. Each RPC consists of two bakelite sheets separated by a narrow gas region filled with tetrafluorethane gas. An electric field is applied for collection of ionized gas. On the other hand, Thin Gap Chambers (TGCs) are used in the end-cap regions and are very similar to CSCs in design. These triggering chambers also provide a measurement of the muon track coordinate in the direction orthogonal to the one determined by the precision chambers.

5 Run 2 Upgrades

Extending from 2010 to 2012, in Run 1 of data taking for the LHC and the associated experiments, the accelerator ran at a lower centre of mass energy and instantaneous luminosity than its design capabilities. This was mainly due to the catastrophic quench in the superconducting wire at LHC in 2008 [24]. Therefore, in the Long Shutdown 1 (LS1), LHC has upgraded its infrastructure for safe and reliable running at its design capabilities. Moreover, the experiments, especially ATLAS, are taking the opportunity to improve the detector and the analysis procedures for Run 2.

5.1 LHC Upgrades

One of the major LHC upgrades is the addition of the bus-bar splices around the superconducting wire interconnects [25]. This is to prevent a repeat of the 2008 incident. Moreover, 19 cryo-magnets will be replaced and strategic radiation shielding around important electronics will be placed [25, 26]. This is to ensure minimal downtime during the data taking period. The associated accelerator complex, such as SPS and PS, are also undergoing upgrades for operations at increased instantaneous luminosity levels in Run 2 [26].

5.2 ATLAS Upgrades

ATLAS had a wide program for upgrades in LS1, which can be broadly classified into hardware and software level upgrades. Section 5.2.1 and Section 5.2.2 list some of the upgrades which affect the $H \to ZZ^* \to 4\ell$ analysis.

5.2.1 Hardware Upgrades

Due to the increased instantaneous luminosity of LHC in Run 2, it was expected that the high radiation dosage to the ID would significantly damage and limit its performance, ultimately affecting the physics capabilities of the entire detector. To minimize the risk, a new ‘insertable B-Layer’ (IBL) is designed to be placed inside the ID as a fourth inner layer [27]. This is accomplished by reducing the beam pipe diameter and developing novel ways to support the module [27]. The IBL consists of silicon pixel sensors, which will provide another point measurement for tracks, improving vertexing and particle identification.

Another major upgrade for ATLAS is to its trigger system. The level 2 (L2) and event filter (EF) farms are consolidated into a high-level trigger (HLT) to allow for similar performance as offline reconstruction. Its output is also increased to ≈ 1 kHz, allowing for more data to be stored [28]. However, even with the increased rate, the Run 1 trigger menu would saturate the output. Therefore, new triggers are designed to cope with this [28]. It is accomplished by raising p_{T} thresholds, tightening isolation requirements and requiring additional complex objects in a given event.

Other ATLAS upgrades include consolidation of different parts of the detector for improved performance, such as replacing power supplies for LAr calorimeters to reduce trips. Also, staged muon chambers are installed, increasing muon efficiency and background rejection. The gas mixture in the TRT is replaced with a xenon-argon mixture to reduce operational cost of leaky chambers but at the expense of reducing electron identification performance.

5.2.2 Software Upgrades

In Run 1, the data followed a very specific route before being analyzed by users. The raw detector level data was stored in a RAW format, which was used to reconstruct the event. The output was saved in a Event Summary Data (ESD) format which contained detailed information about physics objects, tracks, etc [29]. This was then processed into a Analysis Object Data (AOD) format, reducing the information to a level suitable for user analysis. However, the standard practice for physics groups was to process these into a D3PD format [30], before using them to create ntuples (or minitrees) for final analysis and plotting.

The AOD files are written inside ATHENA [31] via a persistency framework based on Storegate [32]. This format was originally envisioned to be used for final analysis and significant work was put into I/O optimization and schema evolution. However, AODs are highly dependant on the ATHENA framework, non-navigable in simple fashion and only support container level reading of data. [32, 33]. Moreover, the perceived difficulty of ATHENA also prevented the general adoption of this format [33].

The D3PDs were developed to overcome some of the limitations of the AOD format. This simple ntuple based format was easily navigable and usable in the widely adopted ROOT framework [34]. Almost all physics groups created custom D3PD versions which contained information relevant to their analysis and to apply Combined Performance (CP) group recommendation. Furthermore, during the D3PD processing, physics objects, such as different jet collections, and auxiliary data were computed to lessen the computational strain on the end user.

Despite the advantages, the D3PD format is not maintainable for Run 2 [32]. A reprocessing of the data, due to updating of a frozen Tier-0 release, triggered D3PD reproduction which significantly increased the time required for changes to propagate to the final analysis stage. Furthermore, due to proliferation of D3PD formats, the needed disk space, computational overhead and maintenance time were not sustainable with the pledged Run 2 computing resources. Lastly, due to lack of standardized validation procedure, many D3PD productions were unusable due to bugs or missing information.

As a solution to these problems, a new event data model (EDM) is being developed and it is labelled as 'xAOD' [32]. It combines the optimized data storage of AODs with the simple interface of the D3PD format. This is accomplished by storing the data in an auxiliary base class using the *DataVector*[3] format developed for ATHENA while using simple interface classes to access the AuxStore. ATHENA specific dependancies and functions are hidden from the ROOT based user by employing conditional compilation. Lastly, schema evolution is natively supported in the new format.

This new EDM is complimented by other software developments. All tools provided by the CP groups are dual use and support both ATHENA and ROOT based user analysis software [35]. Moreover, due to class based xAOD accessors, the CP tool interfaces are simpler, reducing the effort and debugging required from users. An ATLAS wide 'Analysis Release' system, containing all EDM and CP tools packages, is developed to ease collaboration. Lastly, a 'Derivation Framework' is created to reduce the information in a full xAOD down to a manageable size by only keeping the information used by a specific analysis. This smaller data format is referred as derived xAOD (DxAOD).

In addition to changes to the EDM, significant software updates have also been made to the simulation and reconstruction frameworks. The biggest changes for simulation include switching default generators, moving to new a Geant4 version and geometry update to match the Run 2 detector. For track construction, a factor of 3 speedup is gained by migrating the base code to EIGEN [36] and new vertex seeding is

[3]DataVector format is similar to *std::vector* but it additionally owns the vector elements.

developed for improved scaling with pileup. Muon, electron and jet reconstructions are updated improve resolution and apply new offline methods developed in Run 1.

Due to the changes in the EDM, Run 1 user analysis software is incompatible with the data expected to be collected in the upcoming run. This prompts the development of a new framework for $H \rightarrow ZZ^{*} \rightarrow 4\ell$ analysis to process xAODs to minitrees. Furthermore, this complete rewrite provides an opportunity to incorporate the successes and correct the failures of the old code base. Lastly, by employing the new framework, changes to the ATLAS software can be validated and their impact on the final measurements can be quantified.

6 Analysis Framework

6.1 Role of Analysis Framework

For a physics group, the need for an analysis framework arises to aid in reducing AODs/D3PDs/xAODs to a manageable size and format that then can be subsequently used for producing final results and plots. This is accomplished by imposing specific analysis cuts and vetoing events that do not pass the full selection. Moreover, in the final output, only variables needed for the final analysis and plotting are saved by the framework.

Table 4 summarizes the cuts used by $H \to ZZ^* \to 4\ell$ (H4l) for its nominal analysis. These are designed to select events with 4 leptons[4] which match the kinematic distribution of leptons originating from a $H \to ZZ^*$ decay. Furthermore, these cuts are also designed to reduce background events where the final state is created by $pp \to ZZ^* \to 4\ell$ decays or $pp \to ZZ^* \to 2\ell 2q$ decays, where the jets are misidentified as leptons.

In addition to the nominal cutflow, many other cutflows which contain slightly varied cuts are also implemented in a framework for the H4l analysis. For estimating backgrounds, control regions are created by either inverting and relaxing isolation requirements, relaxing identification requirements on electrons or selecting quadruplets with different charges and flavor compositions. For estimating the impact of systematics on the final results, identical cuts to the nominal ones are used but the calibration and smearing applied to leptons/jets are varied according to CP group recommendations. Moreover, for deriving resolution functions, $Z \to 2\ell$ events with similar pre-selection and isolation requirements are selected. Lastly, truth level cutflow for Monte Carlo (MC) events are used for fiducial studies.

Once the events have been selected, variables important to the final analysis are calculated and stored in minitrees. These variables range from calculating invariant mass of selected objects to calculating boosted decision tree scores to simply storing lepton kinematic information. For MC events, weight variables for overall normalization and truth matched information is also calculated and saved in the final output.

6.2 Run 1 Model

In Run 1, H4l analysis had a 'distributed' framework model, as shown in Figure 14. While the same input D3PDs/AODs were by used the entire group, separate codes were written to perform a specific type of cutflow. However, for many of these, multiple frameworks existed which performed either the same exact selection or a slightly varied one. For example, to create background minitrees, separate code bases existed for each specific method, such as the $3\ell + X$ or $Z + XX$ method for estimating $\ell\ell ee$ background [37]. On the other hand for calculating systematic uncertainties, different frameworks produced outputs appropriate for a specific analysis, such as histograms for m_{4l} shape analysis and minitrees for SpinCP/Fiducial analysis. Lastly, for the nominal analysis, many frameworks were written that performed the identical selection and created output minitrees in a standardized format.

While the original design considerations are not known, many benefits of this model have been observed. Due to splitting of specialized cutflows, significantly less strain is placed on one software architecture to provide tools for the entire analysis. Hence, multiple structured and procedural softwares, with simple data flow and control path, were developed. This allowed new users to quickly understand the

[4]Leptons only include muons or electrons.

Table 4: Summary of the nominal H4l event selection requirements.

Event Pre-selection	
Veto	
Veto any event where detector is not working properly	
Triggers	
Single electron, single muon, di-electron, di-muon and electron-muon triggers	
Electrons	
Calibrated Loose Likelihood quality electrons with $E_{\mathrm{T}} > 7$ GeV and $\lvert\eta\rvert < 2.47$	
Muons	
Smeared combined or segment-tagged muons with $p_{\mathrm{T}} > 6$ GeV and $\lvert\eta\rvert < 2.7$	
Maximum one calo-tagged or standalone muon in the quadruplet	
Smeared calo-tagged muons with $p_{\mathrm{T}} > 15$ GeV and $\lvert\eta\rvert < 0.1$	
Smeared standalone muons with $p_{\mathrm{T}} > 6$ GeV, $2.5 < \lvert\eta\rvert < 2.7$ and $\Delta R > 0.2$ from closest segment-tagged	
Jets	
Calibrated R = 0.4 Anti kT jets with $p_{\mathrm{T}} > 25$ GeV and $\lvert\eta\rvert < 2.4$ or $p_{\mathrm{T}} > 30$ GeV and $2.4 < \lvert\eta\rvert < 4.5$	
Overlap removal	
Remove overlap between different physics objects	
Event Selection	
Kinematic Selection	Require at least one quadruplet of leptons consisting of two pairs of same-flavour opposite-charge leptons fulfilling the following requirements: p_{T} thresholds for three leading leptons in the quadruplet 20, 15 and 10 GeV Select best quadruplet to be the one with the leading dilepton mass being the one closer to the Z mass and the second mass closer to the Z one, to be the subleading one. Leading di-lepton mass requirement 50 GeV $< m_{12} <$ 106 GeV Sub-leading di-lepton mass requirement $m_{threshold} < m_{34} <$ 115 GeV Remove quadruplet if alternative same-flavour opposite-charge di-lepton gives $m_{\ell\ell} < 5$ GeV $\Delta R(\ell,\ell') > 0.10(0.20)$ for all same (different) flavour leptons in the quadruplet.
Isolation	Isolation cut applied on all leptons of the quadruplet Contribution from the other leptons of the quadruplet is subtracted Lepton track isolation ($\Delta R = 0.20$): $\Sigma p_{\mathrm{T}}/p_{\mathrm{T}} < 0.15$ Electron calorimeter isolation ($\Delta R = 0.20$) : $\Sigma E_T/E_T < 0.20$ Muon calorimeter isolation ($\Delta R = 0.20$) : $\Sigma E_T/E_T < 0.30$ Stand-Alone muons calorimeter isolation ($\Delta R = 0.20$) : $\Sigma E_T/E_T < 0.15$
Impact Parameter Significance	Apply impact parameter significance cut to all leptons of the quadruplet. For electrons : $d_0/\sigma_{d_0} < 6.5$ For muons : $d_0/\sigma_{d_0} < 3.5$
Event Categorization	
VBF	$N_{jets} > 2$ and $M_{jj} > 130$ GeV
VH-Hadronic	$N_{jets} > 2$ and 40 GeV $< M_{jj} <$ 130 GeV and passes hadronic BDT score
VH-Leptonic	At least 1 isolated lepton
ggF	Any event failing above cuts

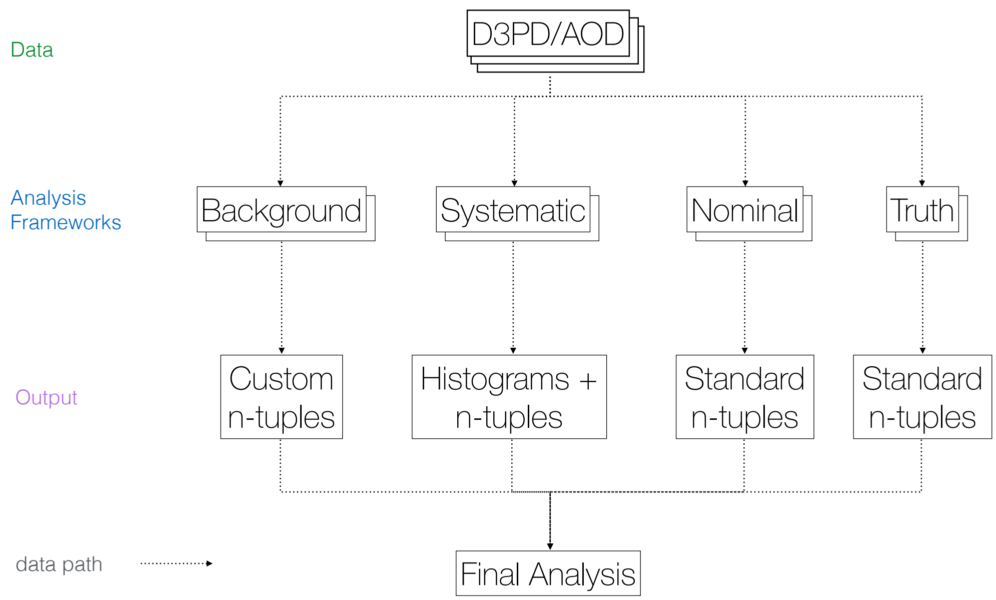


Figure 14: H4l Run 1 analysis framework model. Smaller analysis frameworks, such as for deriving efficiency factors for isolation cuts, tag and probe, etc., are omitted.

underlying logic and appropriately modify it for their use. Moreover, as these frameworks are written and used by researchers whose work directly depends on it, they are constantly maintained and kept updated with the current recommendations.

Due to the duplication of frameworks performing similar tasks, H4l sub-analyses were able to do 1-to-1 testing to find and fix coding errors. Typically in ATLAS, differences of a few percent in the cutflow are considered acceptable; however, by directly comparing the outputs, different analysis frameworks were able to achieve an accuracy of 1 part per million at a precision of 0.1% and use a common method to calculate variables. This ensured that similar results were obtained irrespective of which software was used to produce the minitrees.

Despite the success, many failures of this model were also seen in Run 1. As many parts of the 'Event Pre-selection' and 'Event Selection' (see Table 4) are similar between the different cutflows, there is significant code duplication. This leads to increased maintenance overhead as changing CP recommendations and selection cuts have to be propagated individually to each framework and validation of one software does not carry over to others. Furthermore as these codes are written by different people, they have varying data structures and interfaces, forcing users to understand different implementations to perform the needed tasks.

Another failure observed in Run 1 is the widespread usage of *god classes* and *functions*. In this case, typically 10 000 line long classes or functions are written which perform the entire chain of selection with only small tasks delegated to sub-routines. This type of approach arose due to the rate of changing recommendations outpacing any software architectural changes and the constant pressure to produce results. While not all god objects are harmful [38], in the H4l framework, these led to interdependent and fragile systems where any changes, such as crucial improvements to memory management, broke the entire framework in non-trivial ways.

6.3 Run 2 Model

6.3.1 Requirements and Constraints

Based on the major design failures and successes, as mentioned in Section 6.2, and from the practical experience of using the Run 1 model, the requirements and constraints to guide the Run 2 model design are derived as follow[5]:

- The Run 2 analysis model must fulfill the roles described in Section 6.1. It must be able to implement the major cutflows needed for the H4l analysis, which include, but are not limited to, nominal, background and systematic cutflow. The output of the designed framework must be a flat ntuple as described in [39]. This is based on the observed usability of minitrees and to limit the changes required for other Run 1 H4l softwares.

- The designed framework must have increased modularity when compared to the Run 1 model. Consistent parts of the cutflow must be grouped together and implemented in separated modules. These should be as isolated and objects definition independent as possible. This will ensure the ability to replace certain parts of the cutflow, if the need arises. The expected use cases are implementing modified cuts for H4l high mass and width analysis and addition of a control region selection for background estimation.

- The analysis framework must be flexible enough to incorporate future changes to cutflow and analysis procedures. While it is impossible to predict the changes, use cases that should be considered are the ability to incorporate new weights, inclusions of corrected ID/MS muon mass variables, truth information and systematic variations of selected quadruplets under nominal selection. Other possible changes may include addition of tools for smearing physics objects and vetoing events.

- The designed framework should encourage sharing of tools and modules implementing cuts to reduce maintenance costs, coding bugs and increase homogeneity between different cutflows. The expected use cases are event selection cuts used by the H4l nominal and background cutflows, $Z \to 4\ell$ and $Z \to 2l$ selection.

- The framework should have a simple procedural control and data flow; similar to the Run 1 code base. However, the flexibility and modularity of the system must be preferred over architectural simplicity.

- The designed framework should be extendable to other $H \to ZZ^*$ analyses, especially $H \to ZZ^* \to 2\ell 2\nu$ (H$2\ell 2\nu$) which has similar event selection criteria. It should also have the ability to be used for performance studies such as deriving efficiency factors for isolation cuts.

- The designed framework may include a global container that can be used to share limited information between modules for initialization and job dependant setup.

[5]The key words 'must', 'must not', 'required', 'shall', 'shall not', 'should', 'should not', 'recommended', 'may', and 'optional' in Section 6.3.1 are to be interpreted as described in RFC 2119.

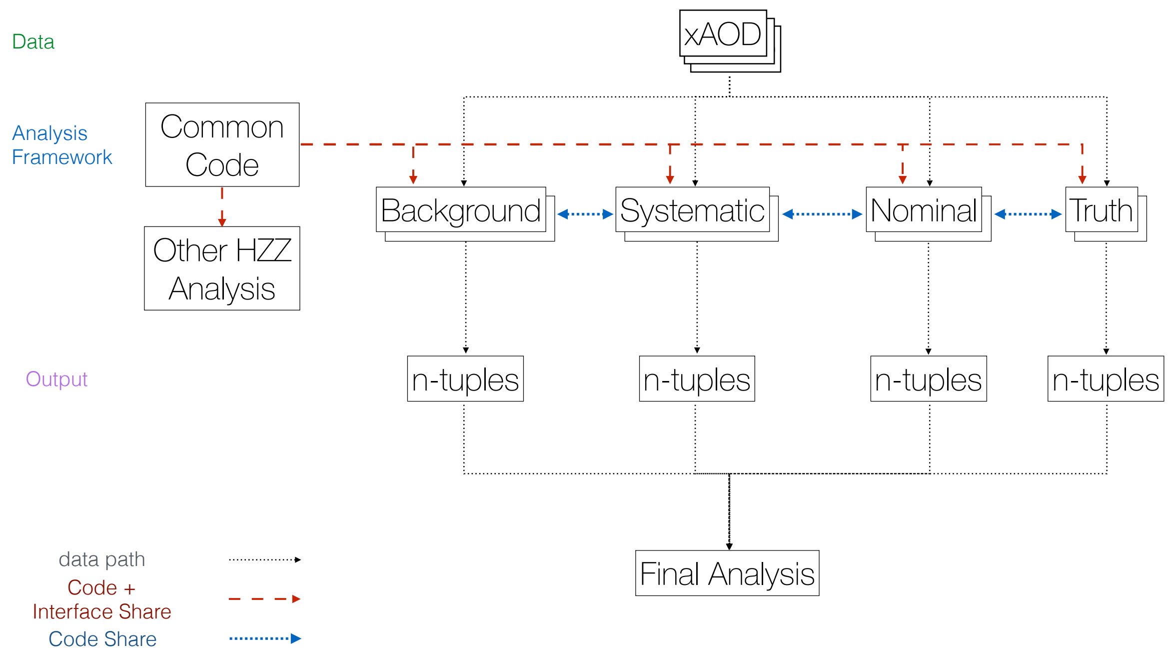


Figure 15: Designed H4l Run 2 analysis framework model. Smaller analysis frameworks, such as for deriving efficiency factors for isolation cuts, tag and probe, etc., are omitted.

- The framework should have a simple interface that can be used by the end user. Minimum information should be required from the user to run the cutflow as initialization and setup should be data-driven.

- Memory management should be explicitly specified within the code to limit memory leaks.

6.3.2 Package Design

Based on the requirements and constraints listed in Section 6.3.1, the overarching Run 2 model for H4l is modified and designed according to Figure 15. The biggest and most important change, with respect to the Run 1 model, is the inclusion of a common code base. This part of the analysis framework will provide interface classes that will create a point of commonality between the different cutflows but are abstract enough to be used by all $H \to ZZ^*$ analyses. Moreover, this will subsequently simplify the process of sharing coded analysis modules as users will have well-defined interaction guidelines and can swap any classes that obey similar procedures.

The common code base will also provide concrete implementation of classes that are required to be similar between all cutflows and different $H \to ZZ^*$ analyses. These typically include modules implementing smearing/calibration and weights such as cross-sections, lepton efficiencies and pile up weight.

For specific analyses, such as H4l or H2ℓ2ν, each will inherit the needed classes and provide concrete implementation for their respective analysis cuts. For example, the common code will provide an event

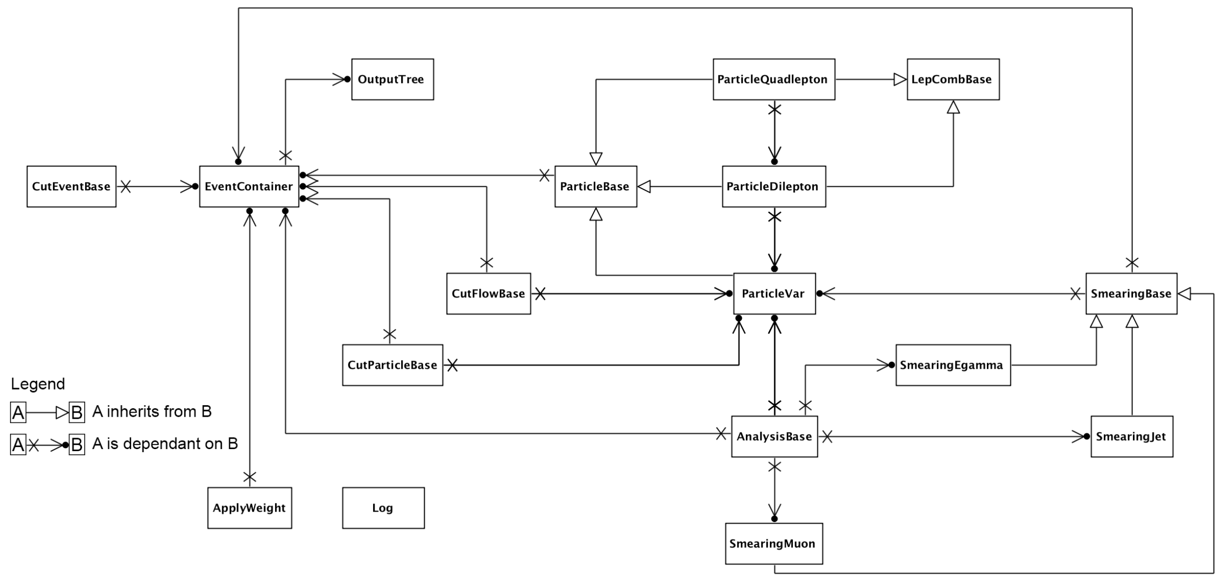


Figure 16: Class structure of the HZZCutCommon package. The dependancy and inheritance tree of each class is graphically outlined.

loop interface which the H4l/H2ℓ2ν analysis will extend by implementing procedures that need to be carried out at each stage of the loop. Furthermore, different cutflows for an analysis will be grouped together in a single package. This is required to reduce code duplication as almost all variations of the nominal cutflow can be implemented by extending relevant classes and modifying a few functions, while still using most of the other modules.

6.3.3 Common Framework

The common code base for the $H \to ZZ^*$ analysis is implemented in a RootCore package [40] named 'HZZCutCommon'. This design choice is due to RootCore being used and supported as an ATLAS wide package manager in Run 2. It provides a standard prescription to organize header and C++ files and provides a dynamic method to generate dependancies and compile the code into a shareable library.

To increase modularity of the framework and encourage code sharing, as specified in the requirements, the simplest way to accomplish this is to implement similar tasks in separate C++ classes. Moreover, to ensure that each user is using the outlined procedures and interfaces when implementing variations to a class, a standard object oriented design (OOD) pattern called *Template Method Pattern* [41] is employed. In this case, abstract classes are provided which implement the flow of an algorithm, where specific virtual functions are called in a defined sequence. The derived classes provide concrete implementations for these functions, and thereby modify the behaviour of the algorithm.

Based on this template method pattern, the common $H \to ZZ^*$ analysis code package is designed. Figure 16 graphically presents the classes and their dependancies. A description of these is as follow:

Figure 17: Design of the AnalysisBase class. Orange highlighted functions are for user interaction and yellow highlighted entries are abstract interface functions for cuts, while non-highlighted variables and functions aid in execution of the algorithm.

AnalysisBase

The AnalysisBase class, as shown in Figure 17, provides the interface that end users will use to configure job options and the main event loop algorithm where the cutflow will be implemented.

To interact with the program, the interface is designed to fulfill the sequence outlined in Figure 18a. This sequence is directly taken from the Run 1 frameworks due to its general applicability. To run any type of cutflow, the user is expected to initialize the AnalysisBase class, provide information about the sample and files to run over and configure the type of cutflow. Based on this, the framework will initialize other objects, process the events and save the output file.

Inside the AnalysisBase class, the algorithm that processes each event is outlined in Figure 18b. This is designed to follow the simple logic flow of Run 1 analysis, as per the requirements in Section 6.3.1. However to increase the flexibility and avoid the previous failures, the class provides a virtual function for each process; these are called sequentially inside the main loop. It is expected that each cutflow will inherit this class and provide concrete implementation for these functions, modifying the resulting behaviour.

Moreover, due to the changes in CP tools and their handling of systematic uncertainties, the AnalysisBase class also provides an option to iterate through all systematics inside the event loop. Therefore, rather than creating a separate framework to handle these variations, the AnalysisBase class can provide systematically varied cutflows using the same concrete implementation. This will reduce code duplication and coding bugs.

EventContainer

The EventContainer class is designed to hold variables that are specific to the sample, control the analysis configuration and to pass information between different classes. As such, almost all other classes in the common analysis package are dependant on this as a pointer to this class is held by them. Moreover, this class also provides the interface to the AnalysisBase class for loading a given event into memory.

ParticleBase

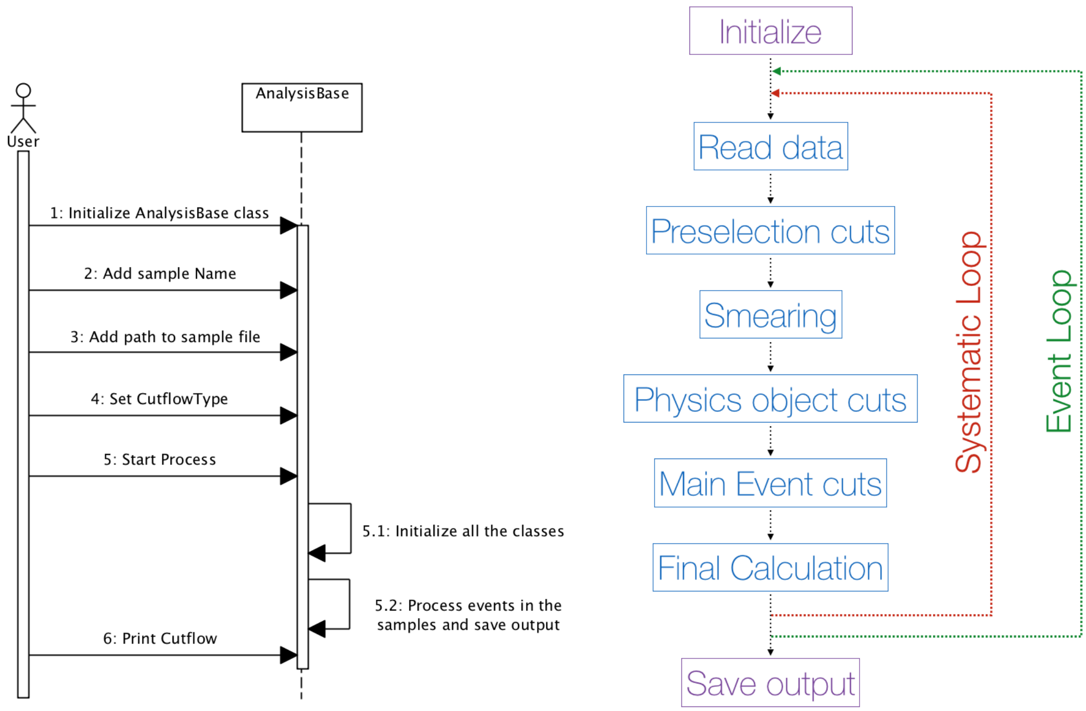
In Run 1, it was observed that a variety of objects, such at multiple TLorentzVectors (TLVs) and error matrices, needed to be associated to a single lepton/jet or to a specific combination to these. As such ParticleBase implements an abstract container class for these objects and provides getter/setter functions. It is designed to be extendable in incorporating any objects that will be required to be paired with a particle.

ParticleVar

ParticleVar extends the ParticleBase class and provides a concrete implementation for single leptons, jets and photons. As such, it stores the xAOD IParticle base pointer [32] for each object and provides a method to associate other IParticles within a single instance of this class. This is designed to ensure a possibility of storing systematic variation of a given particle within a single abstract object. This is to aid in organization and to simplify the cutflow for creating minitrees used in deriving resolution models for the per-event fit method [18]. It is expected that in most standard cutflows, ParticleVar will simply act as a container class for single leptons.

LepCombBase

LepCombBase is a pure virtual class, as shown in Figure 19, that provides a generic interface to loop over a collection of ParticleVar pointers. In 'Event Selection' (see Table 4), for analyses with different



(a) The designed user interaction sequence

(b) Event Loop Algorithm

Figure 18: Interfaces and event loop algorithm provided by AnalysisBase class

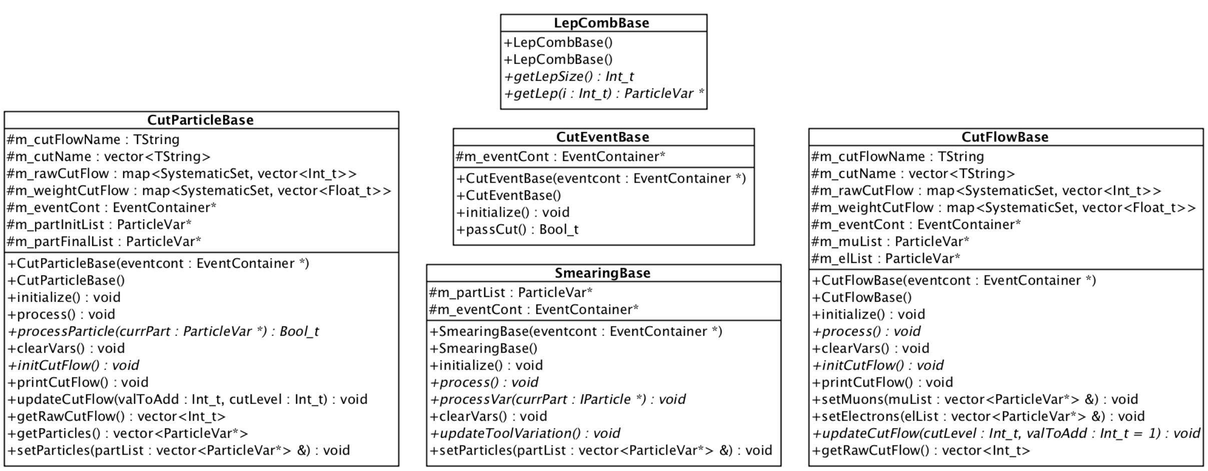


Figure 19: Abstract interfaces provided in the HZZCutCommon package.

particles in the final state but with similar cuts, the exact definition of many cuts is only dependant on the number of particles. Track isolation, calorimeter isolation and d0 significance cuts are examples of this. The interface allows the cuts to be written only assuming the details of this abstract class and, thus, can be used by any other class inheriting from it.

ParticleDilepton and ParticleQuadlepton

In almost all $H \to ZZ^*$ analyses, at a given point in the cutflow, multiples of 2 and/or 4 leptons need to be created to reconstruct the Z and Higgs boson respectively. As such, the ParticleDilepton acts a container class for two ParticleVar pointers and provides simple accessors for them. Moreover, it inherits from ParticleBase to for allow storage of combined properties of leptons, such as TLV and charge. ParticleDilepton can also associate another ParticleVar for final state radiation (FSR) photon [18].

Similarly, the ParticleQuadlepton extends the ParticleBase class and is designed to be a container for two ParticleDilepton pointers. This is explicitly chosen as in $H \to ZZ^*$ analyses, the final state particles are grouped into dilepton pairs to exploit the discriminating properties of the intermediate Z boson. These classes also inherit from and provide a concrete implementation for LepCombBase.

Lastly, as these classes only act as containers, it is explicitly outlined that these do not own the pointers to input ParticleVar. However, any other variable added onto the heap memory is expected to released when the class is deleted.

SmearingBase, SmearingMuon, SmearingJet and SmearingEgamma

SmearingBase class, as shown in Figure 19, provides an interface to aid with smearing and calibrating and to ensure a common prescription for different cutflows. This interface allows users to provide and store a vector of ParticleVars. It contains pure virtual functions to loop over the ParticleVar vector and process each individual object. Furthermore, the design of this class allows external CP tools to be initialized in the constructor and systematic variations to be configured from outside the class. It is also expected that any systematic dependant setup will be implemented in the 'initialize' function.

To increase homogeneity and reduce maintenance cost, the smearing of particles is expected to be exactly the same between all $H \to ZZ^*$ cutflows. As such, the HZZCutCommon provides concrete implementation for smearing muons, photons and electrons and jets in SmearingMuon, SmearingEgamma and SmearingJet classes respectively. To reduce run time memory costs, these classes also provide interfaces to allow for shallow copies of xAOD containers to be used in smearing. As shallow copies are dynamically created, these classes are responsible for releasing this memory. This implies that the ParticleVar class is not allowed to delete the IParticle pointers that are stored within itself.

CutEventBase

In the 'Event Preselection' stage, many cuts, such as trigger veto, are event level cuts. As shown in Figure 19, CutEventBase provides a simple interface. It is designed to ensure that any veto logic is implemented and only accessible through the *passCut* function.

CutParticleBase

In the initial part of the cutflow, cleaning cuts are applied to particles to reduce fakes and misreconstructions. Typically, the procedure followed for these cuts is to loop over the lepton/jet container, veto any particle which fails the cuts and save the rest into another container. CutParticleBase class provides the framework to accomplish this (see Figure 19). Users interact with the class by providing a vector of ParticleVars and after finishing the processing, the class allows the user to retrieve the vector of ParticleVar passing the cuts. Internally inside the class, ParticleVar loop is implemented which calls the 'passEvent' pure virtual function on each vector element. This class is designed under the assumption that any inheriting class will only have to provide a concrete implementation for the 'passEvent' function which contains the relevant cuts.

In comparing different cutflow codes, a typical aspect that is compared is the number of particles passing each cut. As such, this class also provides a method to the user to increment a count vector at each cut. It is internally designed to keep track of different systematically varied cutflows.

CutFlowBase

CutFlowBase is designed to implement most of the 'Event Selection' cuts. As there are great variations at this stage of cutflow between different $H \to ZZ^*$ analysis, this class has attributes of an interface class where the user is expected to provide a concrete implementation of the 'process' function. However, it also provides helper functions to store number of events passing each cut (see CutParticleBase) and setters for the user to provide a list to leptons to use in the cuts. Figure 19 shows a UML class diagram for CutFlowBase.

ApplyWeight

An important aspect of running the cutflow code on MC samples is the application of weights to each event to ensure correct overall normalization and distribution of poorly modelled variables with respect to data. As such, ApplyWeight provides an implementation of common weights such as cross-section, branching ratio and pileup reweighing for all $H \to ZZ^*$ analyses to use. It is expected that analysis specific weights will be implemented by inheriting this class and adding other functionalities.

OutputTree

The final output of the cutflow code for $H \to ZZ^*$ analyses is designed to be a minitree as per the requirements in Section 6.3.1. To aid in this, this class provides helper functions to the user. Typically, to save a flat ntuple, a single variable needs to be mapped to a specific branch in the tree [34]. This inevitably leads to a long list of variables explicitly initialized on the stack memory. However, this is hard to maintain when there are many output variables. To ease this, OutputTree links each variable to a std::map. This allows variables to be initialized on the heap memory and dynamically booked to a single branch in an output tree. This class also provides setters and getters to interact with the map and save output variables. Finally, there are pure virtual functions to control the exact trees that are saved as there are significant variations between different analyses.

Log

The Log class provides a method to output information using the C++ output stream. However, it extends the standard 'cout' function by allowing the user to set different levels for the output to be printed depending on a global setting. These levels include an 'Error', 'Warning', 'Info' and 'Debug' stream depending on the severity of the message.

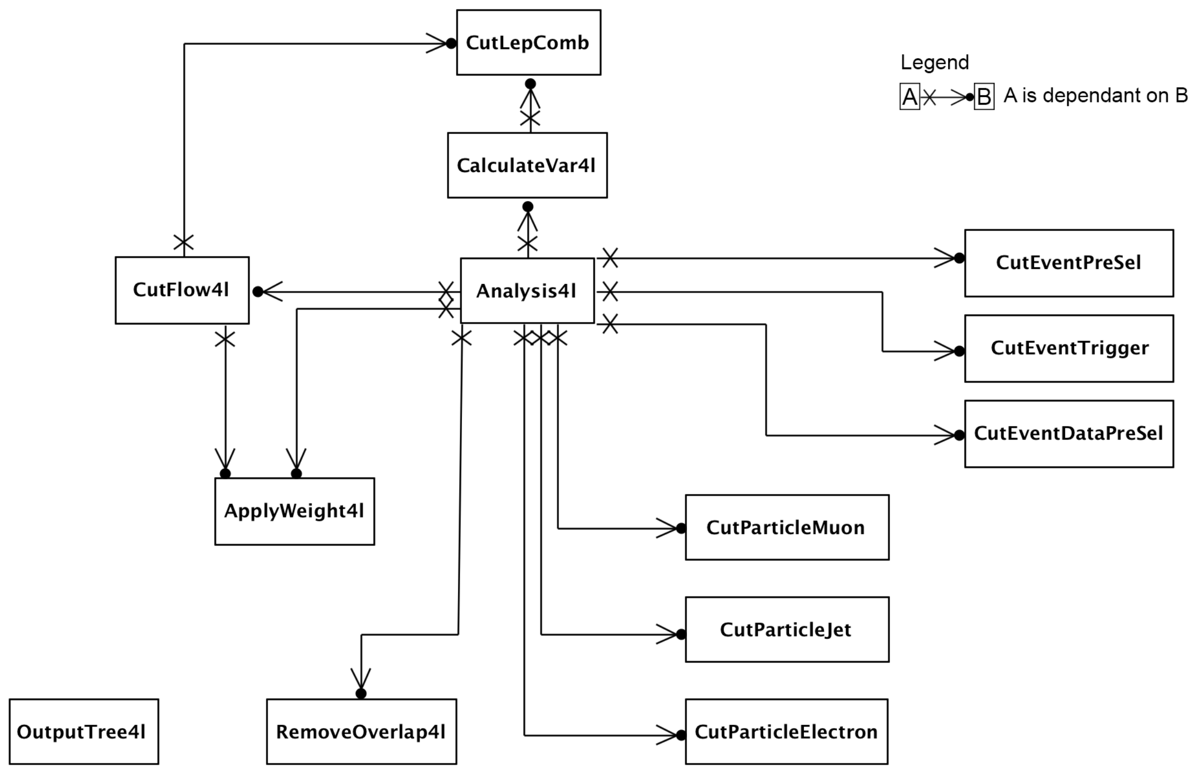


Figure 20: Class structure of the H4lCutflow package. The dependancy structure of each class is graphically outlined.

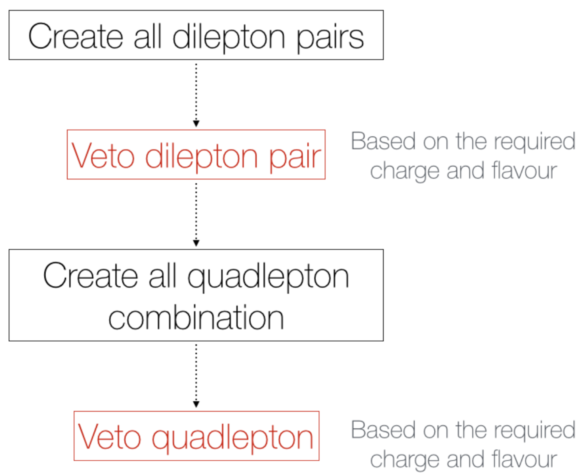


Figure 21: Dilepton and Quadlepton combination algorithm.

6.3.4 H4l Nominal and Systematic Framework

To implement the nominal section (see Table 4) for the $H \rightarrow ZZ^* \rightarrow 4\ell$ analysis, the HZZCutCommon package, as described in Section 6.3.3, is used and a concrete implementation for all virtual classes is provided. This code is grouped in a ROOTCore package called 'H4lCutCommon'. Its dependancy structure is shown in Figure 20.

The 'Event Preselection' cuts in Table 4 are subdivided into further categories to allow for easy maintenance and future extension. These categories are data only preselection cuts, other preselections cuts and trigger cuts and are implemented in *CutEventDataPreSel*, *CutEventPreSel* and *CutEventTrigger* classes respectively. As these are event level cuts, all classes inherit from 'CutEventBase' in the HZZCutCommon package and adhere to the rule that veto logic must only be accessible from the 'passCut' function. However, specific logic is implemented in private helper functions and only combined together in the 'passCut' function.

For the next stage of the cutflow, the particle level cuts are implemented by inheriting the 'CutParticleBase' class and providing a concrete implementation specifically and separately for muons, electrons and jets. Since, each IParticle pointer is wrapped by a ParticleVar class, type checking is explicitly performed to ensure that different particles and selection criteria are not mixed together.

The overlap removal between leptons and jets is implemented in a separate standalone class. This takes vectors of muons, electrons and jets and allows the user to retrieve the ones passing the overlap cuts. This design was explicitly chosen such that it mimics the tool being developed by the 'Harmonization task force' [42]. Once the official tool is ready, the overlap removal implementation will be replaced and removed.

In the package, analysis dependant implementation of weights and outputs is provided in ApplyWeight4l and OutputTree4l classes. For the weights, specific functions are added to calculate lepton reconstruction efficiencies. When the trigger information is added to the MC and data samples, trigger efficiency weights will also be added. For the output minitrees, the class adds variables for lepton kinematics, reconstructed mass, weights, angular variables and event type/categorization and allows the variables to be saved in a single TTree [34].

For the main H4l event selection, a concrete implementation is provided by extending the CutFlowBase class to CutFlow4l. In this class, functions are provided to create dilepton and quadlepton pairs in an algorithmic method as shown in Figure 21. At first, all possible dilepton pairs are created and unwanted ones are vetoed by using a configurable implementation of cuts based on charge and flavour. Similarly, all possible quadlepton pairs are created and then vetoed based on charge and flavour. This is designed to simplify the implementation of other cutflows as discussed in Section 6.3.5.

Once all the quadlepton pairs have been created, the event selection cut logic flow is implemented in the 'process' function of CutFlow4l. However, the exact definition of cuts is implemented in an another helper class called 'CutLepComb'. This separation is chosen as certain cuts are required in the categorization of events, which is implemented in another class. After the cutflow has been performed, CutFlow4l allows the user to retrieve the pointer to ParticleQuadlepton and the last cut it passed.

In the last stage of the cutflow, to calculate all the variables to save in the output file, a helper class called 'CalculateVar4l' is created. It requires a pointer to ParticleQuadlepton, leptons and jets in the event and calibrated xAOD containers from the user. Using this information, it performs FSR recovery, ZMass Constraint [16], obtains BDT scores, categorizes events and calculates other variables. It finally calls upon the OutputTree4l class to save the variables in a output file.

Finally, to bring all these elements together, Analysis4l simply extends the AnalysisBase class by providing a concrete implementation for the virtual functions highlighted in Figure 17. Hence, as shown in Figure 20, the Analysis4l class is dependant on all the other classes in the H4lCutCommon package.

6.3.5 Other H4l Frameworks

While the exact coding of other H4l frameworks is beyond the scope of this thesis, specific procedures are outlined below to aid in future implementation:

High Mass and Width It is believed that there will be homogeneity and consistent selection cuts between the nominal low mass, high mass and width analysis. However, due to experience gained in the final Run 1 analysis, small analysis specific changes, such as the application of a ZMass Constraint [16] only to the leading dilepton pair, may be introduced. It is expected that these changes will be implemented alongside the nominal selection. The application of these cuts will be controlled through configurable boolean flags.

Background To estimate reducible backgrounds in the H4l analysis, the analysis framework must be able to provide minitrees for a variety of control regions (CRs). For estimating $\ell\ell\mu\mu$ backgrounds [37], CRs with either inverted or relaxed isolation and/or d0 significance are required. Other CRs with same sign and opposite flavours quadleptons are also needed. These can simply be created by modifying the cuts in CutLepComb and the quadruplet formation algorithm in CutFlow4l through a global configurable flag. Moreover, support for analysis specific output variables must be added into CalculateVar4l and OutputTree4l.

For estimating $\ell\ell ee$ backgrounds, two separate CRs, $3\ell + X$ and $Z + XX$, are used [37]. In these, electron ID, isolation and/or d0 significance requirements are inverted for one and both subleading leptons respectively. Therefore, to have loose ID selection electron for the quadruplet formation, CutParticleElectron needs to be modified to allow for relaxed cuts. Furthermore, as multiple quadruplets can be saved per event for these CR, a completely new derived CutFlowBase class must be written with a different quadruplet formation algorithm and logic flow from CutFlow4l. Finally, CalculateVar4l and OutputTree4l must be modified to include analysis specific variables.

Lastly, for estimating transfer factors [37], a $Z + X$ CR is required. As the final state has 3 leptons, a new container class must be added to HZZCutCommon package. Moreover, specific CutFlowBase, CalculateVar4l and OutputTree classes must be developed for this final state as well.

$Z \to 4\ell$ The selection for $Z \to 4\ell$ is very similar to the nominal selection as it is based on the H4l analysis [43]. The differences lie only in a loosening of muon p_{T} and leading/subleading dilepton mass cuts. As the definition of these cuts in CutLepComb class is modifiable, this analysis cutflow can be implemented alongside the nominal selection by controlling the specific values of these cuts through configurable flags.

$Z \to 2\ell$ As the $Z \to 2\ell$ (Z2l) cutflow only has two leptons in the final state, a new implementation of the CutFlow4l, CalculateVar4l and OutputTree4l must be created for this analysis. For the main event selection, the derived CutFlowBase class for Z2l will be very similar to CutFlow4l as it will use all cut functions from CutLepComb due to their generic implementation. Furthermore, the modified version of CalculateVar4l and OutputTree4l classes will provide analysis specific output variable calculations and tree structure.

Truth and Fiducial To implement truth selection, modifications are required in both HZZCutCommon and H4lCutFlow packages. Support for different truth particle types must be added into the ParticleVar constructor. Simply creating a separate enum for all the truth particles will create an unmaintainable list, therefore a combination of enums and truth ID numbers must be used to identify different types of truth particles. Truth specific particle cuts must be implemented in a CutParticleBase derived class. Cut definitions in CutLepComb must be extended to include truth particle type. Lastly, CalculateVar4l and OutputTree4l must be modified to include support for truth level variable output.

Performance studies In Run 2, for many performance studies required for the H4l analysis, it is expected that results will be provided by the CP groups. However, if the need arises in the future for a specific study, it is expected that the required classes and parts from HZZCutCommon and H4lCutFlow will be used in a standalone code. No specific designed algorithm or classes are provided for these types of cutflow due to a wide variety used techniques used and small probability that this type of analysis will be implemented using the code base.

6.3.6 Other HZZ frameworks

For other $H \to ZZ^*$ analyses, a similar type of package as H4lCutflow can be developed based on the design considerations mentioned in Section 6.3.4 and Section 6.3.5. For all algorithmic aspects, logic and data flow, the different analyses are exactly the same and, hence, can provide a concrete implementation for the HZZCutCommon package to implement their specific cutflows. However, HZZCutCommon must be extended to incorporate some of the jet and neutrino final state specific variables.

7 Validation studies

Before the framework specified in Section 6 can be used to produce minitrees for the analysis, the common and H4l specific code base needs to validated. Two approaches were employed: a simple cutflow comparison to mainly find coding bugs and a detailed comparison with Run 1 output to understand the differences created by the software changes mentioned in Section 5.2.2.

7.1 Cutflow Comparison

A standard procedure used in ATLAS to validate a user analysis code is to compare the number of events passing each analysis cuts for a selected sample. In the HZZ group, this is typically organized in an official 'Acceptance Challenge' where people with independent analysis codes are invited to post and compare their cutflow numbers.

For the latest comparison [44], the first 5000 events of a data challenge (DC14) sample were used, namely '*mc14_13TeV.167892.PowhegPythia8_AU2CT10_ggH125_ZZ4lep_noTau.merge.AOD.e3292_s1982 _s2008 _r5787_r5853*'. The acceptance challenge validation consisted of comparing the cutflow in each H4l analysis category and overall lepton selection cutflow. A typical example of the event selection numbers compared is shown in Table 5. Using this validation technique, small coding bugs were found and fixed in the designed analysis framework.

Table 5: Cutflow for $H \to ZZ^* \to 4\mu$ analysis produced by the designed analysis framework.

Selection cut	No smearing	With smearing	With smearing and weights
Total	5000	5000	5000
Data Preselection	5000	5000	5000
Preselection	5000	5000	5000
Trigger	5000	5000	5000
Leptons	585	586	586
SFOS	582	583	583
Kinematic	557	558	558
Trigger match	557	558	558
Z1 mass	552	553	551.5
Z2 mass	506	506	504.6
J/ψ veto	497	497	495.7
Track Isolation	254	256	255.3
Calo Isolation	239	241	240.3
Impact Parameter	219	221	220.3
ggF category	208	184	183.5
VBF category	10	29	29.8
VH-Lep category	1	7	1.0
VH-Had category	0	1	6.9

7.2 Run 1 Comparison

To understand the differences created by changing the event data model and the reconstruction software updates for Run 2, a direct comparison of variable distributions is performed for similar MC samples created under the different run conditions. The sample used as the Run 1 reference is *mc12_8TeV.167892. PowhegPythia8_AU2CT10_ggH125_ZZ4lep_noTau.merge.NTUP_HSG2.e1622_s1771_s1741 _r4829 _r4540 _p1344*. This D3PD was used in the final Run 1 H4l analysis and was processed using the Run 1 model with all final CP recommendations. For the Run 2 sample, *mc14_13TeV.167892.PowhegPythia8 _AU2CT10_ggH125_ZZ4lep_noTau.merge.AOD.e3292_s1982_s2008_r5787_r5853* xAOD is used. This sample was produced using AtlasProduction 19.1.1.5 cache, which contained all the major updates to the EDM and reconstruction but still used the Run 1 detector geometry. As a major bug for isolation variable was found but fixed in later derivation framework releases, the HIGG2D1 derivation of the sample was also employed for validation.

7.2.1 Uncalibrated distributions

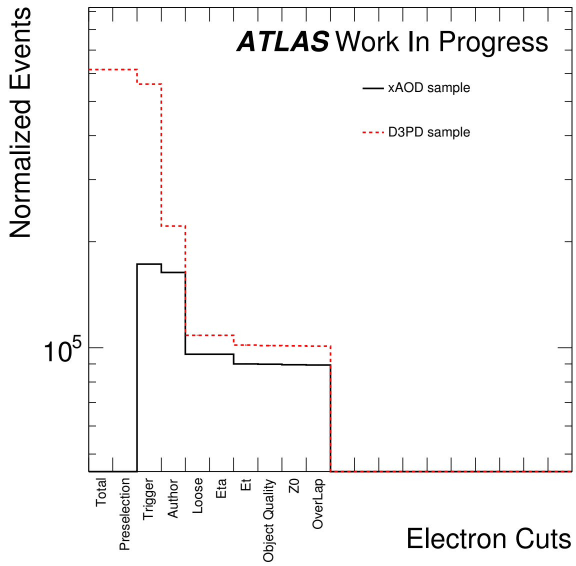
For the initial validation study, a comparison of the variables without any smearing or calibration was performed. This was done to investigate the changes in the raw output of the event reconstruction and understand any underlying differences. As a starting point, the particle and event cutflows between the two samples are compared, shown respectively in Figure 22 and Figure 23.

From the muon and electron cutflows in Figure 22, it is apparent that there are fewer leptons being reconstructed on average in a given event, but after the initial cuts, the level of agreement between two samples is much better. For muons, this difference was traced back to the definition of the 3^{rd} muon chain in xAODs. In this, the overlap between the STACO and CALO reconstruction algorithms is removed from the beginning. While in the D3PD analysis, both algorithms are simply combined, leading to double counting. Lastly, it is observed that in the xAOD sample more muons pass the cuts: this was found to be related to missing smearing in the reconstruction algorithm, leading to better-than-normal p_{T} resolution.

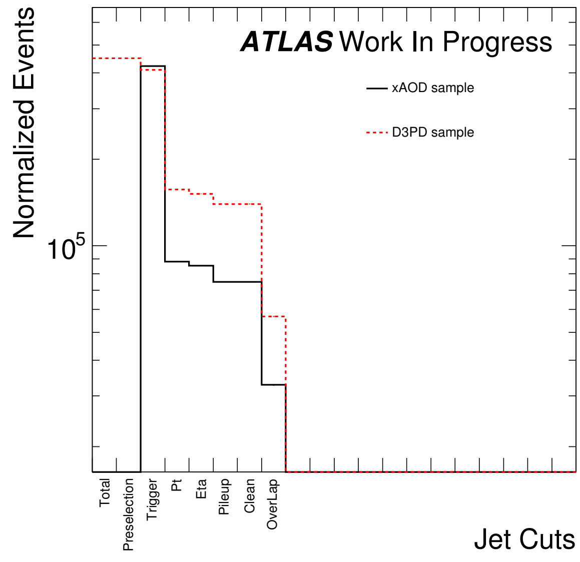
For electrons, the initial mismatch was found to be due to the Egamma CP group splitting them into two separate containers based on the η angular variable. This is easily seen in Figure 24, where in the xAOD sample all electrons with $|\eta| > 2.47$ are stored in the 'FwdElectron' container rather than in the standard 'Electron' collection. However, this change does not affect the final selection for the H4l analysis as all electrons in this region are vetoed. Moreover, it is observed for the xAOD sample, the cleaning cuts have a lower acceptance. This is due to bugs in the electron reconstruction algorithm, which were found and fixed after this sample was created.

The cutflow differences in jets are attributed to a different type of calibration applied in the online reconstruction software. Specifically, in Run 1, the pileup contributions were removed in the offline code; however, in the Run 2, these corrections are performed at the reconstruction stage. This creates a shifted p_{T} spectrum, as shown in Figure 25, leading to a lower acceptance for the jet p_{T} cleaning cut.

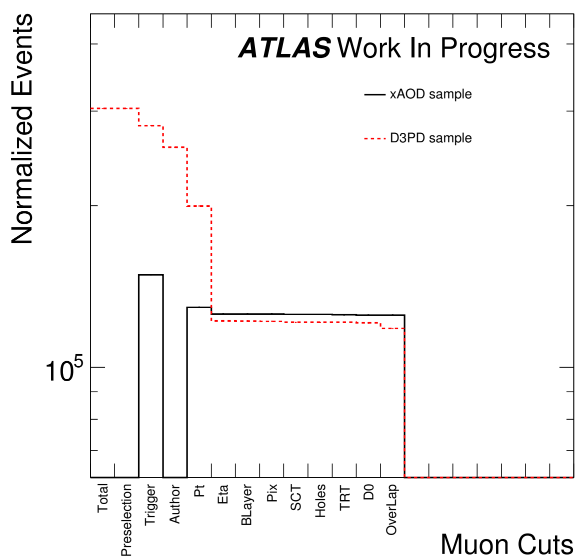
In the event cutflow, the effect of lower acceptances are seen for channels containing electrons. While for a single particle this difference is a few percent, when greater than 3 leptons are required in the 'Lepton' cut, this effect multiplies and significantly fewer events pass this cut. For most intermediate parts of the selection, the relative number of events passing the cuts for both samples are approximately the same. However in the xAOD sample, for the isolation cuts, the acceptance is small compared to the D3PD. This was found to be due to a bug in the code calculating the variables used in the track isolation cut.



(a) Electron cutflow

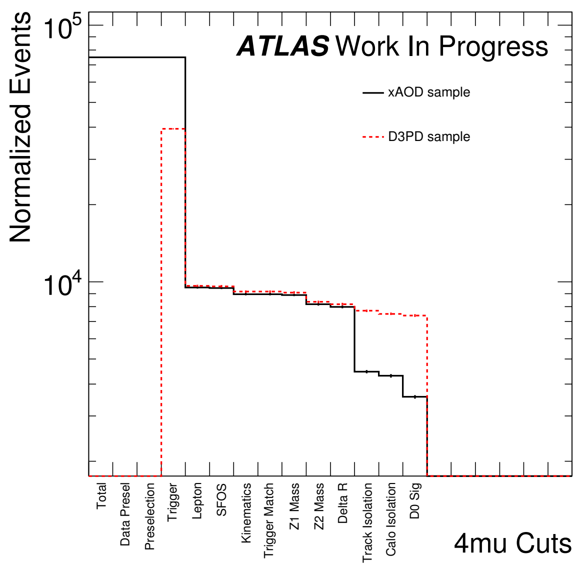


(b) Jet cutflow

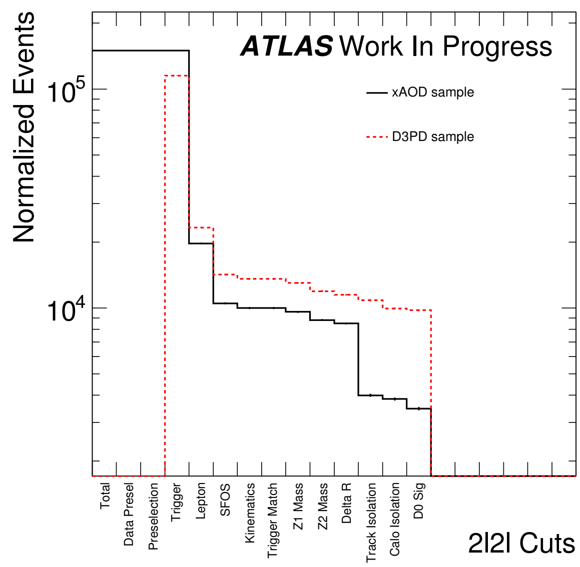


(c) Muon cutflow

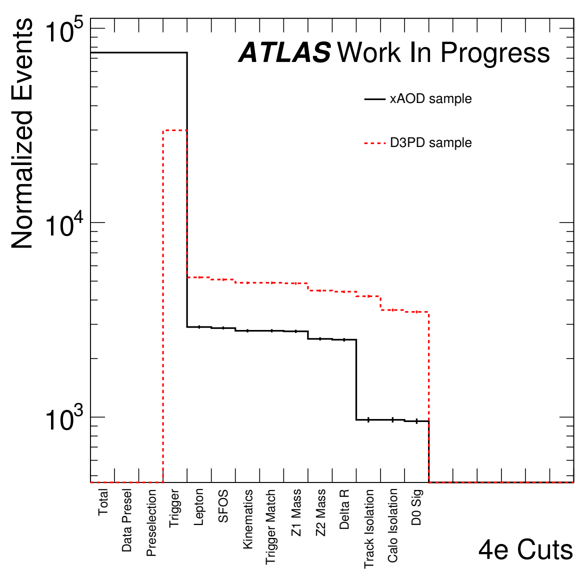
Figure 22: Uncalibrated particle cutflow. Unfilled bins in xAOD histograms are due to different event counting procedures.



(a) 4μ cutflow



(b) $2\mu 2e + 2e2\mu$ cutflow



(c) $4e$ cutflow

Figure 23: Uncalibrated event cutflow. Unfilled bins in D3PD histograms are due to different event counting procedures.

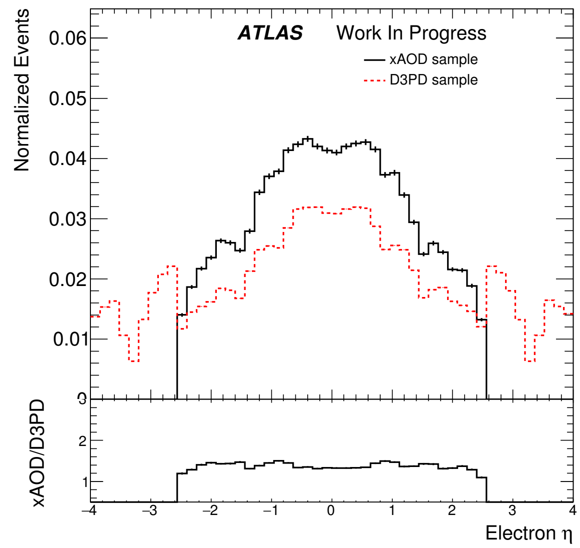


Figure 24: η distribution for electrons in the standard container. Histograms are normalized to 1.

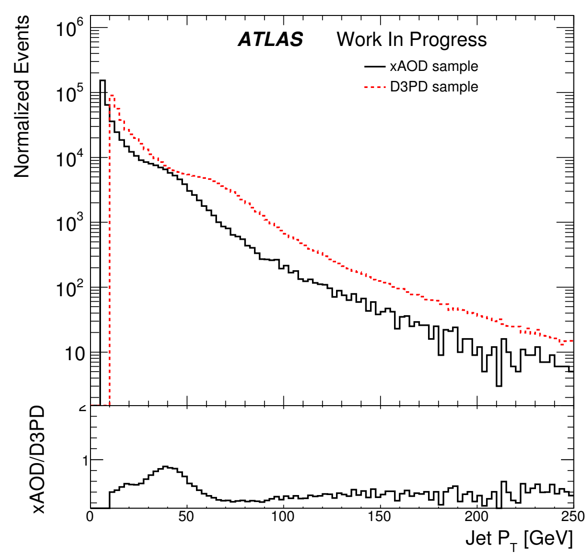


Figure 25: Uncalibrated jet p_{T} distribution.

7.2.2 Calibrated distributions

The comparison of cutflow when the leptons and jets are calibrated according to CP group recommendation is shown Figure 26 and Figure 27. For all plots in Section 7.2.2, two xAODs samples are used as mentioned in Section 7.2. The one labelled as 'xAOD with Iso fix' in the plots is exactly the same as the sample labelled 'xAOD sample', with the only difference being that it has been subsequently processed using the derivation framework. This post-processing thins the containers and provides a fix for the isolation variables. It also vetoes events based on a selection designed to only rejects events that will not pass the offline selection.

From the electron cutflow in Figure 26, it can be observed that there is a better agreement between the xAOD and D3PD sample when compared to the uncalibrated cutflow. This is due to the inclusion of a new electron author in the selection cuts and new Likelihood PDF derived specifically for the DC14 samples by the Egamma CP group. It is expected that with all the fixes in the next MC campaign, the remaining differences will be fixed.

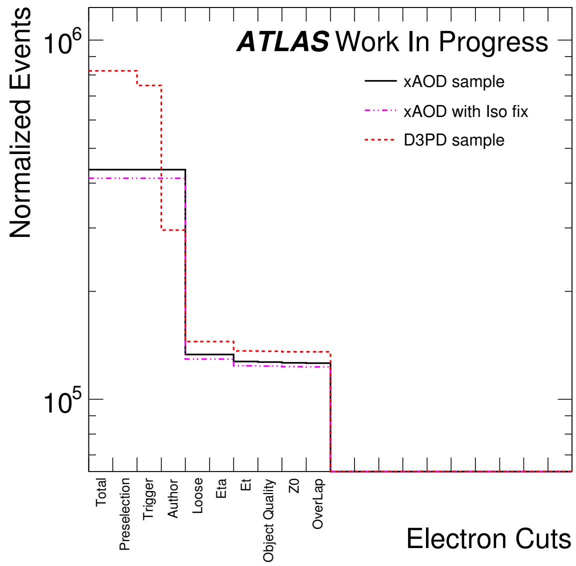
For muons, the calibration provides a fix for the resolution differences, which allows the xAOD muon cutflow to better match the D3PD one. The very small difference in the final selection cut is believed to be due to CALO muons not being reconstructed properly in the DC14 sample. As with the electron reconstructions, this bug is expected to be fixed in the next MC production samples.

Unlike the uncalibrated jet cutflow, for the xAOD sample, a larger fraction of the calibrated jets pass the selection when compared to the D3PD one. This is again attributed to the jet p_{T} spectrum, as shown in Figure 28. For the xAOD jets, the p_{T} spectrum is much harder. This is believed to be due to the increase in the CM collision energy.

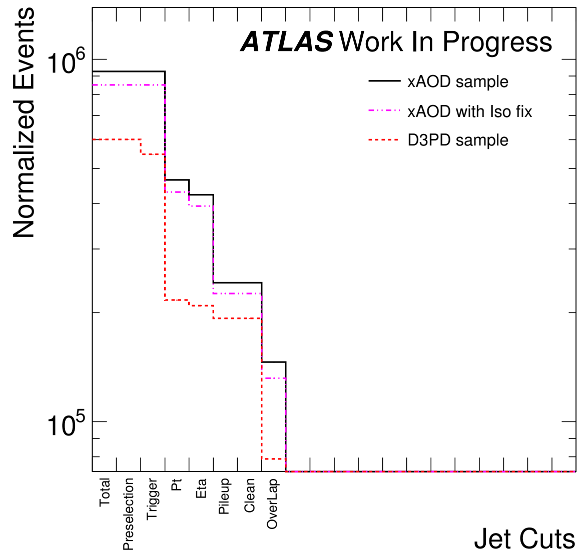
For the calibrated event cutflow, the differences due to lower lepton acceptances are still present. It is also interesting to note that for the two xAOD samples, the cutflows match exactly after the 'Lepton' cut but only until the isolation cuts. This shows that the event veto implemented in the derivation framework is performing as intended. For the isolation cuts, as the post-processed sample contain fixes for the relevant variables, the acceptance for these cuts is much higher. Figure 29 plots the variables related to track isolation cuts for all samples. It can be easily observed that the isolation fix allows the xAOD variables to closely match the D3PD ones.

Finally, for a final level of validation, the variables calculated by the analysis framework and subsequently used in the final analysis are also compared. Mass, p_{T} and η distributions of the selected quadleptons are plotted in Figure 30, Figure 31 and Figure 32 respectively. For the mass and η variables the two samples have a good agreement; however, in xAOD, the p_{T} spectrum is systematically shifted to higher values. This is again believed to be due to the increase in CM collision energy.

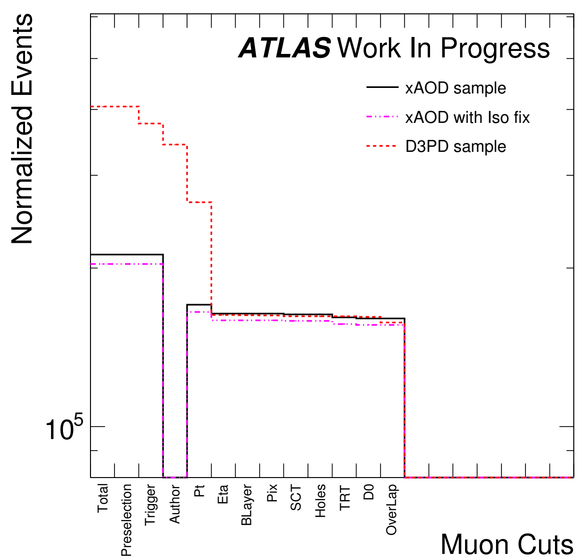
The BDT scores used to separate background, VBF signal and VH hadronic signal are plotted in Figure 33, Figure 34 and Figure 35 respectively. Since the training sets used for these scores are still from Run 1 analyses, the change in p_{T} spectrum creates a shift in the BDT score. In the BDT used to separate the Higgs signal and background, the P_T of the 4-lepton system is used. For this variable, typically the signal spectrum is shifted to higher values when compared to the background. Since, for the xAOD sample, a increase in this variable is seen when compared to D3PD, the BDT score for the new sample is biased towards a more signal-like region. Similarly for the other two BDT variables which use the jet p_{T}, the output score is shifted to a more VBF-like selection.



(a) Electron cutflow

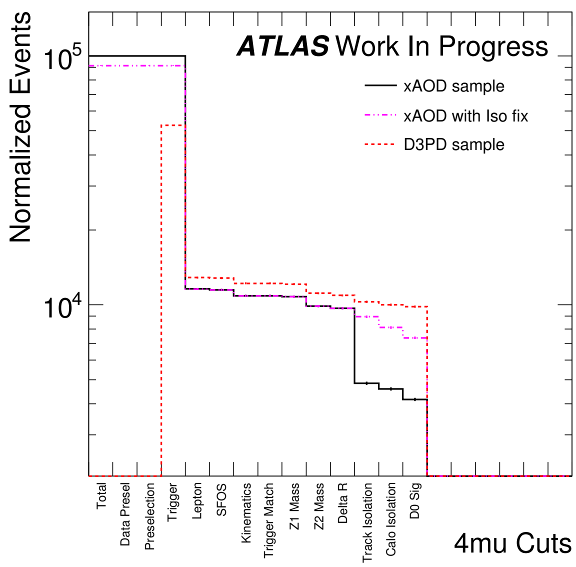


(b) Jet cutflow

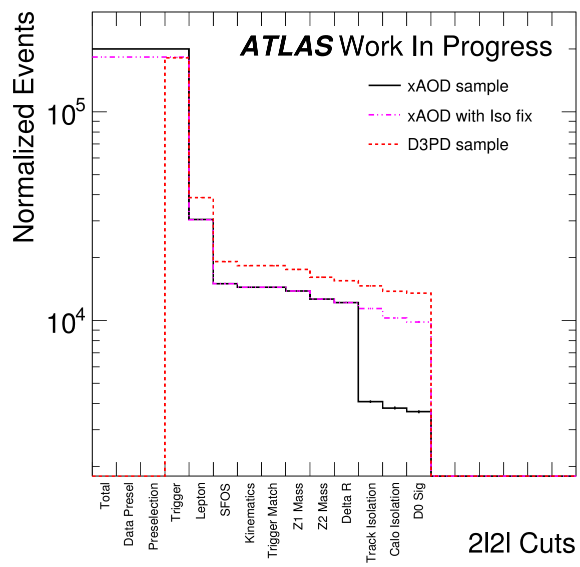


(c) Muon cutflow

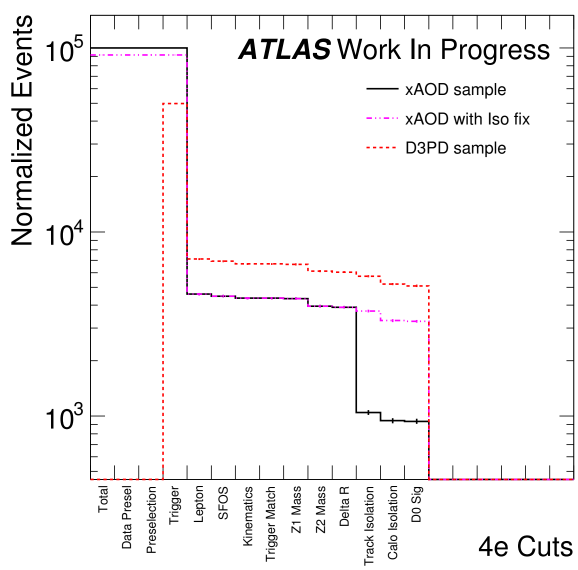
Figure 26: Calibrated particle cutflow. Unfilled bins in xAOD histograms are due to different event counting procedures.



(a) 4μ cutflow



(b) $2\mu 2e + 2e2\mu$ cutflow



(c) $4e$ cutflow

Figure 27: Calibrated event cutflow. Unfilled bins in D3PD histograms are due to different event counting procedures.

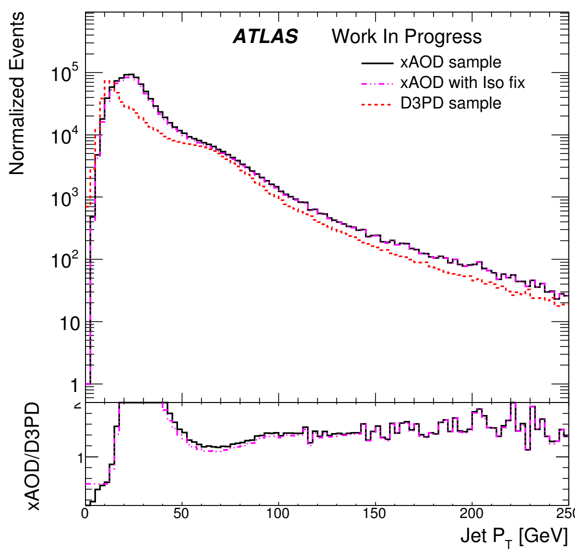
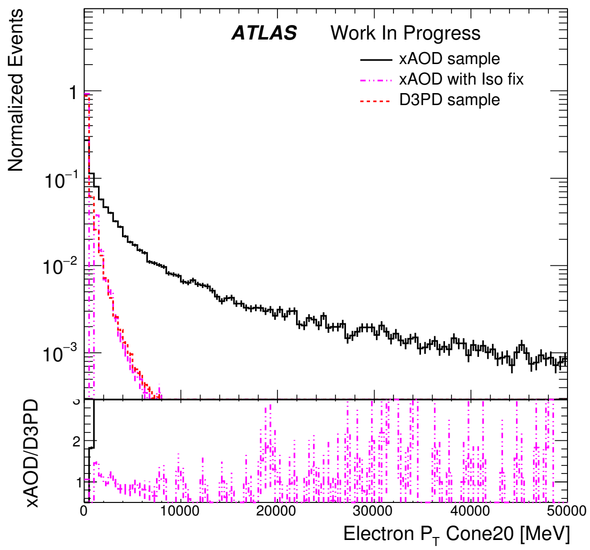
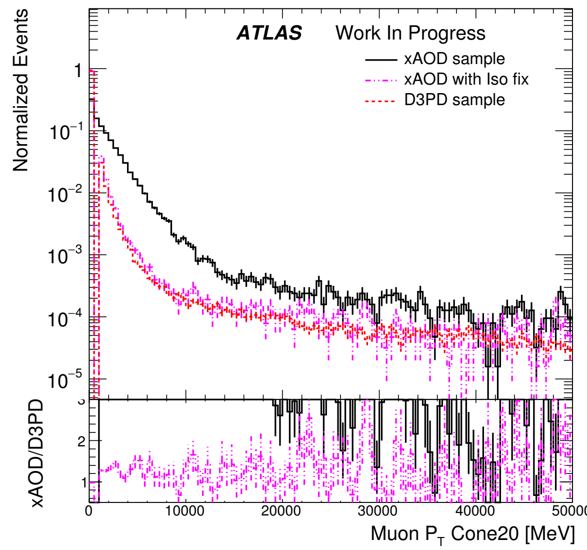


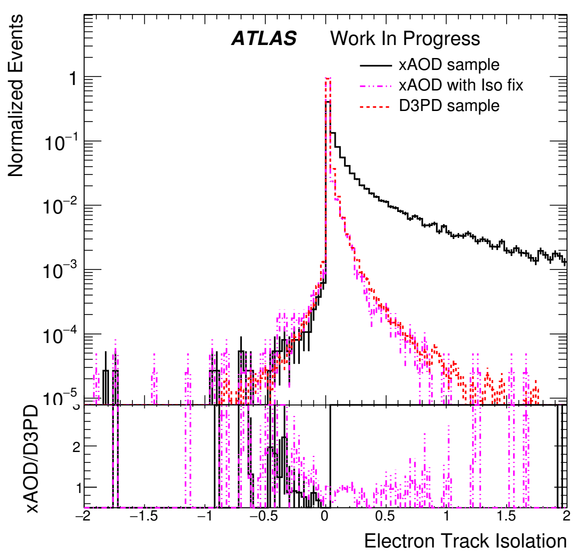
Figure 28: Calibrated jet p_{T} distribution.



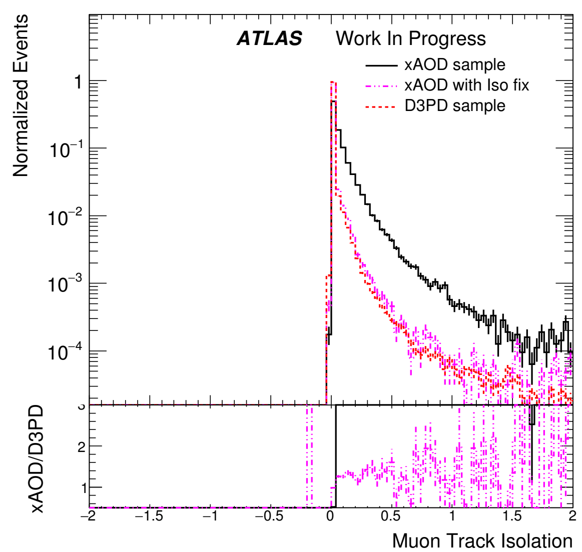
(a) Electron p_{T} cone20 distribution



(b) Muon p_{T} cone20 distribution

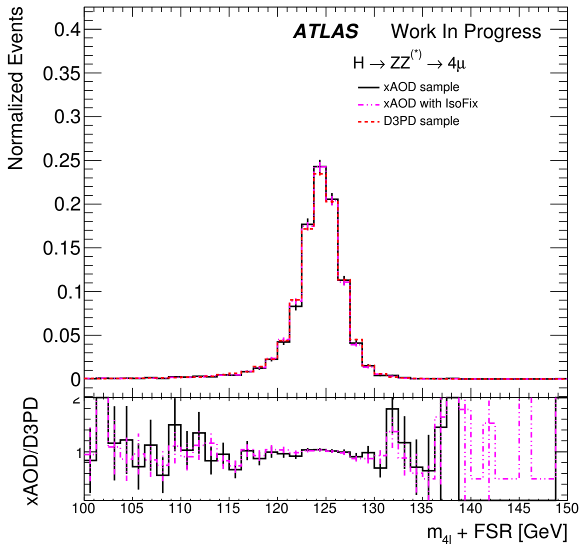


(c) Electron track isolation distribution

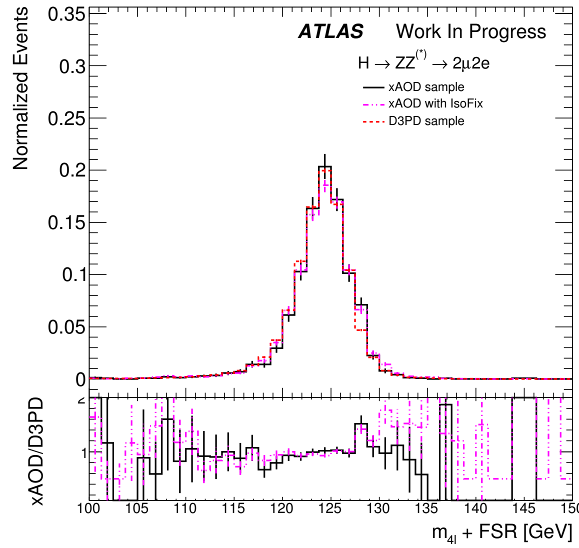


(d) Muon track isolation distribution

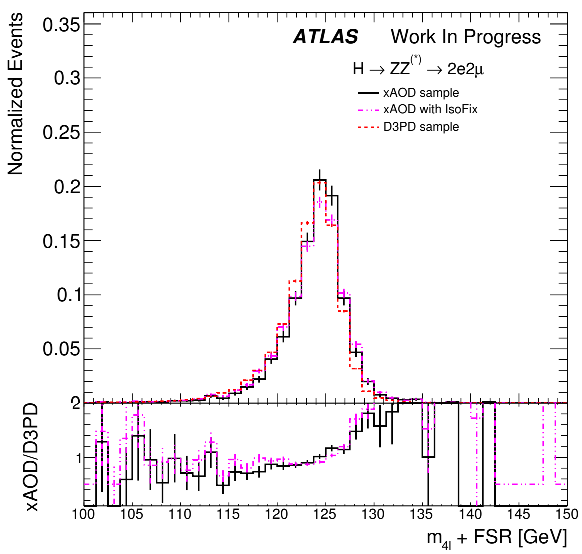
Figure 29: Track isolation variables. All distributions are normalized to 1.



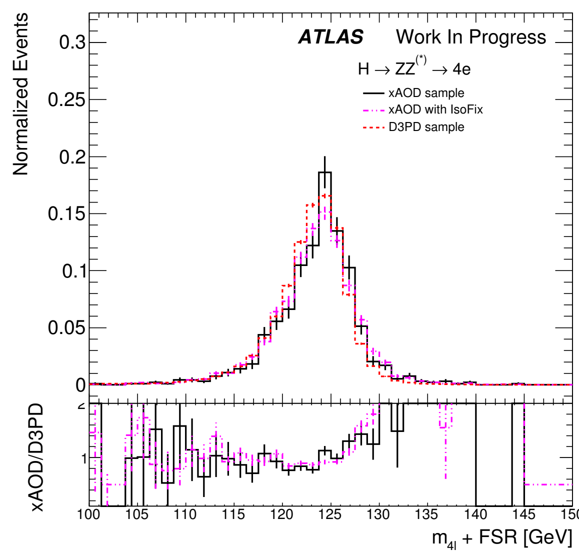
(a) 4μ channel



(b) $2\mu2e$ channel

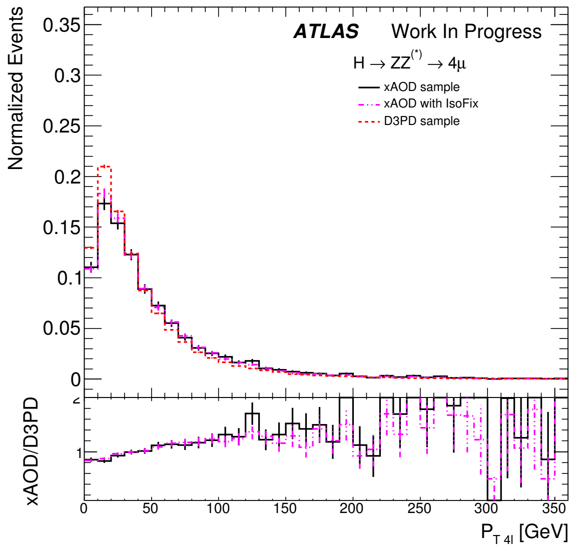


(c) $2e2\mu$ channel

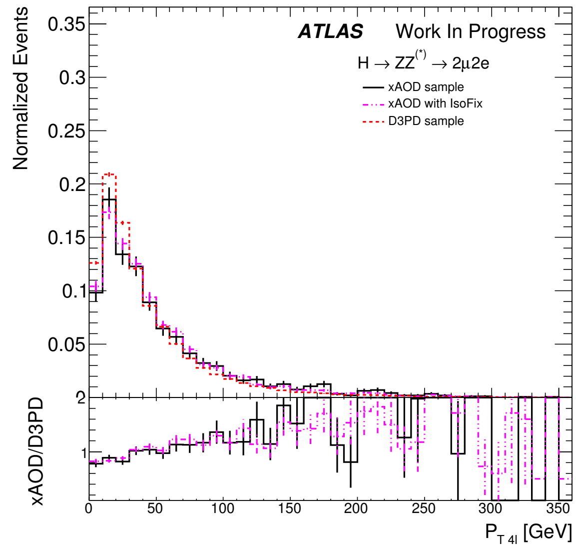


(d) $4e$ channel

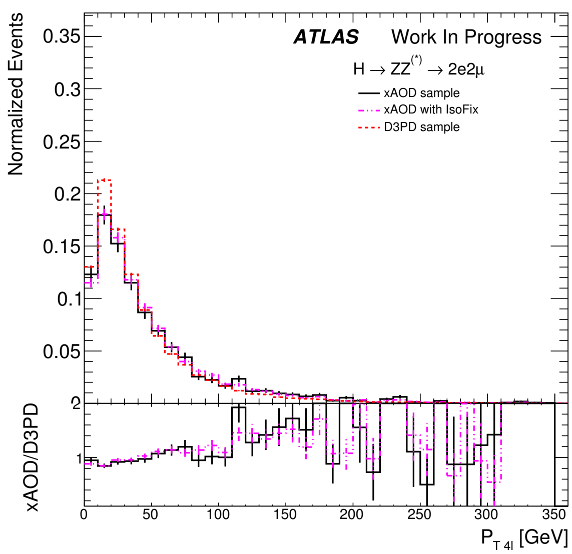
Figure 30: Distributions for final m_{4l} + FSR correction variable used in the H4l analysis. All distributions are normalized to 1.



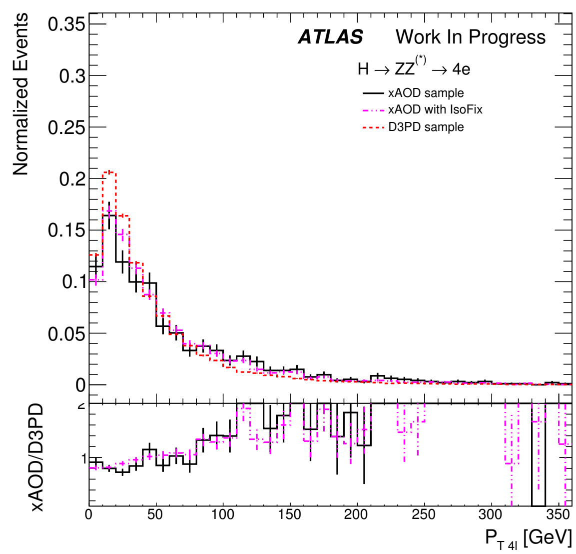
(a) 4μ channel



(b) $2\mu 2e$ channel

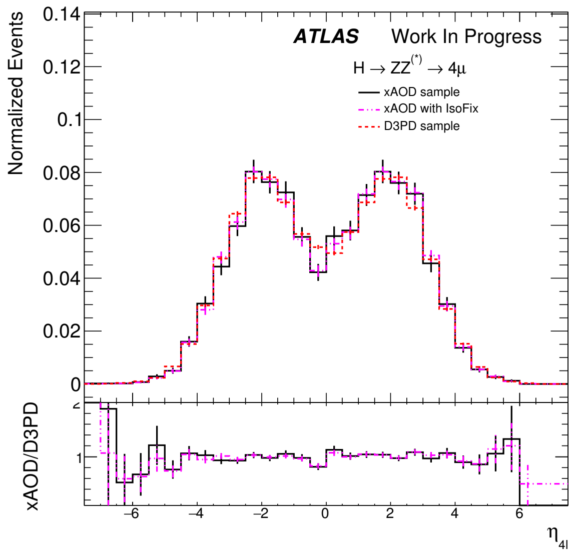


(c) $2e2\mu$ channel

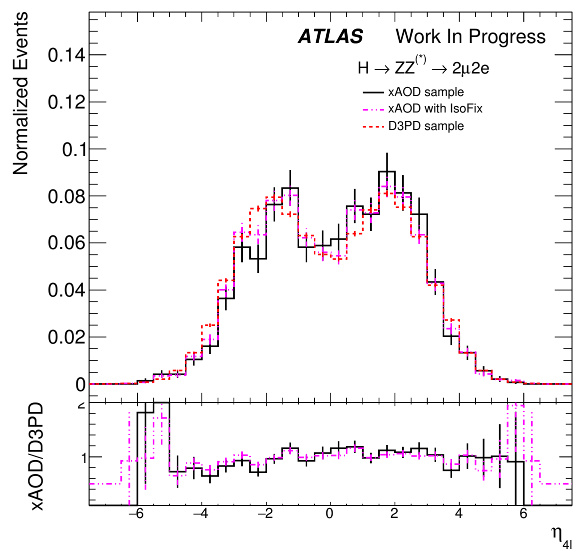


(d) $4e$ channel

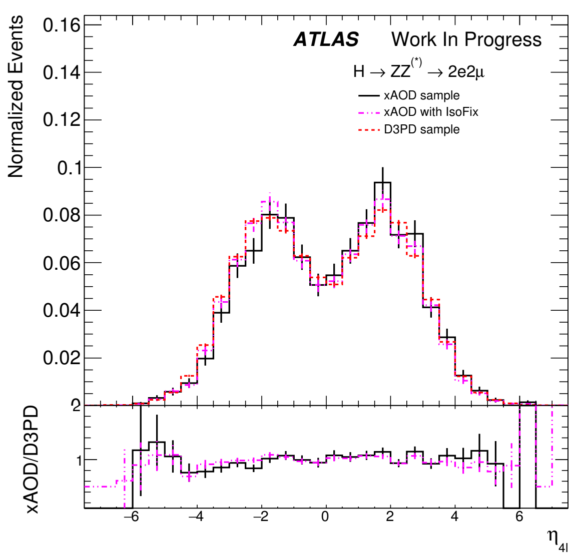
Figure 31: Distributions for final P_{T4l} variable used in the H4l analysis. All distributions are normalized to 1.



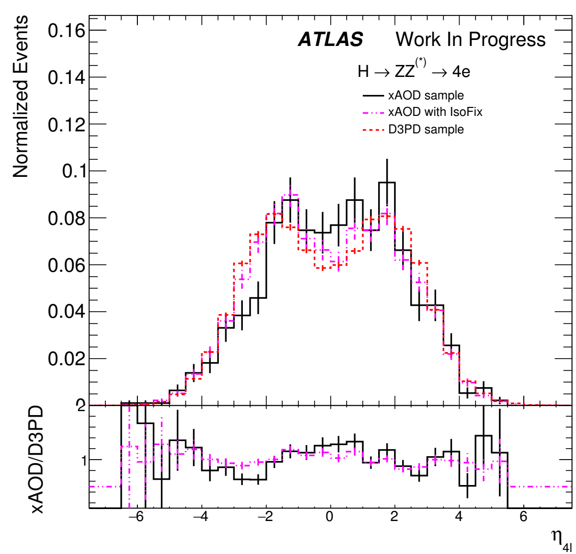
(a) 4μ channel



(b) $2\mu 2e$ channel

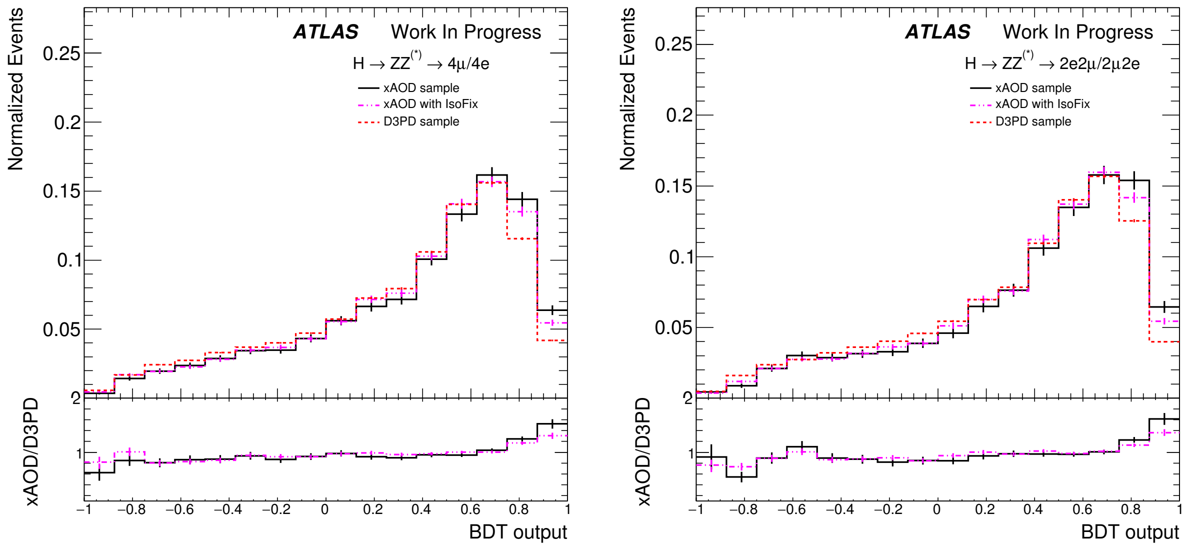


(c) $2e2\mu$ channel



(d) $4e$ channel

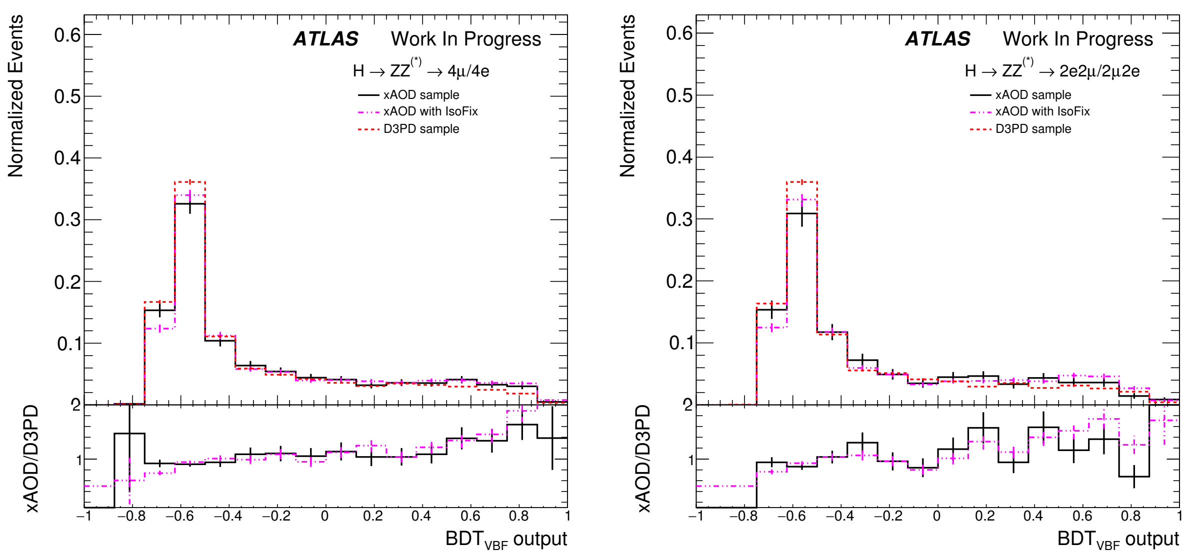
Figure 32: Distributions for final η_{4l} variable used in the H4l analysis. All distributions are normalized to 1.



(a) $4\mu + 4e$ channel

(b) $2\mu 2e + 2e2\mu$ channel

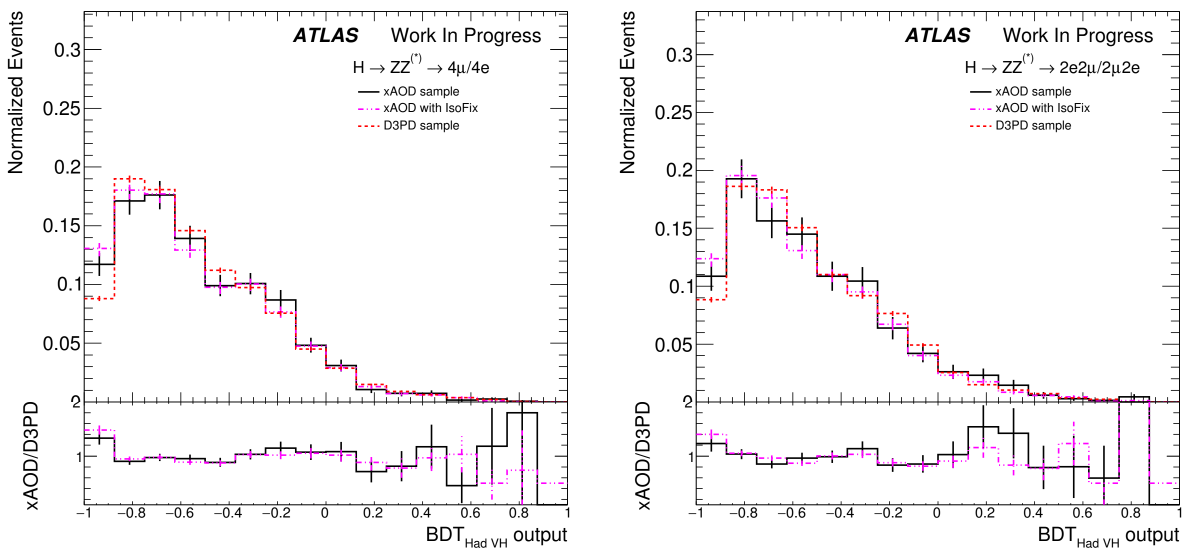
Figure 33: Distributions for final BDT variable used in the H4l analysis. All distributions are normalized to 1. Channels with similar expected distributions are combined to increase statistics.



(a) $4\mu + 4e$ channel

(b) $2\mu 2e + 2e2\mu$ channel

Figure 34: Distributions for final BDT_{VBF} variable used in the H4l analysis. All distributions are normalized to 1. Channels with similar expected distributions are combined to increase statistics.



(a) $4\mu + 4e$ channel

(b) $2\mu 2e + 2e 2\mu$ channel

Figure 35: Distributions for final BDT_{HadVH} variable used in the H4l analysis. All distributions are normalized to 1. Channels with similar expected distributions are combined to increase statistics.

8 Conclusion

In Run 1 of the LHC, ATLAS collected sufficient data to observe a Higgs boson. This particle is a prediction of the Higgs mechanism in the Standard Model, which allows massive gauge bosons in the theory while still preserving the underlying gauge symmetry. Currently, with the collected statistics, it is impossible to pin down whether this newly found particle is consistent with the SM or with some BSM extension of it. It is expected that in Run 2 of the LHC, ATLAS will collect significantly more data and answer some of the remaining questions.

To create a maintainable analysis model in Run 2, the core ATLAS software and EDM underwent major and backwards incompatible upgrades. As such all user analysis software has to be updated. For the $H \to ZZ^* \to 4\ell$ analysis, this allowed a unique opportunity to analyze the Run 1 model and redesign the code base, which is presented in this thesis.

The new analysis framework is designed to ensure flexibility and modularity to incorporate future analysis changes. It, however, trades architectural simplicity to accomplish this. The framework provides classes designed to implement specific parts of the cutflow. The common code package is created such that any $H \to ZZ^*$ analysis can use it by providing a concrete implementation for their own specific analysis cuts. In this thesis, the specific implementation for the H4l analysis is presented and used for initial validation studies.

References

[1] Particle Data Group Collaboration, K. Olive et al., *Review of Particle Physics*, Chin.Phys. **C38** (2014) 090001.

[2] D. Tong, *Lectures of Quantum Field Theory*,. Avaliable at http://www.damtp.cam.ac.uk/user/tong/qft.html.

[3] D. Griffiths, *Introduction to Elementary Particles*. Physics textbook. Wiley, 2008.

[4] M. Thomson, *Modern Particle Physics*. Modern Particle Physics. Cambridge University Press, 2013.

[5] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, *A Historical Profile of the Higgs Boson*, arXiv:1201.6045 [hep-ph].

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

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

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

[9] I. van Vulpen, *The Standard Model Higgs Boson*,. Avaliable at http://master.particles.nl/LectureNotes/2011-PPII-Higgs.pdf.

[10] A. Pich, *The Standard model of electroweak interactions*, arXiv:hep-ph/0502010 [hep-ph].

[11] L. Alvarez-Gaume and J. Ellis, *Eyes on a prize particle*, Nat Phys **7** no. 1, (2011) 2–3.

[12] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Ludwig, et al., *Updated Status of the Global Electroweak Fit and Constraints on New Physics*, Eur.Phys.J. **C72** (2012) 2003, arXiv:1107.0975 [hep-ph].

[13] LEP Working Group for Higgs boson searches, ALEPH , DELPHI , L3 , OPAL Collaboration, R. Barate et al., *Search for the standard model Higgs boson at LEP*, Phys.Lett. **B565** (2003) 61–75, arXiv:hep-ex/0306033 [hep-ex].

[14] Tevatron New Physics Higgs Working Group, CDF , D0 Collaboration,, *Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data*, arXiv:1207.0449 [hep-ex].

[15] LHC Higgs Cross Section Working Group Collaboration, S. Heinemeyer et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, arXiv:1307.1347 [hep-ph].

[16] ATLAS Collaboration, G. Aad, B. Abbott, J. Abdallah, S. Abdel Khalek, O. Abdinov, R. Aben, B. Abi, S. Abidi, et al., *Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data*, Phys.Rev. **D90** no. 5, (2014) 052004, arXiv:1406.3827 [hep-ex].

[17] CMS Collaboration, V. Khachatryan et al., *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, arXiv:1412.8662 [hep-ex].

[18] ATLAS Collaboration, G. Aad et al., *Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, arXiv:1408.5191 [hep-ex].

[19] ATLAS Collaboration, G. Aad et al., *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys.Lett. **B726** (2013) 120–144, arXiv:1307.1432 [hep-ex].

[20] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN, Geneva, 2004.

[21] ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[22] ATLAS Collaboration,, *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997.

[23] H. Spieler, *Semiconductor detector systems*. Semiconductor Science and Technology. Oxford Univ. Press, Oxford, 2005.

[24] M. Bajko, F. Bertinelli, Catalan-Lasheras, et al., *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.

[25] K. Foraz, M. Arnaud, M. Barberan Marin, et al., *LHC Long Shutdown: A Parenthesis for a Challenge*, Conf. Proc. **C130512** no. CERN-ACC-2013-0276, (2013) TUPFI013. 4 p.

[26] K. Foraz, S. Baird, M. Barberan, M. Bernardini, J. Coupard, N. Gilbert, D. Hay, S. Mataguez, and D. McFarlane, *LS1 First Long Shutdown of LHC and its Injector Chains*,.

[27] M. Capeans, G. Darbo, K. Einsweiller, et al., *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19, CERN, Geneva, Sep, 2010.

[28] O. Igonkina, *ATLAS Trigger Menu and Performance in Run I and Prospects for Run II*, Tech. Rep. ATL-DAQ-PROC-2013-042, CERN, Geneva, Nov, 2013.

[29] E. Moyse, *The ATLAS Event Data Model*, Conf. Proc.CHEP 06.

[30] W. Bhimji, J. Cranshaw, P. van Gemmeren, D. Malon, R. D. Schaffer, and I. Vukotic, *The ATLAS ROOT-based data formats: recent improvements and performance measurements*, Journal of Physics: Conference Series **396** no. 2, (2012) 022006. http://stacks.iop.org/1742-6596/396/i=2/a=022006.

[31] ATLAS Collaboration Collaboration,, *ATLAS Computing: technical design report*. Technical Design Report ATLAS. CERN, Geneva, 2005.

[32] A. Buckley, T. Eifert, M. Elsing, D. Gillberg, K. Koeneke, A. Krasznahorkay, and E. Moyse, *Report of the xAOD design group*, Tech. Rep. ATL-COM-SOFT-2013-022, CERN, Geneva, Sep, 2013.

[33] J. Catmore, M. Elsing, E. Lipeles, D. Rousseau, and I. Vivarelli, *Report of the Analysis Model Study Group*, Tech. Rep. ATL-COM-SOFT-2013-005, CERN, Geneva, Apr, 2013.

[34] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, in *AIHENP'96 Workshop, Lausane*, pp. , 81–86. 1996. Avaliable at http://root.cern.ch/.

[35] D. Adams, P. Delsart, M. Elsing, K. Koeneke, E. Lancon, W. Lavrijsen, S. Strandberg, W. Verkerke, I. Vivarelli, and M. Woudstra, *Dual-use tools in ATLAS*, Tech. Rep. ATL-COM-SOFT-2013-125, CERN, Geneva, Dec, 2014.

[36] G. Guennebaud, B. Jacob, et al., *Eigen v3*, 2010. Avaliable at http://eigen.tuxfamily.org.

[37] HSG2, *Event selection and background estimation for the measurement of the properties of the Higgs particle in the four lepton decay channel with the ATLAS detector*, Tech. Rep. ATL-COM-PHYS-2013-1599, CERN, Geneva, Dec, 2013.

[38] S. Olbrich, D. Cruzes, and D. I. Sjoberg, *Are all code smells harmful? A study of God Classes and Brain Classes in the evolution of three open source systems*, pp. , 1–10. Sept, 2010.

[39] HSG2, *H→ZZ→llll Summer 2013 analysis twiki*,. Avaliable at https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HiggsZZllllSummer2013.

[40] N. Krumnack, *RootCore*,. Avaliable at https://twiki.cern.ch/twiki/bin/view/AtlasComputing/RootCore?redirectedfrom=Atlas.RootCore.

[41] R. C. Martin, *TEMPLATE METHOD & STRATEGY*,. Avaliable at http://www.objectmentor.com/resources/articles/inheritanceVsDelegation.

[42] D. Adams, C. Anastopoulos, A. Andreazza, M. Aoki, L. Asquith, et al., *Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014*, Tech. Rep. ATL-COM-PHYS-2014-451, CERN, Geneva, May, 2014.

[43] L. Aperio Bella, J. R. Batley, M. Baumer, T. Dai, A. Daniells, H. Feng, S. French, S. Hou, D. Levin, B. Li, X. Li, J. Liu, N. Lu, E. Mountricha, R. Mudd, K. Nikolopoulos, F. Tarrade, A. Wilson, Y. Wu, L. Xu, H. Yang, Z. Zhao, and B. Zhou, *Measurements of the* $Z \to 4\ell$ *branching fraction and cross-sections in 7 and 8 TeV pp collisions with the ATLAS detector*, Tech. Rep. ATL-COM-PHYS-2013-1088, CERN, Geneva, Aug, 2012.

[44] HZZ, *Run2_AcceptanceChallenge_5k*,. Avaliable at https://docs.google.com/spreadsheets/d/1iCuPbZDMOTF9ltqiyRNZhOVPUrvuUwN64CVurBJJrcg/edit#gid=2011625908.