The ABCStar readout ASIC for the ATLAS Inner Tracker upgrade and evidence for the decay of the Higgs boson to two muons

by

Bryce Norman

A thesis submitted to the
Office of Graduate Studies
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy in Physics

Department of Physics
Carleton University
Ottawa-Carleton Institute of Physics
Ottawa, Canada

January 15, 2026

Abstract

The decay of the Higgs boson into a muon-antimuon pair is a rare process that provides the best opportunity to probe the Higgs Yukawa coupling to a second-generation fermion. This thesis presents an analysis that provides evidence for this decay using proton-proton collision data recorded by the ATLAS detector during 2015–2024. The $H \to \mu\mu$ process has a very small signal to background ratio, so events are divided based on their Higgs production mode and placed into mutually exclusive categories using machine learning methods to enhance sensitivity. This thesis presents the novel use of neural networks trained using a distance correlation (DisCo) loss term, which suppresses unwanted correlations with the dimuon mass and enables the use of neural networks for this analysis, which was previously not feasible. The neural networks with the DisCo loss term are shown to outperform traditional machine learning methods when targeting Higgs boson production modes with complex final states in the $H \to \mu\mu$ analysis.

Using categories defined using machine learning techniques, including neural networks with a DisCo loss function, $305\,\text{fb}^{-1}$ of data collected by the ATLAS detector is analyzed for the $H \to \mu\mu$ signal. The observed significance over the background-only hypothesis for a Higgs boson with a mass of 125.09 GeV decaying to a muon-antimuon pair is 3.4σ (2.5σ expected). This provides stand-alone evidence for the $H \to \mu\mu$ decay with the ATLAS detector.

To further enhance the discovery potential of rare decays such as $H \to \mu\mu$, the Large Hadron Collider will enter the high luminosity phase and the ATLAS detector will undergo upgrades, including the new Inner Tracker. This thesis presents on-

wafer testing and troubleshooting of the ABCStar readout ASIC for the Inner Tracker. Testing was completed through a partnership with an industrial ASIC testing company, DA-Integrated, which was a first for a particle physics detector project. After on-wafer testing, 85.55% of ABCStar chips are approved for use in the Inner Tracker.

Acknowledgments

This thesis has been the culmination of many years of hard work, which would not have been possible without the support of many people.

First, I would like to thank my supervisors, Dag and Thomas, for their support over the years. Thank you for your guidance and for pushing me to strive for excellence throughout my PhD. And thank you for choosing to take me on as a Master's student back in 2019, which set off this journey.

Thank you to everyone that I had the pleasure of working with throughout my PhD. Thank you to James Botte, who taught me most of what I know about ASICs. Thank you to the Carleton physics department and the students I shared an office with over the years. Alex, Callan, Dylan, Ishan, Jeremie, Laura, Matt, Owen, Sam and Zeke, you helped to make the office a more enjoyable place to come in to. A special thanks to two big wankers, Brandon and Ian, who both started this journey with me and were always down to go to Mike's Place for drinks and trivia at the end of the day. Both of you (and Kelsey and Cass) have made these years in Ottawa a time to remember.

Thank you to my parents and to Kerri and Sam for your continued support throughout my PhD. I am grateful for the times that you would drive down to Ottawa to watch a hockey game and pay for a meal for a starving PhD student in need. To all of my friends back home in Kincardine and my friends from McMaster, thank you for all of the shenanigans and fun times over the years.

And of course, thank you to my wife Rachael, who has been by my side the entire time. Thank you for always believing in me and making sure I enjoyed a life outside of my thesis. I couldn't have done it without you.

Statement of Originality

The LHC and ATLAS collaborations consist of thousands of faculty, students, engineers, and technicians from many institutions across the globe, and the work in this thesis is a result of this combined effort. This includes the operation of the LHC and ATLAS detector for data collection and the production of simulated samples used in this thesis. I contributed to jet calibration and uncertainty calculation for Run-3 data and simulation, which is not documented in this thesis.

I had significant involvement with on-wafer testing of the ABCStar ASICs at DA-Integrated which is documented in Chapter 4. Testing of earlier prototypes, the ABC130 and ABCStarV0 was conducted by Thomas Koffas and James Botte from Carleton, together with a test engineer Rob Scouten from DA-Integrated. When I joined Carleton in Fall 2019, I began analyzing data collected testing the ABCStarV0 and later became heavily involved in the test development and data analysis for the ABCStarV1 along with Thomas, James and Rob. This included deciding on pass/fail criteria for chips, implementing analog tests, and adapting digital tests made for a simulation of the ABCStar to be implemented as tests for physical chips at DA-Integrated. All analysis of on-wafer testing data collected at DA-Integrated documented in this thesis was performed by me. I had some involvement in the radiation-hardness testing of the ABCStar mentioned in this thesis, which is also documented in Ref. [1]. I contributed to the formatting and upload of data from on-wafer testing of the ABCStar to the ITk Production Database. This included deciding which test data would be uploaded, aiding with scripts to upload data, and virtually shipping components through the database.

Investigation into the SRAM issue encountered with the ABCStar was a joint collaboration between Carleton University, Rutherford Appleton Laboratories, and the chip designers at CERN. A majority of the work studying the impact of digital voltage and duty cycle on the SRAM performance was performed by me. This information was provided to the chip designers to identify the root cause of the SRAM issue. A paper documenting the ABCStar SRAM issue is currently under ATLAS review to be submitted for publication.

The highly anticipated result establishing evidence for the $H \to \mu\mu$ decay [2] was the work of a large team of over 70 ATLAS researchers across the world. I performed cross checks of the samples produced for use in this analysis and some studies on FSR recovery. I developed NNs trained with a DisCo loss function (Chapter 6), which were found to provide the best significance for the $t\bar{t}H$ categories. I also contributed to the signal modelling, background modelling and calculation of systematic uncertainties and significance for the $t\bar{t}H$ categories included in this paper. The framework for modelling the signal and background distributions and calculating the final expected significance for categories outlined in this thesis and in the paper was developed by other members involved with the paper and built on the framework used for the Run-2 paper [3].

All figures used in this thesis were created by me, unless otherwise indicated with a citation.

Contents

Conventions and Acronyms

The conventions and acronyms used in this thesis are listed below.

Conventions

Due to the energy scale of particle physics measurements, energy is reported in units of the electron volt (eV). This thesis uses natural units, where the speed of light and the reduced Planck constant are set equal to unity, $c = \hbar = 1$. This results in the momentum (eV/c), energy (eV), and mass (eV/c^2) all being in units of energy (eV).

The charge of a particle is expressed in units of the elementary charge. In the Standard Model, electrons, muons and tau leptons all have a charge of -1, while antielectrons, antimuons and antitau leptons have a charge of $+1$. Thesis thesis often does not distinguish between particles and antiparticles, which is a common notation in particle physics. Most notably this applies decays of the Higgs boson or Z boson to pairs of muons or electrons (e.g. $H \to \mu\mu$). This always refers to a decay to a muon-antimuon pair ($H \to \mu^-\mu^+$) or an electron-antielectron pair to obey conservation of charge unless otherwise indicated. This also applies to other cases, such as WH production, where there is a W^+ or W^- boson.

The Standard Model only contains a single scalar Higgs boson. Extensions to the Standard Model occasionally contain multiple Higgs boson particles, sometimes with a charge, denoted with variations of h, H, h^0, H^+. This thesis only considers the Standard Model Higgs boson and is denoted with h or H interchangeably.

Data for this thesis is collected from proton-proton collisions at a circular collider. The energy for these collisions is expressed as the centre of mass energy and written in terms of the Mandelstam variable, s,

$$s = (p_1 + p_2)^2,$$

where p_1 and p_2 are the four vectors of the two incoming particles. For a circular collider, such as the Large Hadron Collider used in this thesis, the centre-of-mass energy of the colliding particles can be expressed in terms of s as

$$E_{\mathrm{CM}} = \sqrt{s}.$$

In particle physics, establishing a signal with 3σ (one-sided) statistical significance over the background for a process is considered "evidence", and establishing 5σ significance is considered "discovery". The probability for the background-only hypothesis to fluctuate to the observed measured signal or higher is given by $p = 1-\Phi(Z)$, where Φ is the cumulative distribution function of a Gaussian distribution. This gives $p = 0.00135$ for 3σ ($Z = 3.0$) and $p = 2.9 \times 10^{-7}$ for 5σ.

Acronyms

ABCStar - ATLAS Binary Chip - Star Version

ALICE - A Large Ion Collider Experiment

AMUX - Analog multiplexer

AMVF - Adaptive multi-vertex fitter

ASIC - Application Specific Integrated Circuit

ATE - Automated test equipment

ATLAS - A Toroidal LHC ApparatuS

$b\bar{b}H$ - Higgs production in association with a bottom quark pair

BC - 40 MHz Bunch crossing clock

BCID - Bunch crossing ID

BCR - Bunch counter reset

BDT - Boosted decision tree

BR - Branching ratio

CERN - European Organization for Nuclear Research

CLK - 160 MHz readout clock

CMD - Command

CMOS - Complimentary metal-oxide-semiconductor

CMS - Compact Muon Solenoid

DA - DA-Integrated

DAQ - Data acquisition

DisCo - Distance correlation

DNN - Deep neural network

DSCB - Double-sided crystal ball

DY - Drell-Yan

ECAL - Electromagnetic calorimeter

EF - Event filter

EMB - Electromagnetic Barrel Calorimeter

EMEC - Electromagnetic Endcap Calorimeter

EW - Electroweak

FCal - Forward Calorimeter

FSR - Final state radiation

ggF - Gluon fusion

GRL - Good runs list

HCCStar - Hybrid Controller Chip - Star Version

HEC - Hadronic Endcap Calorimeter

HL-LHC - High Luminosity Large Hadron Collider

HLT - High level trigger

HPR - High priority register

IBL - Insertable B-Layer

ID - Inner Detector

ISR - Initial state radiation

ITk - Inner Tracker

JSON - JavaScript Object Notation

L0A - Level-0 accept

LAr - Liquid Argon

LCB - L0A/CMD/BCR

LDO - Low-dropout

LHC - Large Hadron Collider

LHCb - Large Hadron Collider beauty

LO - Leading order

LP - Low priority readout request

LUCID - LUminosity Cherenkov Integrating Detector

MC - Monte Carlo

MDT - Monitored drift tube

MIP - Minimium ionizing particle

ML - Machine learning

MS - Muon Spectrometer

NLO - Next to leading order

NN - Neural network

NSW - New Small Wheel

PCB - Printed chip board

PDB - ITk production database

PDF - Parton distribution function

PFO - Particle-flow objects

PMT - Photomultiplier tubes

PMU - Parametric measurement units

PR - Priority readout request

PS - Proton Synchrotron

PSB - Proton Synchrotron Booster

PSU - Programmable power supplies

PV - Primary vertex

QCD - Quantum chromodynamics

QED - Quantum electrodynamics

R3 - Regional readout request

RAL - Rutherford Appleton Laboratories

RAM - Random access memory

RoI - Region of interest

RMS - Root mean square

RPC - Resistive plate chamber

SCT - Semiconductor Tracker

SEU - Single event upset

SLVS - Scalable low-voltage signalling

SM - Standard Model

SN - Serial number

SPS - Super Proton Synchrotron

SR - Signal region

SRAM - Static random access memory

SS - Spurious signal

sTGC - Small strip thin gap chamber

SW - Signal window

TDAQ - Trigger and data acquisition

TGC - Thin gap chamber

tH - Higgs production in association with a top quark

TID - Total ionizing dose

$t\bar{t}H$ - Higgs production in association with a top quark pair

TRT - Transition Radiation Tracker

VBF - Vector boson fusion

VCD - Vector change dump

vev - vacuum expectation value

VH - Higgs production in association with a vector boson

VT50 - Threshold voltage at which a channel has a 50% chance of registering a hit

WP - Working point

Chapter 1

Introduction

The Higgs boson was first predicted in 1964 [4, 5] and was thought to be the last missing piece of the Standard Model of particle physics. The discovery of the Higgs boson came almost half a century later in 2012 [6, 7] at the Large Hadron Collider by the ATLAS and CMS collaborations. Since its discovery, measurements of the Higgs coupling to many different particles in the Standard Model have been performed. The strength of the coupling between the Higgs boson and fermions is proportional to the fermion's mass. This means that a fermion with larger mass will tend to interact more frequently with Higgs bosons, and such an interaction will be easier to identify. Particles in the Standard Model can be grouped into different generations depending on their masses, and previous measurements of Higgs couplings focused on the third generation, which is the most massive. One of the next benchmark results in Higgs physics is to measure the Higgs coupling to a less massive, second generation fermion called the muon.

This thesis probes the coupling between the Higgs boson and the muon by searching for

the Higgs decay to a muon-antimuon pair using proton-proton collision data collected by the ATLAS detector. The interaction between the Higgs boson and fermions is the only mechanism in the Standard Model that distinguishes between fermion generations, so this study provides insights into the foundations of the Standard Model and the origin of the fermion mass hierarchy.

The previous ATLAS search for the Higgs boson decaying to a muon-antimuon pair had an observed (expected) significance over the background-only hypothesis of 2.0σ (1.7σ) [3]. A similar CMS search observed (expected) a significance of 3.0σ (2.5σ) [8]. The expected significance is determined from a fit to simulated Asimov datasets assuming the Standard Model signal strength, representing the sensitivity that would be expected in the absence of statistical fluctuations, while the observed significance reflects the actual result obtained from data. Building on these previous searches and establishing *evidence* (3.0σ significance) for the Higgs boson decaying to two muons with the ATLAS detector is the primary goal of this thesis. This is done by combining data collected from 2015–2018 with new data collected by the ATLAS detector from 2022–2024.

Since the Higgs boson's coupling to a particle depends on the particle's mass, its decay to a muon–antimuon pair is extremely rare at the Large Hadron Collider. To enhance the sensitivity to this process, advanced machine learning techniques are employed to divide the dataset into mutually exclusive categories. A fit to the invariant mass of the muon–antimuon system is then performed to search for a peak at the Higgs boson mass. Neural networks are one such technique that can effectively distinguish Higgs signal events from background, but their use has traditionally been limited by unwanted correlations with the dimuon invariant mass, which can distort the signal extraction. This thesis presents a novel application of neural networks trained with a

distance correlation (DisCo) loss term, which mitigates these correlations and enables neural networks to be used effectively in the search for the Higgs boson decaying to two muons.

Due to the rarity of the Higgs boson decay to a muon-antimuon pair, additional data will need to be collected by the ATLAS detector to observe (5σ significance) this process. Starting in 2026, the Large Hadron Collider will undergo major upgrades to enter the High-Luminosity phase, increasing the intensity of collisions and providing improved access to rare decay processes. To accommodate these extreme conditions, the ATLAS detector is also being upgraded, including a state-of-the-art silicon Inner Tracker. This thesis presents the characterization and on-wafer testing of the ABCStar readout chips for the Inner Tracker, which was achieved in collaboration with the Ontario-based industrial partner and a leader in ASIC quality control, DA-Integrated.

An overview of the Standard Model of particle physics, the Higgs boson, and proton-proton collisions is presented in Chapter 2. Chapter 3 presents the Large Hadron Collider and the ATLAS detector, which are the machines used to produce the Higgs bosons and detect the muons used for this analysis. This chapter also introduces future upgrades for the ATLAS detector, including the new Inner Tracker. Chapter 4 gives an overview of the ABCStar readout chip for the Inner Tracker upgrade, presents the results from on-wafer testing of the chips, and describes the problems encountered during testing. Reconstruction of objects in the ATLAS detector and an overview of the samples used in this analysis is given in Chapter 5. Chapter 6 outlines the analysis strategy used to search for the Higgs boson decaying to a muon-antimuon pair, and presents the novel approach of using neural networks with a DisCo loss term. The results from this approach are presented in Chapter 7 and compared with results from a recently published ATLAS paper on this Higgs decay process. Finally, Chapter 8

presents the concluding remarks and the outlook for future analyses.

Chapter 2

Theoretical Background

Our current understanding of the fundamental particles of the universe and the interactions that govern them is described by the Standard Model of particle physics (SM). This chapter gives an overview of the SM in Section 2.1 and introduces the Higgs boson, how it is produced in proton-proton collisions, and the Higgs decay to two muons in Section 2.2. Section 2.3 describes how proton-proton collisions are modelled and finally, Section 2.4 introduces the important background processes for this analysis.

The description of the SM in Section 2.1 largely draws on information from Ref. [9–11] unless otherwise stated.

2.1 The Standard Model

The Standard Model of particle physics is a description of the elementary particles in our universe and the fundamental forces that govern them. The SM includes three

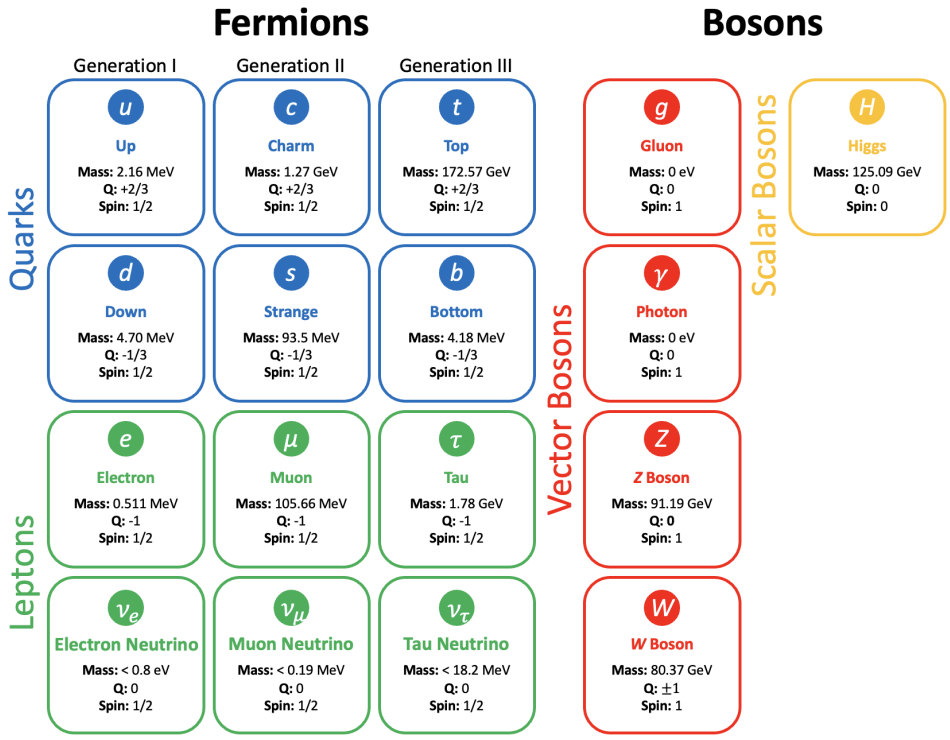


Figure 2.1: A summary of the Standard Model particle content. Particle masses are obtained from [12] and the Higgs mass is obtained from [13].

of the four fundamental forces: the electromagnetic force, the strong force and the weak force. The gravitational force is not included in the SM, but its affects are negligible within the realm of particle physics. The SM has been thoroughly tested experimentally and can accurately describe most natural phenomena.

The contents of the SM are shown in Figure 2.1. There are *fermions*, which are spin-1/2 particles that compose matter, and *bosons*, which are force-carrying particles. There are three generations of fermions, where each successive generation has a larger mass. It should be noted that although neutrinos have been determined to have very small masses experimentally [14–16], in the SM neutrinos are considered to be massless[1].

[1]The origin of neutrino masses is currently being investigated and the SM will need to be revised to include neutrino masses.

The SM Lagrangian is invariant under transformations of,

$$\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y. \tag{2.1}$$

These symmetries lead to conserved quantities within the SM according to Noether's theorem [17]. The $\mathrm{SU}(3)_C$ group describes the strong interaction, where C represents the colour charge. The colour charge can be red, blue or green and is conserved in all strong force interactions. The $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ group describes the electroweak interaction, where the conserved quantity is the weak hypercharge, $Y = 2(Q - I_3)$. Q is the electric charge, the conserved quantity of $\mathrm{U}(1)_{\mathrm{EM}}$ and I_3 is the third component of the weak isospin, the conserved quantity of $\mathrm{SU}(2)_L$. The L indicates that only fermions with left-handed chirality participate in weak interactions.

2.1.1 Quantum Electrodynamics

The Lagrangian for quantum electrodynamics (QED) is,

$$\mathcal{L}_{\mathrm{QED}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\psi}\not{D}\psi - m\bar{\psi}\psi, \tag{2.2}$$

where $F^{\mu\nu}F_{\mu\nu} = (\partial_\nu A_\mu - \partial_\mu A_\nu)(\partial^\nu A^\mu - \partial^\mu A^\nu)$. Here ψ represents a fermion of mass, m and the A_μ field is the photon. This Lagrangian has the interaction term,

$$\mathcal{L} = -e\bar{\psi}\gamma^\mu A_\mu \psi, \tag{2.3}$$

where e is the coupling constant of QED. The vertex arising from this interaction term is shown in Figure 2.2.

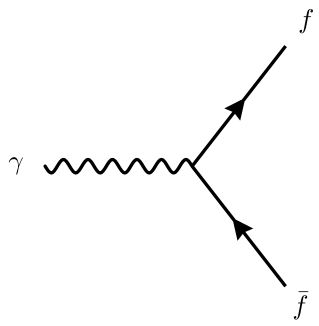


Figure 2.2: A representative QED vertex.

2.1.2 Quantum Chromodynamics

The Lagrangian for quantum chromodynamics (QCD) is,

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G^{a\mu\nu} G^{a}_{\mu\nu} + \bar{\psi}\left(i\not{D} - m\right)\psi, \tag{2.4}$$

where $G^{a}_{\mu\nu} = \partial_\mu G^{a}_{\nu} - \partial_\nu G^{a}_{\mu} - g_s f_{abc} G^{b}_{\mu} G^{c}_{\nu}$ is the field strength tensor. Here, ψ is a quark colour triplet and g_s is the strong coupling constant. The fundamental vertices of QCD that arise from this Lagrangian are shown in Figure 2.3.

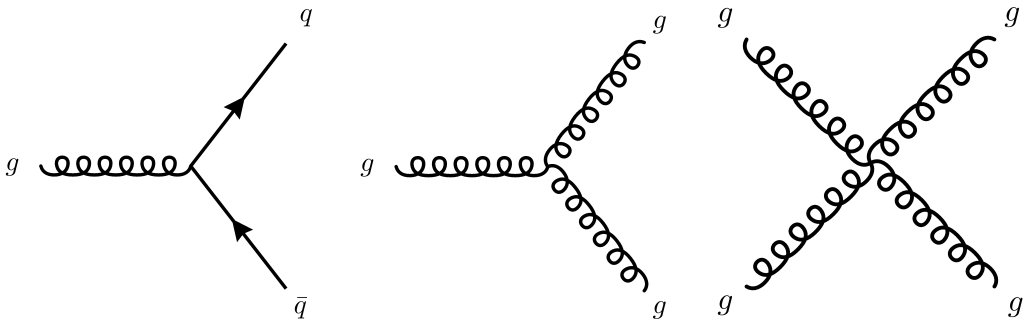


Figure 2.3: Vertices representing the three types of fundamental QCD interactions.

Quarks do not remain isolated and when they are created through interactions in the detector they will *hadronize* into hadrons composed of multiple quarks (see Section 2.3).

The spray of particles that arises when a quark hadronizes in a detector is referred to as a *jet*.

2.1.3 Electroweak Interaction

The electroweak (EW) interaction unifies the electromagnetic and weak forces in one theory [18–20] and is necessary to describe the weak interactions at high energies. The EW interaction is symmetric under $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ transformations, but this symmetry is spontaneously broken via the Higgs mechanism which is discussed in Section 2.2.

The weak interaction is mediated by the $W^{\pm}$ and Z vector bosons. When a W boson interactions with a fermion it changes the particle's species or flavour and charge. Interactions with the Z boson do not change a particle's species, similar to interactions with the photon. The vertices for the EW interactions are shown in Figure 2.4.

2.2 The Higgs Boson

The SM is invariant under transformations of $\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ when combining the strong and EW symmetry groups. This requires that all fermions and bosons in the SM are massless, which contradicts experimental evidence. This is resolved through the so called *Higgs mechanism* [4, 5], where the masses of fermions and bosons are acquired through spontaneous symmetry breaking of $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$. This mechanism predicts a massive boson called the *Higgs boson*, whose interactions are investigated in this thesis.

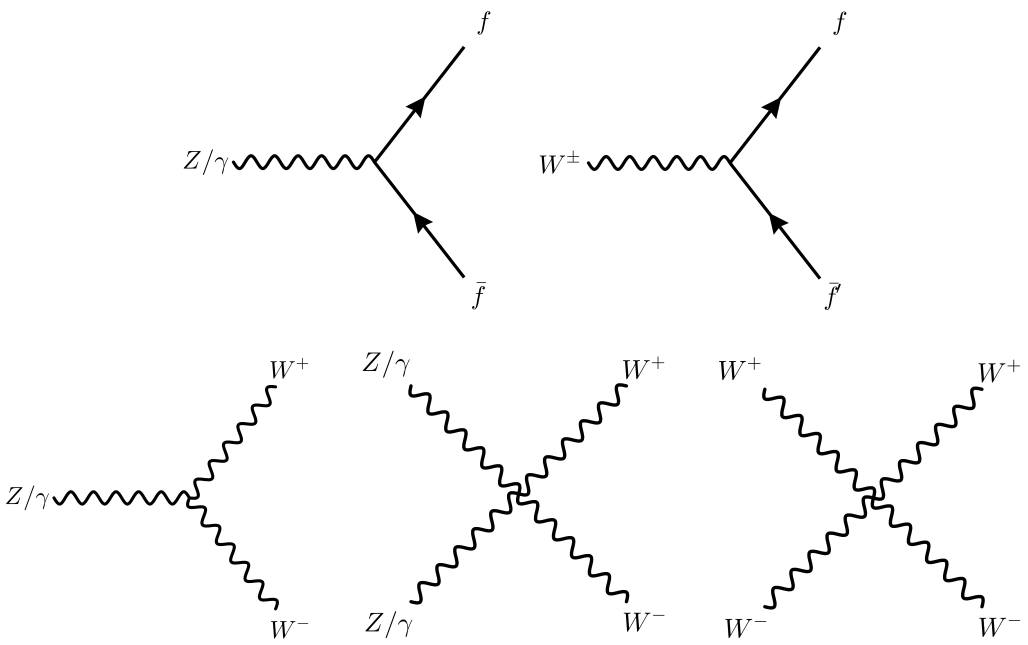


Figure 2.4: Vertices representing the fundamental EW interactions.

2.2.1 Electroweak Symmetry Breaking

Bosons and fermions in the SM acquire mass through a process called spontaneous symmetry breaking. This is introduced through the Higgs field, which consists of two complex scalar fields,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.5}$$

and is symmetric under $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$. The Lagrangian for this Higgs doublet is

$$\mathcal{L} = |D_\mu \phi|^2 + V(\phi) - \bar{\psi}_R \mathcal{M} \psi_L - \bar{\psi}_L \mathcal{M}^\dagger \psi_R, \tag{2.6}$$

where D_μ is the covariant derivative, $V(\phi)$ is the Higgs potential and the final two terms represent the interaction between the Higgs boson and fermions, where $\mathcal{M}$

depends on the type of fermion. The Higgs potential is given by,

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2, \tag{2.7}$$

where we require that $\lambda > 0$ otherwise the potential goes to $-\infty$ for large values of $\phi^\dagger\phi$. If $\mu^2 < 0$ then the Higgs potential has degenerate non-zero minima at

$$\phi = \pm v = \pm\sqrt{\frac{-\mu^2}{\lambda}}, \tag{2.8}$$

as seen in Figure 2.5. This results in symmetry breaking of the Lagrangian, because the potential has an infinite set of minima defined by

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} = v^2, \tag{2.9}$$

where $v = \sqrt{-\mu^2/\lambda}$, is the *vacuum expectation value* (vev).

Without loss of generality, Equation 2.5 can be simplified and defined in terms of a physical Higgs field, H,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \tag{2.10}$$

Using this, the Higgs potential (Equation 2.7) becomes,

$$V(\phi) = \lambda v^2 H^2 + \lambda v H^3 + \frac{\lambda}{4} H^4. \tag{2.11}$$

If we include an extra term from the covariant derivative in Equation 2.6, $\frac{1}{2}(\partial_\mu H)^2$,

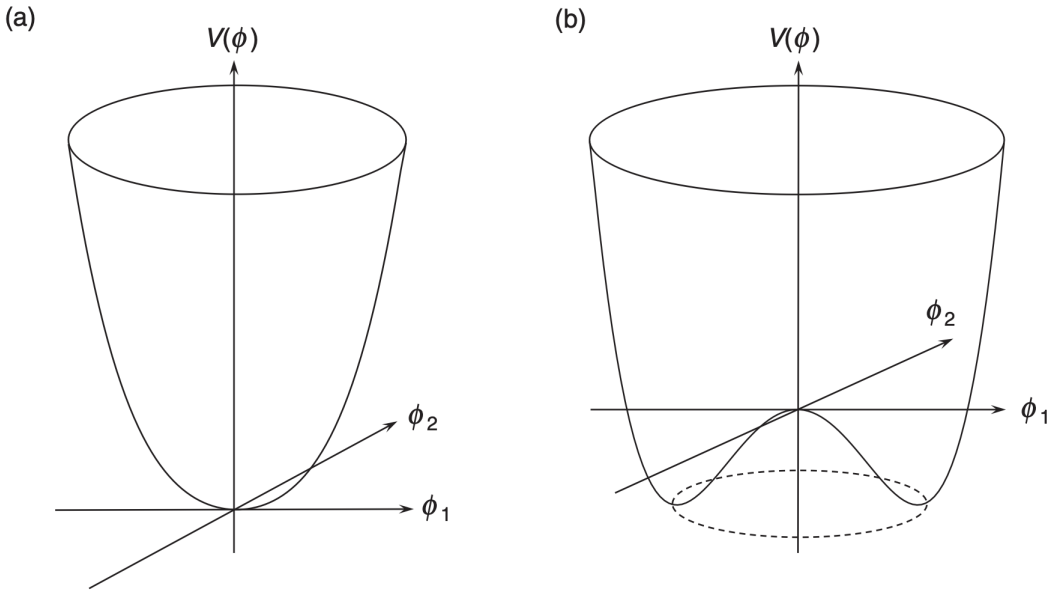


Figure 2.5: The Higgs potential, $V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$ for (a) $\mu^2 > 0$ and (b) $\mu^2 < 0$. Taken from [9].

then the new terms in the SM Lagrangian only containing the H field are,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu H)^2 + \lambda v^2 H^2 + \lambda v H^3 + \frac{\lambda}{4}H^4. \tag{2.12}$$

This represents the self-interacting terms from the SM for the Higgs boson. The term $\frac{1}{2}(\partial_\mu H)^2$ is the kinetic term, $\lambda v^2 H^2$ is the Higgs mass term and the final two terms, $\lambda v H^3$ and $\frac{\lambda}{4}H^4$, are Higgs self-interaction terms. This results in a Higgs mass of,

$$m_H = \sqrt{2\lambda v^2}. \tag{2.13}$$

The Higgs self interaction terms are shown in Figure 2.6.

The covariant derivative term in Equation 2.6 shows how the Higgs boson couples to

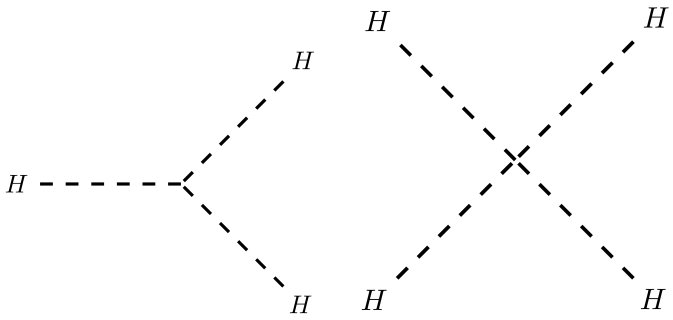


Figure 2.6: Higgs boson self interaction terms.

the EW sector. The covariant derivative acting on the Higgs doublet has the form,

$$D_\mu \phi = \left[\partial_\mu + \frac{1}{2} \left(i g_{\mathrm{W}} \boldsymbol{\sigma} \cdot \mathbf{W}_\mu + i g' B_\mu \right) \right] \phi, \tag{2.14}$$

where $\boldsymbol{\sigma}$ are the Pauli matrices. $\mathbf{W}_\mu$ are massless fields associated with $\mathrm{SU}(2)_L$ symmetry with coupling constant g_{W}, and B_μ is the massless field associated with $\mathrm{U}(1)_Y$ symmetry with coupling constant g'. If we expand the covariant derivative for the Higgs field we get

$$\begin{aligned} |D_\mu \phi|^2 =& \frac{1}{2} (\partial_\mu H)^2 + \frac{1}{8} v^2 g_{\mathrm{W}}^2 \left(W_\mu^{(1)} W^{(1)\mu} + W_\mu^{(2)} W^{(2)\mu} \right) \\ &+ \frac{1}{8} v^2 \left(g_{\mathrm{W}} W_\mu^{(3)} - g' B_\mu \right) \left(g_{\mathrm{W}} W^{(3)\mu} - g' B^\mu \right). \end{aligned} \tag{2.15}$$

If we introduce four new fields

$$W_\mu^{\pm} = \frac{1}{\sqrt{2}} \left(W_\mu^{(1)} \mp i W_\mu^{(2)} \right), \tag{2.16}$$

$$Z_\mu = \cos(\theta_W) W_\mu^{(3)} - \sin(\theta_W) B_\mu, \tag{2.17}$$

$$A_\mu = \sin(\theta_W) W_\mu^{(3)} + \cos(\theta_W) B_\mu, \tag{2.18}$$

and define

$$\tan\theta_W = \frac{g'}{g_W}, \tag{2.19}$$

then we get several important terms in the SM Lagrangian from the covariant derivative,

$$\mathcal{L} \supset \frac{g_{\mathrm{W}}^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{g_{\mathrm{W}}^2 v^2}{4\cos^2(\theta_{\mathrm{W}})} Z_\mu Z^\mu + 0 \cdot A_\mu A^\mu. \tag{2.20}$$

These represent the mass terms for the photon and the W and Z weak vector bosons,

$$m_{W^\pm} = \frac{1}{2} g_{\mathrm{W}} v, \tag{2.21}$$

$$m_Z = \frac{g_{\mathrm{W}} v}{2\cos(\theta_{\mathrm{W}})}, \tag{2.22}$$

$$m_A = 0. \tag{2.23}$$

This shows that the W and Z bosons acquire mass through the introduction of this new Higgs field, but the A field representative of the electromagnetic interaction remains massless, which agrees with experimental data. This also predicts that

$$\frac{m_W}{m_Z} = \cos\theta_{\mathrm{W}}, \tag{2.24}$$

which has been experimentally confirmed. Using measured values for m_W and g_{W} we find that the vev, $v = 246\,\mathrm{GeV}$. This still leaves a free parameter, λ, in Equation 2.13 for the Higgs mass. The latest combined measurement of the Higgs mass with the ATLAS and CMS collaborations is $m_H = 125.09\,\mathrm{GeV}$ [13].

Expanding the covariant derivative also yields several interaction terms between the

Higgs and W and Z bosons,

$$\mathcal{L} \supset \frac{v g_{\mathrm{W}}^2}{2} W_\mu^+ W^{-\mu} H + \frac{v g_{\mathrm{W}}^2}{8\cos^2(\theta_{\mathrm{W}})} Z_\mu Z^\mu H + \frac{g_{\mathrm{W}}^2}{4} W_\mu^+ W^{-\mu} H^2 + \frac{g_{\mathrm{W}}^2}{8\cos^2(\theta_{\mathrm{W}})} Z_\mu Z^\mu H^2. \tag{2.25}$$

These are the *gauge couplings*, which give the following interaction terms (Figure 2.7).

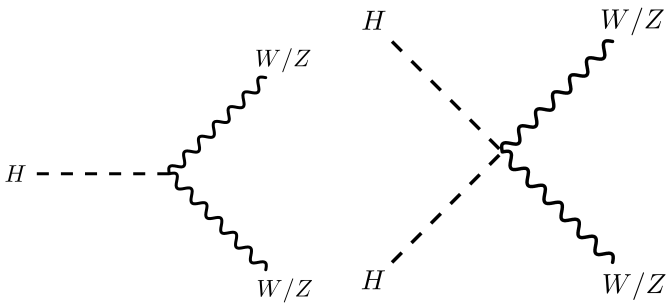


Figure 2.7: Interaction terms between the Higgs boson and W and Z bosons.

The Higgs mechanism is also used to generate mass terms for fermions. A term involving the Higgs doublet and the left-handed doublet ψ_L and right-handed singlet ψ_R will be gauge invariant under $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$. For example, for the muon and muon neutrino doublet, μ, there is a term in the Lagrangian of the form,

$$\mathcal{L}_{\text{Yukawa}} \supset -\frac{g_\mu}{\sqrt{2}} v \left(\bar{\mu}_L \mu_R + \bar{e\mu}_R \mu_L\right) - \frac{g_\mu}{\sqrt{2}} H \left(\bar{\mu}_L \mu_R + \bar{\mu}_R \mu_L\right). \tag{2.26}$$

Here, g_μ is the *Yukawa coupling* of the muon. If we choose the Yukawa coupling to be consistent with the observed muon mass, $g_\mu = \sqrt{2} m_\mu / v$, then Equation 2.26 becomes

$$\mathcal{L}_{\text{Yukawa}} \supset -m_\mu \bar{\mu}\mu - \frac{m_\mu}{v} \bar{\mu}\mu H. \tag{2.27}$$

Here we see that the first term gives the mass of the muon and represents the coupling of the muon to the non-zero vev of the Higgs field. The second term represents a

coupling between the muon and the Higgs boson. The mass terms and couplings for other leptons (electrons and taus) have the same form as Equation 2.27.

The non-zero vev occurs in the lower component of the Higgs doublet (Equation 2.10), so the combination with the left-handed doublet ψ_L and right-handed singlet ψ_R only generates masses for the charged leptons in the lower component of an $\mathrm{SU}(2)_L$ doublet, not the neutrino. In the minimal SM neutrinos are taken to be massless, since there are no right-handed neutrino fields[2].

The Higgs mechanism can also be used to generate the masses for quarks. The down-type quarks follow the same logic as charged leptons, but since the non-zero vev occurs in the lower component of the Higgs doublet the up-type quarks have the same problem as neutrinos. To fix this problem, the conjugate doublet, ϕ_c of Equation 2.5 is formed,

$$\phi_c = i\sigma_2\phi^* = \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -\phi_3 + i\phi_4 \\ \phi_1 - i\phi_2 \end{pmatrix}. \tag{2.28}$$

A gauge invariant mass term for the up-type quarks can then be constructed from $\bar{\psi}_L\phi_c\psi_R + \bar{\psi}_R\phi_c^\dagger\psi_L$,

$$\mathcal{L} = -\frac{g_\mathrm{u}}{\sqrt{2}}v\left(\bar{\mathrm{u}}_L\mathrm{u}_R + \bar{\mathrm{u}}_R\mathrm{u}_L\right) - \frac{g_\mathrm{u}}{\sqrt{2}}H\left(\bar{\mathrm{u}}_L\mathrm{u}_R + \bar{\mathrm{u}}_R\mathrm{u}_L\right), \tag{2.29}$$

where a Yukawa coupling of $g_\mathrm{u} = \sqrt{2}m_\mathrm{u}/v$ gives

$$\mathcal{L} = -m_\mathrm{u}\bar{\mathrm{u}}\mathrm{u} - \frac{m_\mathrm{u}}{v}\bar{\mathrm{u}}\mathrm{u}h. \tag{2.30}$$

[2]The SM can be extended to include massive neutrinos by adding right-handed neutrinos and allowing Dirac or Majorana mass terms, which is required to account for neutrino oscillations.

For all Dirac fermions, a gauge invariant mass term can be constructed from

$$\mathcal{L} = -g_{\mathrm{f}} \left[\bar{\psi}_L \phi \psi_R + \left(\bar{\psi}_L \phi \psi_R \right)^\dagger \right], \tag{2.31}$$

or

$$\mathcal{L} = g_{\mathrm{f}} \left[\bar{\psi}_L \phi_c \psi_R + \left(\bar{\psi}_L \phi_c \psi_R \right)^\dagger \right]. \tag{2.32}$$

The Yukawa couplings of the fermions to the Higgs field are given by

$$g_{\mathrm{f}} = \sqrt{2} \frac{m_{\mathrm{f}}}{v}. \tag{2.33}$$

The interaction terms between fermions and the Higgs boson are shown in Figure 2.8

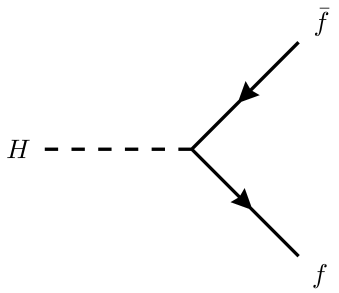


Figure 2.8: Feynman diagram showing the interaction between the Higgs boson and fermions.

2.2.2 Higgs Boson Production Modes

There are several different production modes for the Higgs boson. This thesis focuses on Higgs bosons that are created at the LHC, which occurs through the interaction of gluons or quarks from the pp collisions. The production modes that are relevant to this analysis are gluon fusion (ggF), vector boson fusion (VBF), Higgs-strahlung (VH, Higgs production in association with a vector boson), production of the Higgs boson

Table 2.1: Production cross sections for the Higgs boson with mass, $m_H = 125.09\,\text{GeV}$ at $\sqrt{s} = 13.6\,\text{TeV}$. Values obtained from [21].

Production Mode	Cross Section [pb]	Probability [%]
ggF	52.23	87.18
VBF	4.08	6.81
VH	2.40	4.01
$t\bar{t}H$	0.57	0.95
$b\bar{b}H$	0.53	0.88
tH	0.10	0.17
Total	59.91	100.00

together with a top quark pair ($t\bar{t}H$), production of the Higgs boson together with a bottom quark pair ($b\bar{b}H$) and top quark associated Higgs production (tH). The cross section for each Higgs production mode can be seen in Table 2.1.

2.2.2.1 Gluon Fusion

The most dominant Higgs boson production mode at the LHC is ggF. This involves gluons fusing via a quark loop as seen in Figure 2.9. The Higgs boson has a large coupling with the top quark due to the large top quark mass, so the largest contribution is from a top quark loop. Other quarks, such as the bottom quark, contribute to a much smaller fraction of this production mode, scaling like $(m_b/m_t)^2$ compared to the top quark loop.

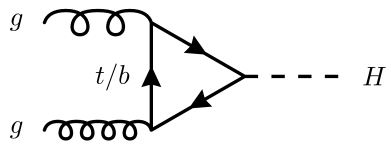


Figure 2.9: The leading order Feynman diagram for gluon fusion (ggF) Higgs production.

At leading order (Figure 2.9), ggF will produce a single Higgs boson with no associated particles or transverse momentum. However, there can be initial-state QCD radiation from the gluons that will frequently result in jets. In general, ggF tends to produce a Higgs boson with low transverse momentum and a few high-momentum jets.

2.2.2.2 Vector Boson Fusion

The second largest Higgs production mode is VBF. Two incoming quarks radiate a pair of vector bosons which fuse to create a Higgs boson as seen in Figure 2.10. VBF produces a distinct signal due to the presence of two quarks that will hadronize into two forward jets in either end of the ATLAS detector. There are rarely additional jets in VBF events as there is no colour charge flow between the two incoming partons. The Higgs boson tends to be in the central region of the detector.

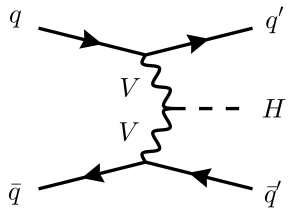


Figure 2.10: A representative Feynman diagram for vector boson fusion (VBF) Higgs production.

2.2.2.3 Higgs-Strahlung (VH)

Higgs-Strahlung production involves quarks interacting to produce a vector boson (W or Z boson), which radiates a Higgs boson (hence the name Higgs-Strahlung) as seen in Figure 2.11. VH events create a unique signal based on the decay of the associated vector boson. The vector bosons can decay into a pair of quarks or leptons.

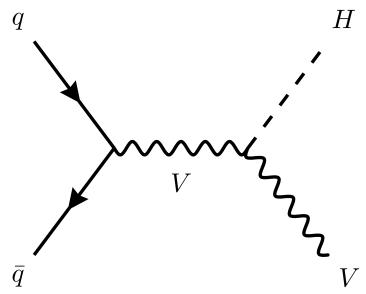


Figure 2.11: A representative Feynman diagram for Higgs-Strahlung (VH) Higgs production. The vector boson can be a W or Z boson.

If any of these leptons are a neutrino, they will escape the detector leading to missing transverse energy. The Higgs boson will have a larger transverse momentum as it will recoil against the vector boson.

2.2.2.4 Higgs Production in Association with Top Quark Pair

$t\bar{t}H$ production occurs when gluons create a pair of top quarks, which radiate a Higgs boson as seen in Figure 2.12a. The Higgs boson often has high p_{T} due to the top quarks recoiling against it. The two top quarks will further decay into a W boson and bottom quark. These two bottom quarks will hadronize into jets. Each W boson can decay hadronically or leptonically. As a result, these top quark decays can produce a final state with zero, one or two additional charged leptons, possible missing transverse energy due to neutrinos, and additional jets.

2.2.2.5 Higgs Production in Association with Bottom Quark Pair

$b\bar{b}H$ production is similar to $t\bar{t}H$ but a pair of bottom quarks are created in association with the Higgs boson as seen in Figure 2.12b. The resulting final state will have two soft jets originating from the bottom quarks. $b\bar{b}H$ has a similar signature to ggF, with

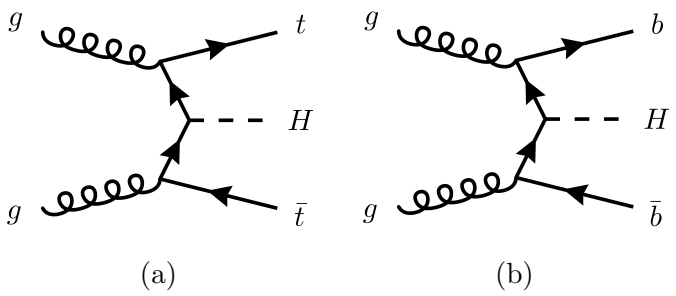


Figure 2.12: A representative Feynman diagram for production of the Higgs boson together with (a) a top quark pair ($t\bar{t}H$) and (b) a bottom quark pair ($b\bar{b}H$).

a much smaller cross section.

2.2.2.6 Higgs Production in Association with a Top Quark

The rarest Higgs production mechanism relevant for this analysis is tH production. The Higgs boson and top quark are produced along with either a quark and bottom quark or with a W boson and bottom quark as seen in Figure 2.13. As with $t\bar{t}H$, the top quark will decay into a W boson and bottom quark. The W boson will further decay into leptons or jets. The resulting final states will have multiple jets (some of which originate from bottom quarks) and possibly leptons from the W boson decay.

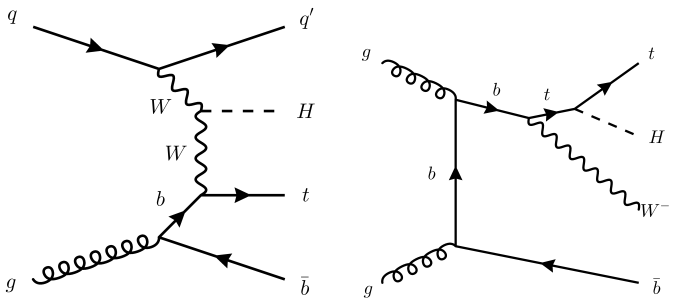


Figure 2.13: Representative Feynman diagrams for production of the Higgs boson together with a top quark (tH).

2.2.3 Higgs Boson Decay Channels

The Higgs boson is an unstable particle and will decay promptly in $\sim 10^{-22}$ s. The Higgs boson can decay via multiple different channels with a probability that is determined by their *branching ratios* (BR). The BR for Higgs decays is predicted by the SM and proportional to the Higgs mass and the coupling to the decay particles. The Higgs coupling is proportional to the decay particles mass (Section 2.2.1), so the branching ratio is larger for more massive particles. This is constrained by conservation of energy, and the Higgs boson cannot decay to a top quark pair as the top quark mass is larger than the Higgs mass. In decays to vector bosons, such as $H \to ZZ^*$, one of the bosons is produced off-shell (denoted Z^*) as the combined mass of the decay products exceeds that of the Higgs boson.

ATLAS measurements of the Higgs BR for the most important decay channels are shown together with SM prediction in Figure 2.14. Decays to massless photons are possible via coupling through a top quark loop. All of the processes in Figure 2.14 have been observed with 5σ significance except for $H \to Z\gamma$ and $H \to \mu\mu$ [22].

2.2.3.1 The $H \to \mu\mu$ Channel

This thesis searches for the Higgs boson decaying into two muons, $H \to \mu\mu$. The most dominant $H \to \mu\mu$ process at the LHC is a Higgs boson created through ggF which promptly decays to two muons as seen in Figure 2.15.

Measurement of the $H \to \mu\mu$ decay is highly anticipated because it would be the first measurement of a Higgs coupling to a second generation fermion. The Yukawa couplings of third generation fermions have been established, but couplings to second

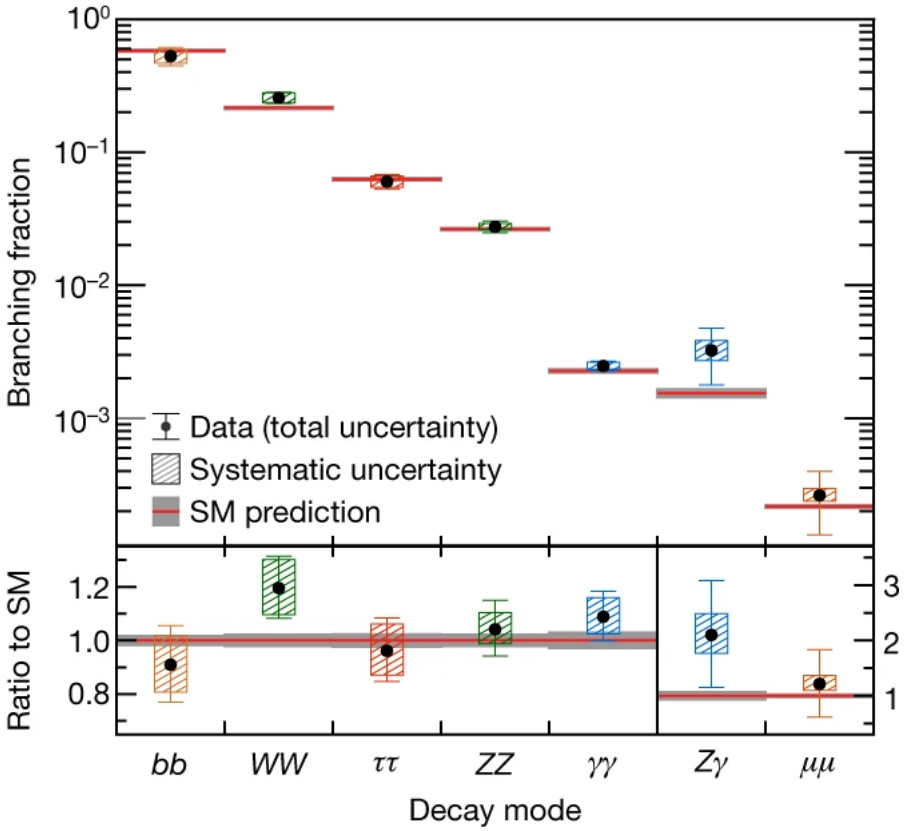


Figure 2.14: Branching ratios for different Higgs boson decay modes measured by the ATLAS detector assuming SM values for the production cross sections. Figure taken from Ref. [22].

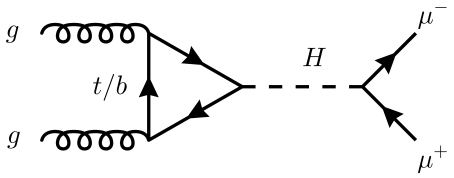


Figure 2.15: A representative Feynman diagram for ggF production of a Higgs boson decaying to two muons.

generation fermions have not yet been measured with 5σ significance. The Higgs boson interactions are the only ones in the SM that distinguish between fermion generations and the $H \to \mu\mu$ process is the most sensitive channel to measure a Higgs coupling to a second generation fermion at the LHC.

Since the Higgs couplings are proportional to the fermion masses, the coupling to the muon is much smaller than the couplings that have been measured experimentally so far. The BR for the Higgs to two muons is dependent on the mass of the Higgs boson. The predicted BR of $H \to \mu\mu$ is $\mathcal{B}(H \to \mu\mu) = (2.17 \pm 0.05) \times 10^{-4}$ [23], assuming a mass of $m_H = 125.09\,\text{GeV}$.

2.3 Modelling *pp* Collisions

When protons collide at the LHC they can undergo *elastic scattering* or *inelastic scattering* [24]. Elastic scattering occurs roughly 30% of the time and the protons collide and deflect, but remain intact. This process is never seen in the ATLAS detector as the protons will scatter forward and miss all detector components. Inelastic scattering occurs about 70% of the time and at least one proton will break apart into quarks which will hadronize into jets. The ATLAS experiment is mainly interested in *non-diffractive* inelastic collisions, where the protons undergo a *hard scatter* and produce interesting particles for studies.

It is important to have theoretical predictions to compare with experimental measurements. Unfortunately, theoretical predictions of *pp* collisions are difficult to compute analytically and are predicted using sophisticated Monte Carlo (MC) simulations [25].

2.3.1 Cross Section Calculations and PDFs

The probability that a specific process will occur during a particle collision is referred to as the *cross section*, σ. The cross section is defined as the number of interactions per unit time per target particle over the incident flux. Predictions for a cross section can be calculated using the *matrix element*, $\mathcal{M}$. For example, for a two-body to two-body process, the cross section is given by

$$\sigma = \frac{1}{64\pi^2 s}\frac{|p_f|}{|p_i|}\int |\mathcal{M}|^2 \mathrm{d}\Omega. \tag{2.34}$$

Here $\sqrt{s}$ is the centre-of-mass energy, $|p_f|$ and $|p_i|$ is the magnitude of the momentum of the final and initial particles and $\mathrm{d}\Omega$ is the solid angle element. For a given process, X, if the integrated luminosity (a measure of the size of the dataset, see Section 3.2.7), $\mathcal{L}$, and the cross section, σ_X, are known, the expected number of events, ν_X, can be calculated with

$$\nu_X = \mathcal{L}\,\sigma_X. \tag{2.35}$$

In order to simulate a process in MC, the matrix element needs to be calculated. This is done by adding the contributions from all possible Feynman diagrams describing a process. If we consider the annihilation of a quark and anti-quark to produce a Z boson as shown in Figure 2.16, there will be Feynman diagrams of different orders. Leading order (LO) diagrams have the least amount of vertices, next-to-leading order (NLO) have diagrams with one additional vertex, next-to-next-to-leading order (NNLO) will have two additional vertices and this pattern continues for even higher order diagrams (e.g. $\mathrm{N^3LO}$ for next-to-next-to-next-to-leading order). In order to calculate the matrix element up to a certain order, all diagrams of that order must be included. Moving

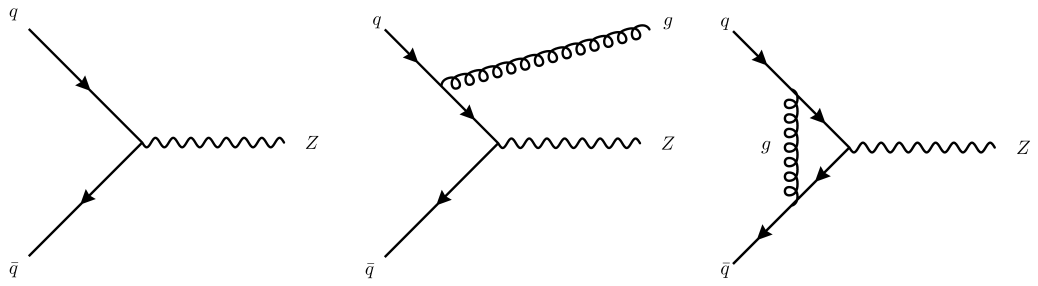


Figure 2.16: Sample Feynman diagrams for a quark and antiquark annihilating to produce a Z boson at (a) LO, (b) NLO and (c) NNLO.

to higher orders involves more vertices, leading to more complex diagrams and a higher number of possible diagrams, which makes calculations more computationally expensive. For this reason, the matrix elements in MC simulations are only calculated based on diagrams up to a certain order, which can vary between samples. The higher the order, the better predictions, but the more computational power required.

When calculating the cross section for processes involved in pp collisions at the LHC, the structure of the protons needs to be taken into account. Hadrons, such as the proton, are composed of quarks and gluons referred to as *partons*. The proton is a bound state of two up quarks and one down quark (uud). These are known as the *valence quarks* of the proton and will interact with each other via gluons which can produce additional quark-antiquark pairs referred to as *sea quarks*. The total momentum of the proton is divided among these partons and this must be taken into account in simulations.

When the protons undergo inelastic scattering, the partons will scatter. To parametrize this, the squared momentum transferred in the collision, Q^2, and the fraction of the proton's momentum carried by the parton involved in the collision, x, are considered. From this, the probability distribution for the momentum of the parton involved in

the collision, referred to as the parton distribution function (PDF), $f(x, Q^2)$ can be constructed. The cross section for a pp collision where partons q_i and q_j go to a final state of products, X, is given by

$$\sigma(pp \to X) = \sum_{i,j} \int_0^1 \mathrm{d}x_i \mathrm{d}x_j f_i(x_i, Q^2) f_j(x_j, Q^2) \mathrm{d}\hat{\sigma}(q_i q_j \to X). \tag{2.36}$$

These PDFs give the probability to find a parton with a momentum fraction x, for a given momentum transfer Q^2. The PDFs are non-calculable, but are universal and can be measured experimentally and applied to different experiments. The evolution of PDFs in Q^2, however, is calculable with Altarelli-Parisi equations [26]. A sample PDF set is shown in Figure 2.17, which shows that most collisions at the LHC will be soft, but occasionally there will be hard collisions. It is also worth noting that in pp collisions there will often be an imbalance in the energy of the scattering partons, which will lead to an asymmetric scatter boosted to the left or right.

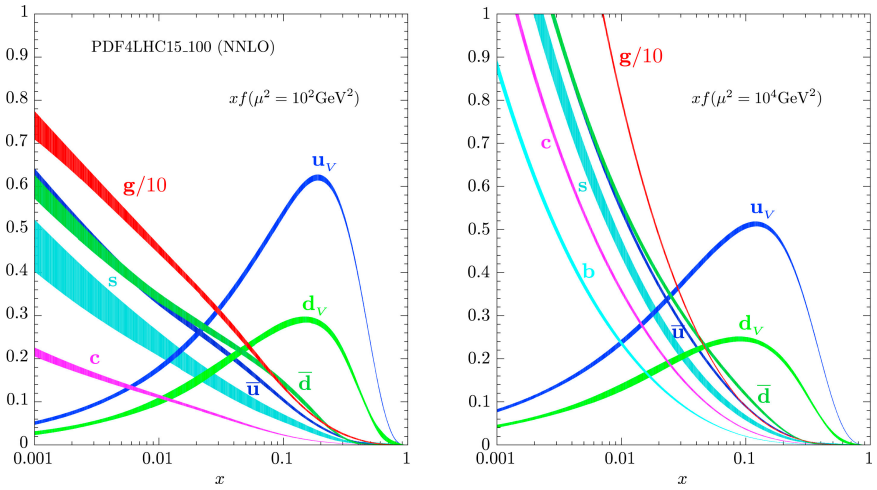


Figure 2.17: PDFs with uncertainty bands shown for $Q^2 = 10\,\mathrm{GeV}^2$ and $Q^2 = 10\,000\,\mathrm{GeV}^2$. Figure taken from Ref. [27].

2.3.2 Parton Showers

Calculating the outgoing partons from a collision is the first step of the simulation, but these partons will radiate additional quarks and gluons in a *parton shower* until their energy is low enough for hadronization to occur. These QCD parton showers are not calculable so they are approximated with simulations.

The parton showering algorithm works by taking the partons from the initial matrix element calculation then determining the probability that these partons will split into additional partons and what energy they will have using a splitting function. Multiple splitting sequences are generated with assigned probabilities so the final cross section of a parton shower is the sum of many different final states of quarks and gluons. Parton showering algorithms deal with soft and collinear parton emissions. For harder or wide angle emissions it is better to model this with further matrix element calculations. Hard splittings are referred to as *final state radiation* (FSR), for example a gluon radiating from a final state quark. QCD radiation occurring from partons before the hard scatter is referred to as *initial state radiation* (ISR). Multiparton interactions can also arise between additional partons from the protons involved in the hard scatter or from interactions with the beam remnants. The final parton showering modelling for a process is a combination of showering and hard splitting calculations, which can be modelled with difference procedures. The main schemes used are CKKW-L [28, 29] and FxFx [30].

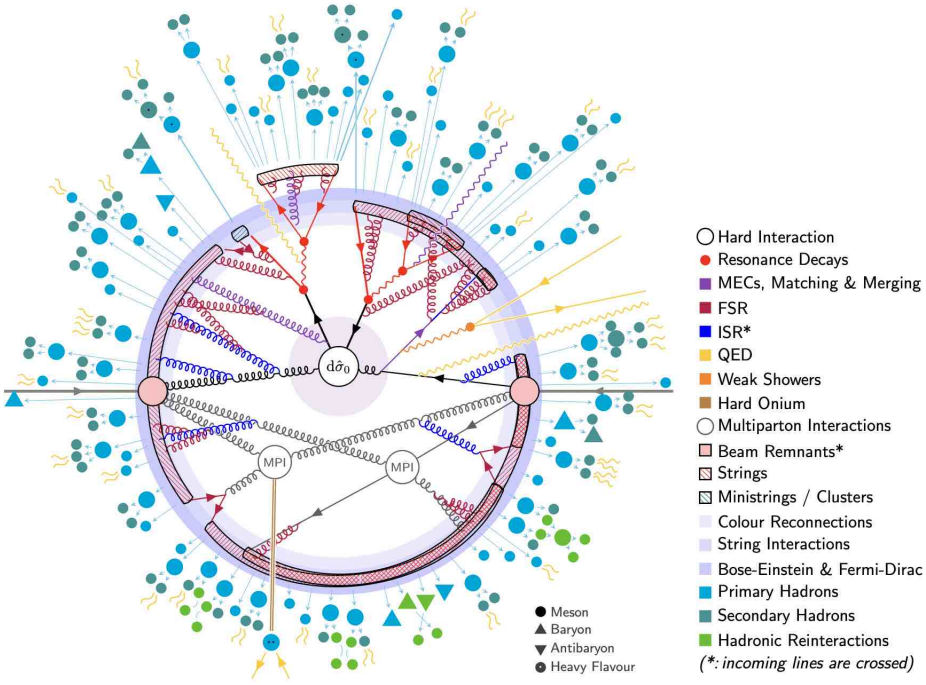


Figure 2.18: A schematic of a $pp \to \bar{t}t$ event showing the parts of an event simulated in an MC generator. The top and antitop quarks are represented by the Fermion lines exiting the initial hard interaction. The initial partons from the pp collision undergo scattering, parton showering and hadronization to produce the final state particles seen in a detector. Figure taken from Ref. [31].

2.3.3 Hadronization

Once the outgoing partons have a low enough energy ($\mathcal{O}$(GeV)), the partons will become confined to hadrons in a process called *hadronization*. Two common hadronization models used are the Lund string model [32, 33] and the cluster model [34]. These models both use input from experimental measurements to determine their parameters. If the hadrons produced from hadronization are unstable, the decays into more stable particles will also be simulated by the MC generator.

A representation of the processes arising from a pp collision that are modelled by the simulation is shown in Figure 2.18. The particles produced by hadronization in

simulation represent particles that are measurable by the detector. The output from a simulation for an event is a set of final state particles and their momentum.

2.4 Dimuon Background Processes in *pp* Colliders

There are a number of processes that can produce a muon-antimuon pair at pp colliders in addition to the $H \to \mu\mu$ process. All of these additional dimuon processes provide a large background to the $H \to \mu\mu$ signal that is being measured in this thesis. Due to the low Higgs production cross section and low BR to two muons, there is a very low signal to background ratio for $H \to \mu\mu$, predicted to be less than 1/500.

The dominant background is from Drell-Yan (DY) $Z/\gamma^* \to \mu\mu$ production. This occurs when a quark and antiquark annihilate to produce a virtual Z boson or photon and decay into a pair of charged leptons (muons for this analysis) as seen in Figure 2.19. Jets can also be produced in the DY process by partons arising from ISR. The DY dimuon invariant mass spectrum has a large resonance around the mass of the Z boson and a large smoothly falling background around the Higgs mass.

Additional important backgrounds include $t\bar{t}$, diboson processes, single-top processes and EW $Z \to \mu\mu$ production. Example Feynman diagrams for these processes can be seen in Figure 2.20. $t\bar{t}$ and single-top processes will produce bottom quarks in addition to a dimuon pair, making them important backgrounds for $t\bar{t}H$. Diboson backgrounds such as $VV \to \mu\mu\ell\ell$ and $VV \to \mu\mu\ell\nu$ are important backgrounds for leptonic VH decays as they will have the same lepton multiplicities. EW $Z \to \mu\mu$ produces two jets in the forward regions of the detector, which can mimic the signal from VBF $H \to \mu\mu$ decays.

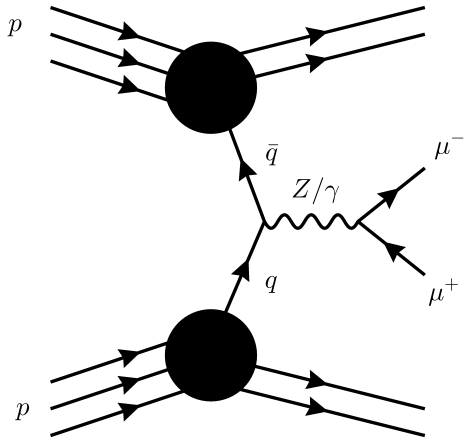


Figure 2.19: Feynman diagram for Drell-Yan $Z/\gamma^* \to \mu\mu$. Additional partons can be produced through initial state radiation, resulting in the production of jets.

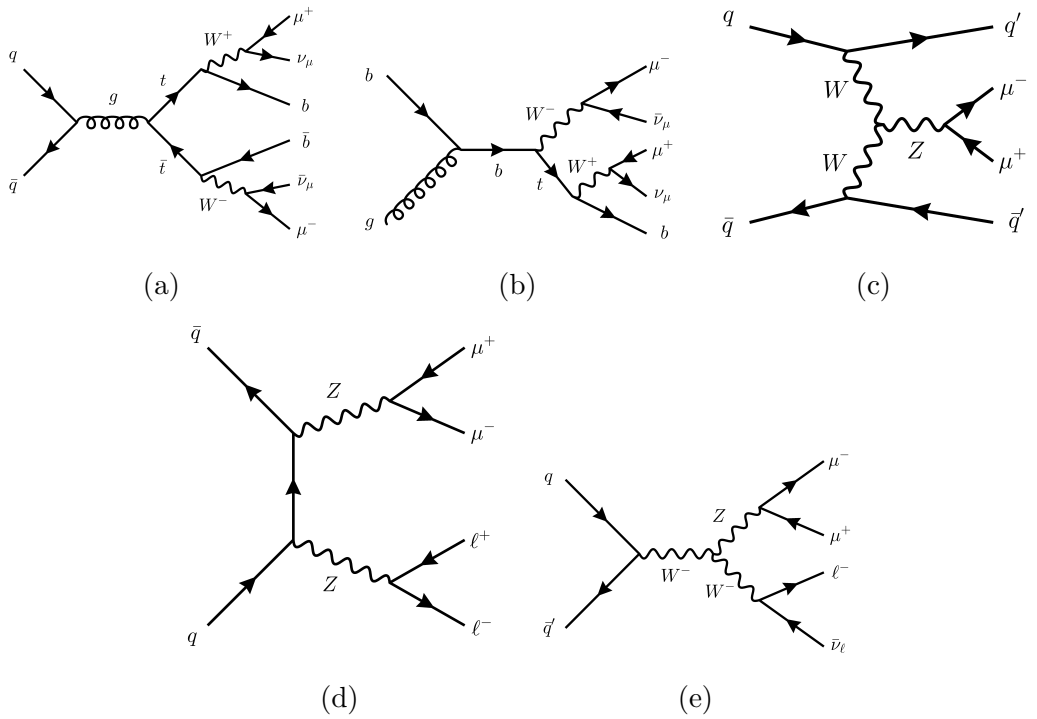


Figure 2.20: Sample Feynman diagrams of background processes for $H \to \mu\mu$: (a) $t\bar{t}$, (b) single-top, (c) EW $Z \to \mu\mu$, (d) $VV \to \mu\mu\ell\ell$ and (e) $VV \to \mu\mu\ell\nu$.

Chapter 3

The LHC and the ATLAS Detector

This chapter introduces the experimental infrastructure used to collect the particle physics data used in this thesis as well as their anticipated upgrades. Section 3.1 introduces the Large Hadron Collider, which is used to accelerate protons to high energies and collide them. Section 3.2 describes the ATLAS detector, a general purpose detector on the Large Hadron Collider used to collect data from the proton collisions. Future upgrades for the ATLAS detector during the next phase of the Large Hadron Collider with increased data taking are outlined in Section 3.3.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [35] is the world's largest particle collider. It is a circular collider that is located 100 m underground at the border of France and Switzerland with a circumference of approximately 27 km. The LHC, operated by the European Organization for Nuclear Research (CERN), is part of the CERN accelerator

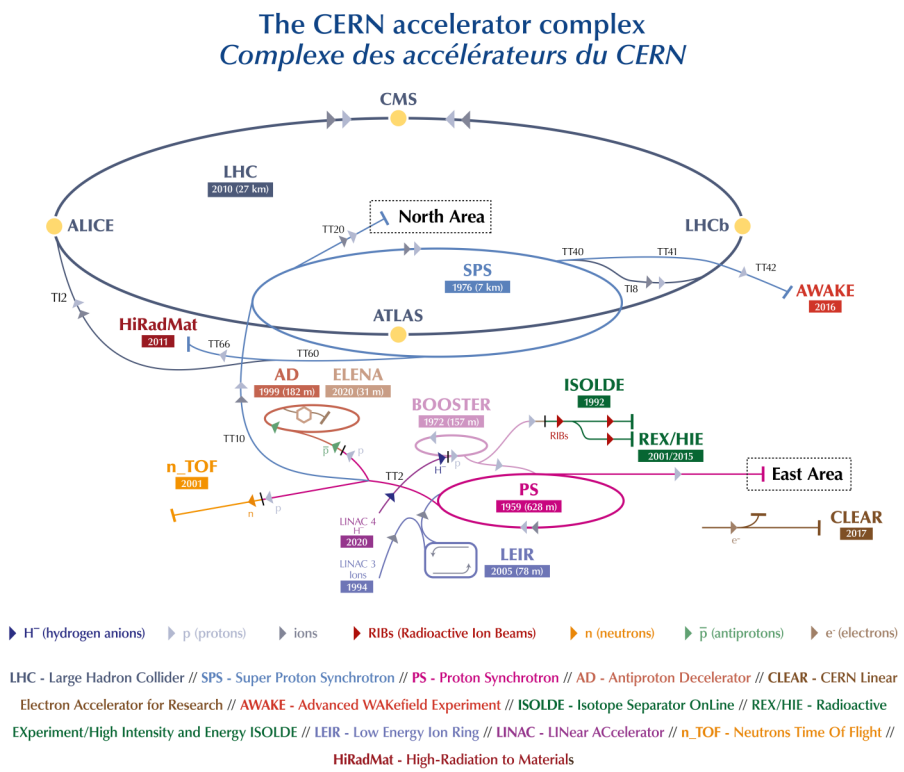


Figure 3.1: A schematic of the CERN accelerator complex in 2019. Protons are accelerated using LINAC4, followed by the PS booster, the PS, the SPS and then finally transferred into the LHC and circulated until they achieve the desired energy. The protons collide at different detectors along the ring, including the ATLAS detector. Figure taken from Ref. [36]

complex, as shown in Figure 3.1.

The primary function of the LHC is to accelerate bunches of protons up to high energies and then collide them at different detectors along the beam line to collect data for particle physics experiments. These protons are created by using an electric field to strip the electrons from hydrogen atoms [37]. The resulting protons are then accelerated using LINAC4, a linear accelerator, and injected into the Proton Synchrotron Booster (PSB) which accelerates them to an energy of 1.4 GeV. After the PSB, the beam of protons is fed into the Proton Synchrotron (PS) and accelerated to

25 GeV. These protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Protons are then fed from the SPS to the LHC into two separate beams travelling clockwise and anti-clockwise until they reach the desired energy.

Data-taking periods at the LHC are divided into *runs* that span for several years. An overview of past and future planned runs at the LHC can be seen in Figure 3.2. This analysis focuses on data collected during Run-3 of the LHC. During Run-3, protons beams were accelerated up to an energy of 6.8 TeV, to achieve a total centre-of-mass energy of $\sqrt{s} = 13.6\,\text{TeV}$.

The LHC beam line is kept at an ultrahigh vacuum pressure of 10^{-13} atm to prevent collisions with gas molecules [37]. The protons are accelerated and bent around the LHC using a series of superconducting dipole magnets, which produce a magnetic field of up to 8.3 T. In order to operate, the magnets have to be cooled down to a temperature of 1.9 K, using a superfluid helium cryogenic system.

When protons are injected into the LHC, they are injected in *bunches* with approximately 10^{11} protons per bunch [37]. These bunches collide at various interaction points around the LHC every 25 ns (frequency of 40 MHz). There are four main detectors at interaction points around the LHC: A Toroidal LHC Apparatus (ATLAS) [39], Compact Muon Solenoid (CMS) [40], A Large Ion Collider Experiment (ALICE) [41] and LHCb [42]. ATLAS and CMS are the two general purpose detectors, designed to collect information from the proton-proton collisions, while ALICE focuses on heavy ion collisions and LHCb focuses on the physics of b quarks.

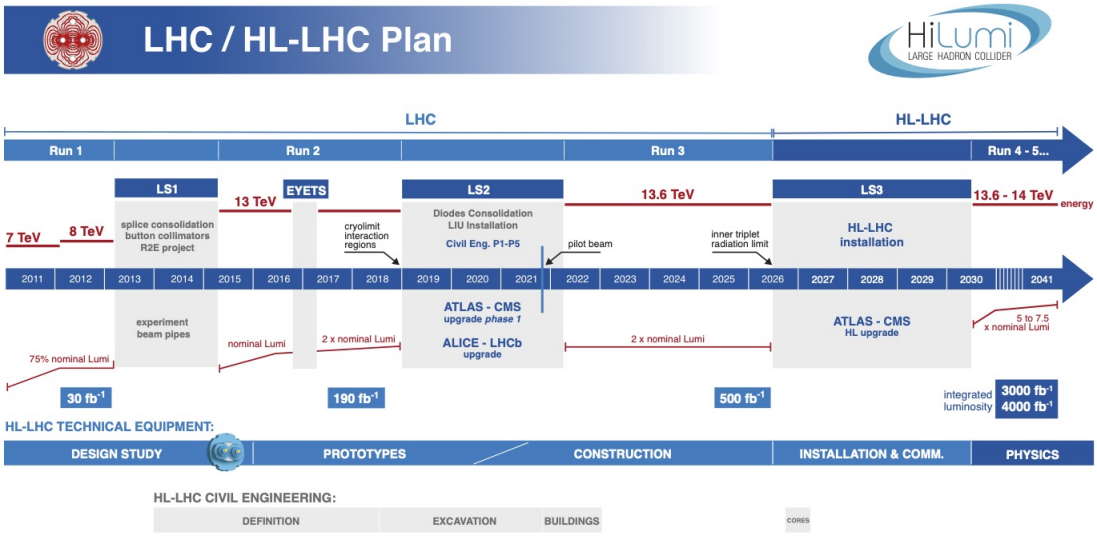


Figure 3.2: Overview of the data taking periods and breaks for maintenance of the LHC. After Run-3, the LHC will be shut down and undergo substantial upgrades for the HL-LHC phase. The analysis presented in this thesis uses data collected during Run-3 of the LHC at $\sqrt{s} = 13.6\,\text{TeV}$. Figure taken from Ref. [38].

3.1.1 The HL-LHC Upgrade

After Run-3 of the LHC, the accelerator will be shut down for Long Shutdown 3 (LS3) and undergo upgrades for the High-Luminosity phase of the LHC (HL-LHC) [43]. This upgrade aims to greatly increase the luminosity being delivered to provide increased statistics for rare physics events. The number of interactions per proton bunch crossing, μ, is going to increase to about $\mu = 200$ at the HL-LHC, approximately 5 times more than in Run-3. The HL-LHC will be able to deliver a peak luminosity $5 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ and an integrated luminosity of about $250\,\text{fb}^{-1}$ each year. The goal is to provide a total integrated luminosity of about $4000\,\text{fb}^{-1}$, which would be six times the expected integrated luminosity from Run-3.

3.2 The ATLAS Detector

The ATLAS detector is a 25 m tall, 44 m long cylindrical detector that weighs almost 7000 tonnes [44], making it the largest detector on the LHC. It is built around the beam line of the LHC, and provides nearly entire solid angle coverage around the interaction point near the center of the detector. ATLAS is composed of multiple subdetectors, which can be seen in Figure 3.3 and will be described in the following sections. Closest to the interaction point at the centre of the detector is the Inner Detector (ID), which provides tracking information for charged particles passing through. Outside of this are the electromagnetic and hadronic calorimeters that provide information on the energy of particles leaving the detector. On either end of the calorimeters is the New Small Wheel (NSW) muon chambers, designed to provide trigger and tracking for muons. The outermost layer of the ATLAS detector is the muon spectrometer, which can measure the momenta of muons.

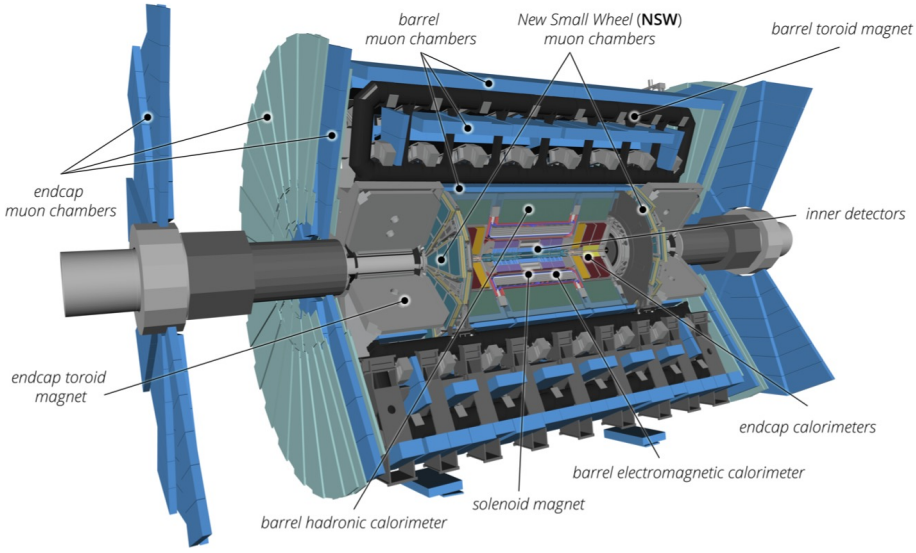


Figure 3.3: The ATLAS detector is cylindrical and composed of many subdetectors built around the LHC beam pipe. Figure taken from Ref. [44].

3.2.1 The ATLAS Coordinate System

In order to describe the physics inside of the ATLAS detector, a coordinate system needs to be established. The origin of this coordinate system is the centre of the detector, located at $x = y = z = r = 0$. The positive x-axis points towards the centre of the LHC and the positive y-axis points the sky. The positive z-axis points anti-clockwise along the beam line when the LHC is viewed from above. The polar angle θ is the angle from the positive z-axis and the azimuthal angle ϕ is the angle from the positive x-axis, with positive (negative) ϕ representing the top (bottom) half of the detector.

Rapidity, y, is commonly used as an angular quantity for particles in the ATLAS detector instead of θ. Rapidity is defined as,

$$y = \frac{1}{2}\left(\frac{E + p_z}{E - p_z}\right), \tag{3.1}$$

and is a useful quantity because differences in rapidity are invariant under *boosts* (a change in the reference frame of an observer due to relative motion along a specific direction) in the z-direction (along the beam line). For the remainder of this thesis, y will refer to rapidity unless otherwise stated. The angular distance between objects in ATLAS can be defined based on the rapidity and azimuthal angle,

$$\Delta R = \sqrt{\Delta y^2 + \Delta \phi^2}. \tag{3.2}$$

Another commonly used coordinate is the pseudorapidity,

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \tag{3.3}$$

The pseudorapidity is approximately equal to rapidity for a massive particle, and is equivalent to y if the particle is massless. Components of the ATLAS detector are commonly described by their location using η, with low $|\eta|$ being the central part of the detector and high $|\eta|$ being the forward part of the detector.

For particles in the ATLAS detector, we are usually interested in the transverse momentum, p_{T}, defined as

$$p_{\mathrm{T}} = \sqrt{p_{\mathrm{x}}^2 + p_{\mathrm{y}}^2}. \tag{3.4}$$

The transverse momentum is useful because the initial longitudinal momentum of particles travelling in the ATLAS detector is unknown, but the particles are always moving parallel to the beam line so the initial transverse momentum before collisions occur is always zero. An illustration of the coordinate system for a particle in the ATLAS detector is shown in Figure 3.4.

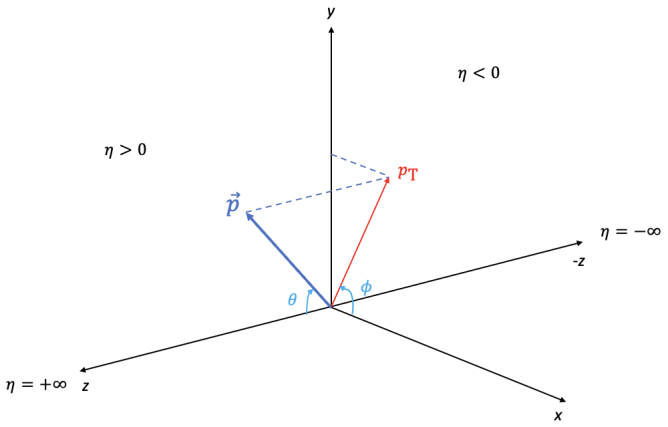


Figure 3.4: An illustration of the coordinate system used in the ATLAS detector.

3.2.2 The Interaction Point

The *pp* interaction point occurs near the centre of the ATLAS detector. At the interaction point, the two proton beams overlap and the bunches of protons circulating the LHC may collide. This region where proton collisions occur is referred to as the *beam spot*. The beam spot has a roughly gaussian distribution in the xy-plane with a width of $\sigma_{x,y} \simeq 20\,\text{mm}$ [45]. The centroid of the beam spot can vary, but is usually located within a few millimetres of the centre of the detector [45].

A point in space where a particle interaction occurs is referred to as a *vertex*. A primary vertex (PV) is a *pp* collision that occurs during a bunch crossing in the beam line. A secondary vertex can occur if longer lived particles are produced in a PV and travel some distance before decaying into secondary particles. Within the beam spot, multiple *pp* collisions can occur at each bunch crossing. Physics analyses are usually focused on a single hard-scatter collision during a bunch crossing. The additional *pp* collisions that occur within the same or neighbouring bunch crossings are referred to as *pileup*. Sample bunch crossings with different numbers of PVs are shown in Figure 3.5.

3.2.3 Inner Detector

The subdetector in ATLAS closest to the interaction point is the Inner Detector (ID). The ID is composed of four subsystems: the Insertable B-Layer (IBL), the Pixel Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT), which can be seen in Figure 3.6. The ID is 7024 mm long, has a radius of 1150 mm and provides tracking information for charged particles within $|\eta| < 2.5$ [44].

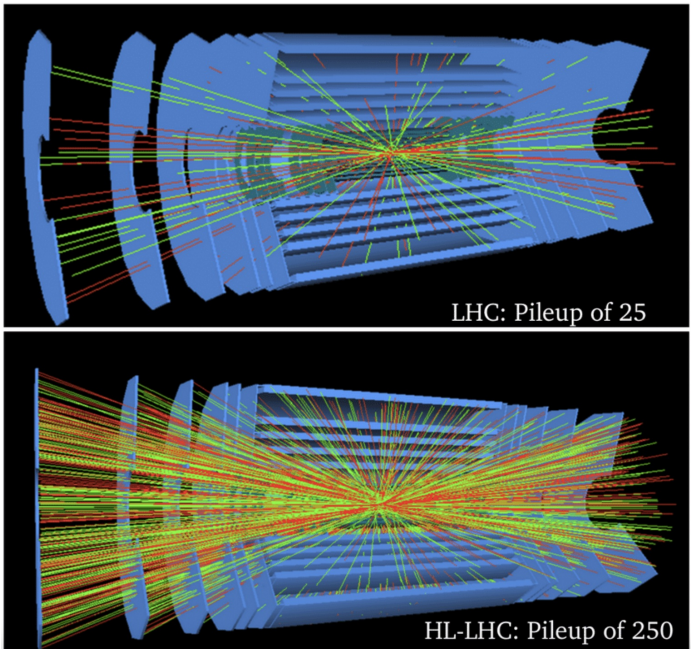


Figure 3.5: A depiction of a bunch crossing with 25 pileup events (top) and 250 pileup events (bottom). The high luminosity phase of the LHC will have many more primary vertices per bunch crossing. Figure taken from Ref. [46].

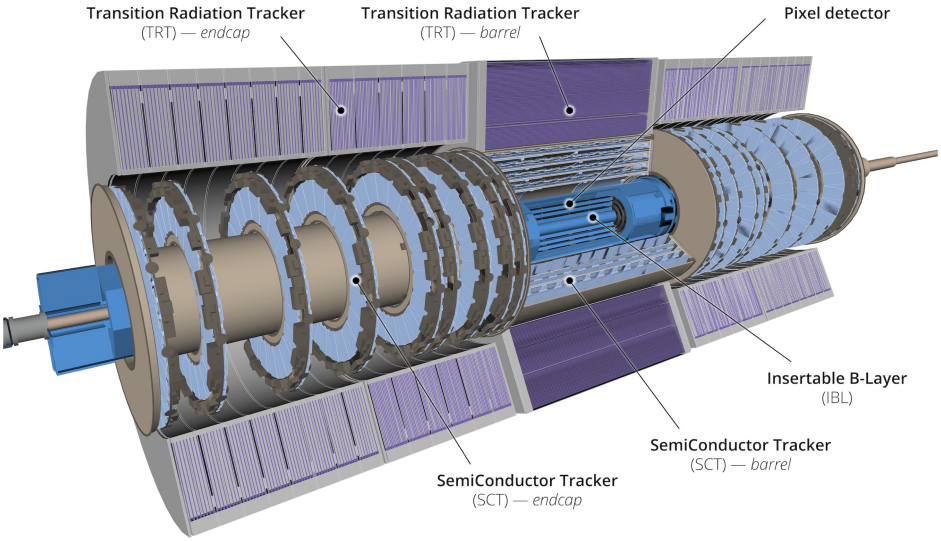


Figure 3.6: The Inner Detector is the innermost component of the ATLAS detector, composed of the Pixel Detector, Semiconductor Tracker, and Transition Radiation Tracker. Figure taken from Ref. [44].

A solenoid provides a 2 T magnetic field through the ID, which bends the trajectory of charged particles as they are passing through. The trajectory of a particle in the ATLAS detector is referred to as the *track*, and the deflection of a particle due to the magnetic field allows for a measurement of the momentum of charged particles. In addition to measuring the momentum of charged particles, the ID is also used for vertex identification. Tracks from charged particles in the ID are used to identify the location of primary and secondary vertices. Some PVs will occur very close together and get merged together when they are identified.

The IBL and Pixel Detector are closest to the beam line and composed of pixel sensors. The IBL is the closest layer, and was added in after Run-1 of the LHC to enhance tracking performance [44]. The IBL forms a cylinder around the beam line, composed of 50 µm × 250 µm pixels. Outside of this are three more layers of 50 µm × 400 µm pixels, which form the barrel region of the Pixel Detector. On either side of the barrel region are disks containing pixel sensors that form the end-cap region of the pixel detector. Charged particles passing through these pixels will create 3-dimensional hits along their trajectory.

Outside of the Pixel Detector is the SCT, which is composed of four layers of silicon microstrips in the barrel region and nine disks on either endcap region [47]. The strips run at a small angle to the z-axis, to provide measurements for r, ϕ and z. Hits in the silicon strips from charged particles provides additional tracking information to complement the Pixel Detector.

The TRT is the last layer of the ID, and is composed of proportional drift tubes called straws [48]. These tubes are 4 mm in diameter and 144 cm long, oriented parallel to the beam pipe. The TRT covers the region $|\eta|$ <2.0, and particles passing through

will cross 35-40 tubes. These straws contain a gas mixture of 70% Xe, 27% CO_2, and 3% O_2, which gets ionized as charged particles pass through and the resulting charge is collected and read out at an anode. The TRT also has a radiator material between the straws, which emit transition radiation when charged particles pass through.

3.2.4 Calorimeters

Calorimeters are located outside of the ID (Figure 3.7) and are designed to measure the position and energy of charged or neutral particles [44]. Most particles will interact with the absorber material of the calorimeter and produce a shower of particles, eventually bringing them to a stop and allowing for a measurement of their energy. ATLAS has electromagnetic calorimeters to measure the energy of electrons and photons, surrounded by hadronic calorimeters for measuring the energy of hadrons.

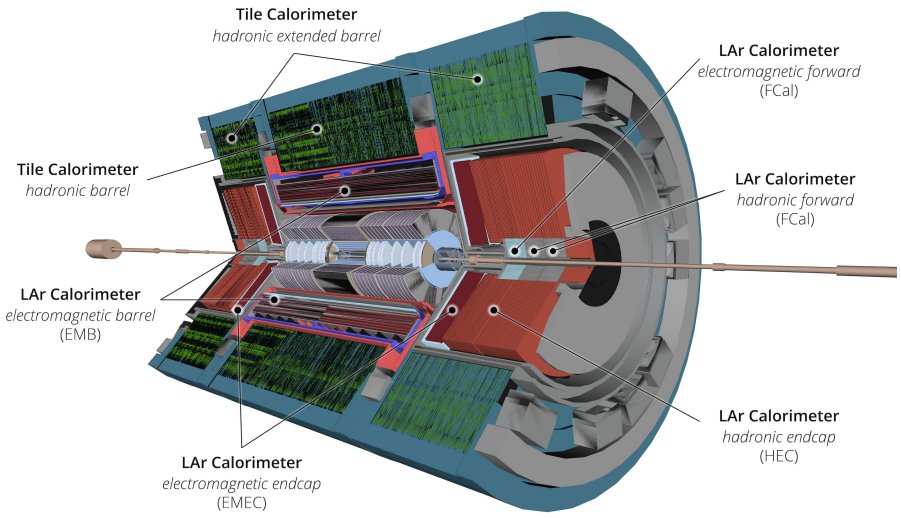


Figure 3.7: The calorimeter system is located outside of the Inner Detector. There are liquid argon electromagnetic calorimeters and hadronic calorimeters composed of scintillating tiles. Figure taken from Ref. [44].

3.2.4.1 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) uses liquid argon (LAr) and consists of a LAr Electromagnetic Barrel Calorimeter (EMB), covering $|\eta| < 1.475$ and a LAr Electromagnetic Endcap Calorimeter (EMEC), covering $1.375 < |\eta| < 3.2$ and the first layer of the Forward Calorimeter (FCal) covering covering $3.1 < |\eta| < 4.9$ [44]. The EMB and EMEC use lead as a high density absorber material to slow down and stop particles, interleaved with LAr as the active material to measure the energy deposited. In the first layer of the FCal, copper is used as the absorber.

There are three layers in the EMB and a presampler layer [49]. The presampler is necessary to account for energy lost by particles before they reach the calorimeter. Particles can pass through multiple radiation lengths of material depending on their pseudorapidity, as they pass through components before the calorimeters, such as the ID. The radiation length (X_0) is the mean length over which an electron loses $1/e$ of its energy due to bremsstrahlung radiation [50]. After the presampler are three layers with progressively coarser granularity, which provides a depth of $22X_0$. The first layer provides the best position resolution and captures the initial part of the electromagnetic showers. The second layer captures the bulk of the shower and provides the best energy measurement. The third and final layer is the thickest and captures any of the remaining shower from very high energy electrons and photons. An accordion-shaped geometry ensures that particles pass layers of absorber materials and active materials symmetrically in ϕ with no gaps.

Similarly to the EMB, the EMEC has three layers with an accordion structure to absorb electromagnetic showers in the forward region; a higher granularity first layer for accurate position measurements, a second layer for absorbing the bulk of the energy

and a third layer to capture the tail-end of high energy showers. The EMEC provides a depth of $24X_0$.

The first layer of the FCal consists of longitudinal channels, which form a honeycomb shape when viewed from the end. These channels are surrounded by copper, which is the absorber material used to cause electromagnetic particles to shower. Inside the channels are rods held at a high voltage to collect the ionization signal. There is a gap of 0.25 mm around the rods filled with the active LAr. The FCal provides a depth of $29X_0$.

3.2.4.2 Hadronic Calorimeter

Hadronic calorimetry is provided by the tile calorimeter, covering $|\eta| < 1.7$, Hadronic Endcap Calorimeters (HEC), covering $1.5 < |\eta| < 3.2$ and the second two layers of the FCal covering covering $3.1 < |\eta| < 4.9$ [49]. The HEC and FCal use LAr and high density absorbers, while the tile calorimeter uses scintillating tiles as an active material and steel as the absorber. For hadronic showers, the nuclear interaction length, λ, is the mean distance over which a hadron has a $1/e$ chance of an inelastic collision.

The tile calorimeter is composed of one barrel and two extended barrel components [51]. These contain alternating sheets of steel tiles and plastic scintillating tiles. When a hadron passes through the steel it will interact and create a shower of particles. These particles then pass through the scintillating tiles and generate photons, which are directed towards photomultiplier tubes (PMTs) by fibres coupled to either side of the scintillator so that a signal can be measured. The tile calorimeter is approximately 7.4λ [52].

The HEC covers $1.5 < |\eta| < 3.2$ and is located directly behind the EMEC [51]. It consists of four wheels (two on either side) with LAr as the active material and copper-plate absorbers. The HEC has a total depth of 10λ on either side.

The second and third layer of the FCal are used to measure the energy of hadrons in the most forward region of the detector. Similar to the first layer of the FCal, the second and third layers have a honeycomb shape with small gaps filled with LAr as an active material. However, tungten is used as the absorber for hadronic jets, instead of copper that is used in the first layer. The two FCal layers provide a total depth of 6.9λ.

3.2.5 Muon Spectrometer

The Muons Spectrometer (MS) [44] forms the outermost part of the ATLAS detector, as seen in Figure 3.8. Muons require a dedicated detector system to be measured, as they are minimum ionizing particles (MIPs) [50], and they pass through most of the previous subdetectors without depositing substantial energy. The MS is composed of a barrel, two endcaps and two "extended" endcaps, which use a multitude of technologies to measure muons in the range $|\eta| < 2.7$. Muons in the MS are deflected by magnetic fields from large superconducting air-core magnets. A large barrel toroid provides the magnetic field for $|\eta| < 1.4$, and two smaller endcap toroids provide a magnetic field for $1.6 < |\eta| < 2.7$. In the region $1.4 < |\eta| < 1.6$, the magnetic field is provided by a mixture of the barrel and endcap toroids.

The MS barrel system is composed of an inner, middle, and outer barrel, which are cylindrical units that use monitored drift tube (MDT) technology for precision measurements and resistive plate chambers (RPCs) for triggering and measuring ϕ [53].

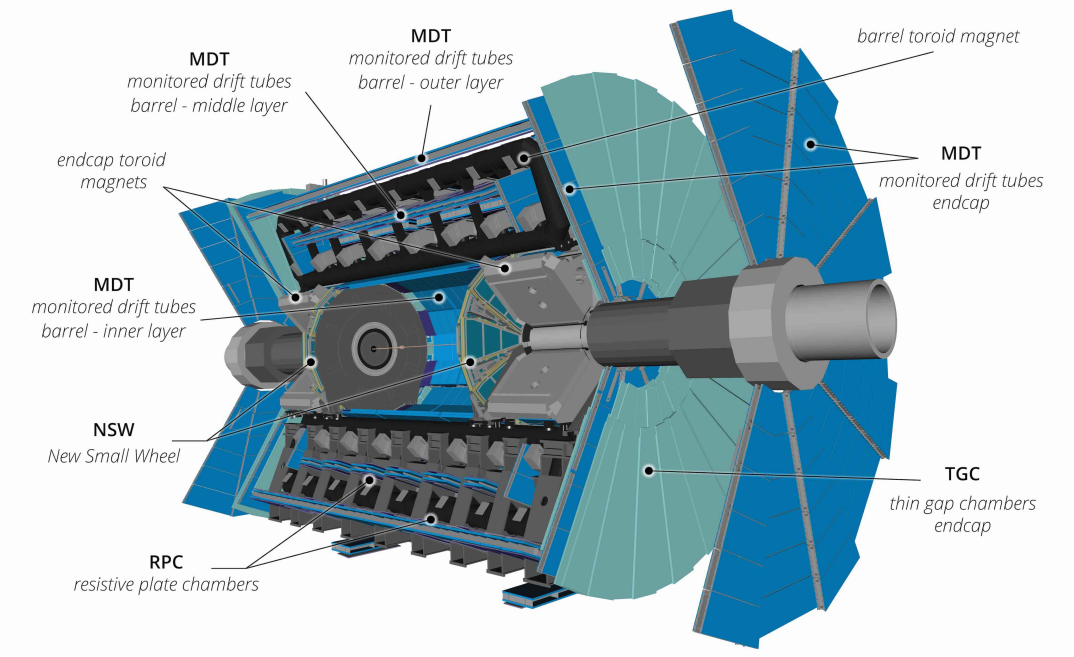


Figure 3.8: The Muon Spectrometer uses five different detector technologies and forms the outer layer of the ATLAS detector. The New Small Wheel (NSW) that was installed for Run-3 can be seen. Figure taken from Ref. [44].

The MDTs are filled with a gas mixture that gets ionized when muons pass through. Free electrons drift towards a wire held at a large potential difference, and produce a pulse. The drift time to the wire is used to determine the position of the muons as they pass through the MDTs. RPCs are formed with two parallel resistive plates with a small gap filled with a gas mixture. When muons pass through they ionize the gas and a strong electric field accelerates the electrons to generate an avalanche of ionization. The RPCs have a time resolution of 1.5 ns, which makes them ideal for triggering on events as they can determine when particles were passed through with great precision.

In the forward region, the endcaps of the MS are composed of an inner, middle and outer wheel, along with extended endcap rings between the middle and outer wheel. The outer wheels contain MDTs and the middle wheels contain MDTs and thin gap

chambers (TGCs). TGCs have gas filling a narrow gap between a cathode plate and a second plate with anode wires. Similar to RPCs, when a muon passes through the TGC it will create an avalanche of electrons that can be detected. TGCs are used for triggering and measuring azimuthal angle of muons.

The inner wheels, often referred to as the small wheels, were replaced in Run-3 with the New Small Wheels (NSWs) [44], which had many components assembled at Carleton University. The NSW was designed for precision tracking, improved p_{T} resolution in the trigger and increased background rejection. The NSWs use small-strip thin gap chambers (sTGCs) and micro-mesh gaseous structure (Micromegas) detectors to provide tracking and triggering in the region $1.3 < |\eta| < 2.7$. sTGCs work similar to the aforementioned TGCs, but with a much smaller strip pitch providing improved resolution [54]. Micromegas consist of a drift region between a planar drift cathode and thin metallic mesh, and an amplification region between the mesh and the readout electrode. There is a gas gap of a few millimeters, so when charged particles pass through they will ionize this gas. Electrons will then drift towards the metallic mesh due to the drift field. A majority of electrons will pass by the mesh into the thin amplification region where a much stronger electric field causes an avalanche which is measured on the readout electrode. Micromegas have a very good spatial resolution and fast response time, allowing them to operate under high particle flux.

3.2.6 Trigger System

Under normal operating conditions, proton bunch crossings happen at the ATLAS detector at a rate of 40 MHz (every 25 ns), which makes saving data from every bunch crossing prohibitively expensive in terms of storage. In addition, a majority

bunch crossings only result in low energy "soft" collisions, where the resulting particle interactions are not interesting for a majority of analyses. The ATLAS trigger system solves these problems by only saving events that look interesting for physics analyses, thus reducing the total amount of storage required. ATLAS uses a two level trigger system [44]: a hardware-based Level-1 (L1) Trigger and software-based High Level Trigger (HLT).

The L1 trigger uses reduced-granularity information from the calorimeters and MS to search for events with high-p_{T} muons, electrons, photons, jets, τ-leptons decaying into hadrons, and events with large missing transverse energy or a large total transverse energy. The L1 trigger will accept events that satisfy requirements on particle type and multiplicity. Regions of Interest (RoI) where possible trigger objects have been located are also established by the L1 trigger. The L1 trigger reduces the rate of detector readout to a maximum of 100 kHz and has a latency of 2.5 µs.

Events accepted by the L1 trigger are passed to the HLT, where algorithms further investigate RoIs for possible trigger objects. The HLT uses the full granularity of the information calorimeters, Muon Spectrometer and tracking systems for improved particle identification. The HLT reduces the rate of data taking to approximately 3 kHz, and all events passing a HLT are saved to permanent storage for later use in physics analyses.

3.2.7 Luminosity Measurements

The LHC delivers a large particle flux to the ATLAS detector. In particle physics, we measure the luminosity, L, which is measured in particles per unit area per unit time and is determined by the LHC beam properties. The luminosity from the LHC

is defined as

$$L = \frac{fN^2 n_b}{4\pi\sigma_x\sigma_y}, \tag{3.5}$$

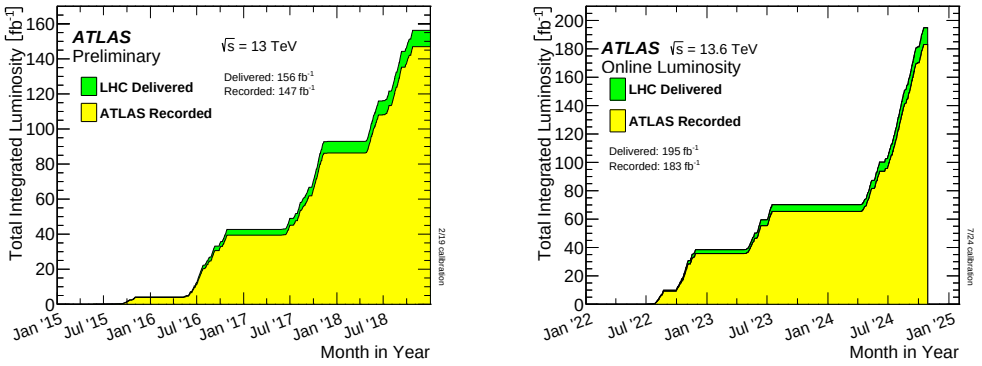
where f is the revolution frequency of the LHC, N is the number of protons in each colliding bunch, n_b is the number of bunches and σ_x and σ_y are the width of the beam spot in the xy plane at the ATLAS detector. All of these values will fluctuate through the course of a run at the LHC, making it difficult to calculate the luminosity being delivered. For this reason, ATLAS has a dedicated detector called LUCID-2 (LUminosity Cherenkov Integrating Detector) to measure the luminosity [55].

LUCID-2 is composed of two detectors on either side of the ATLAS detector on the beam line. Particles from collisions in the detector that travel in the forward regions will enter the detector and produce Cherenkov radiation if they are moving faster than the speed of light in the medium of the detector. This Cherenkov light is detected by PMTs, and the luminosity can be determined from the rate of particles detected by LUCID-2.

When analyzing data from ATLAS we are interested in the total integrated luminosity over that period of time,

$$\mathcal{L} = \int L \, \mathrm{d}t. \tag{3.6}$$

The integrated luminosity for pp collisions delivered during Run-2 and Run-3 is shown in Figure 3.9. The total integrated luminosity suitable for this analysis in Run-2 is $140\,\mathrm{fb}^{-1}$ [56] and for Run-3 from 2022 until 2024 is $165\,\mathrm{fb}^{-1}$.



(a) Run-2 (b) Run-3

Figure 3.9: The total integrated luminosity measured over the Run-2 [57] and part of the Run-3 [58] data-taking periods for *pp* collisions. The total integrated luminosity delivered by the LHC is shown in green and the total integrated luminosity recorded by the ATLAS detector is shown in yellow. Data taking for Run-3 resumed in May 2025, but this data was not included in this thesis.

3.3 ATLAS Phase-II Upgrades

After Run-3, the ATLAS detector is undergoing many upgrades to deal with the conditions of the HL-LHC. The HL-LHC will have up to 200 interactions per *pp* bunch crossing, resulting in higher data rates, radiation levels, and a high-occupancy environment for the detector components. The Phase-II ATLAS detector upgrades include an upgraded trigger and data acquisition system (Section 3.3.1), a new silicon tracking detector (Section 3.3.2), the High Granularity Timing Detector [59], and upgrades to existing detectors [60–62].

3.3.1 HL-LHC ATLAS Trigger Architecture

The ATLAS Trigger and Data Acquisition (TDAQ) system are going to be upgraded for the HL-LHC to deal with the increased instantaneous luminosity. There is a baseline scheme with a single-level trigger architecture, and an evolved scheme with

a split-level trigger architecture if trigger rates or occupancies get too high in the HL-LHC.

In the baseline trigger scheme, a hardware based level-0 (L0) trigger uses information from the calorimeters and Muon Spectrometer. If an event passes an L0 trigger, it is transmitted off to the Data Acquisition (DAQ) system at a rate of 1 MHz. An Event Filter (EF) system will then analyze the events passing L0 triggers and reduce the number of events down to the final output rate. Events passing the EF are moved to permanent storage for analysis.

The evolved trigger scheme has a two-level Level-0/Level-1 hardware system. In the evolved trigger scheme, events passing an L0 trigger may identify RoIs in the ITk that have potentially interesting physics. A Regional Readout Request (R3) is then sent to these parts of the detector which will transfer data for use in the L1 trigger. Events then have to pass an L1 trigger before being passed on to the EF and possibly saved to permanent storage.

3.3.2 The ITk Upgrade

The Inner Tracker (ITk) is replacing the current Inner Detector (Section 3.2.3) at the centre of ATLAS [63]. The increased luminosity at the HL-LHC will make the data rate too large for the current ID, so a new detector is required to maintain tracking performance without getting saturated. In addition, the ID has received extensive radiation damage throughout its lifetime and will not be suitable for operation in the HL-LHC. In order to cope with the increase in collisions and large radiation dose the HL-LHC, the ITk will be an all silicon detector.

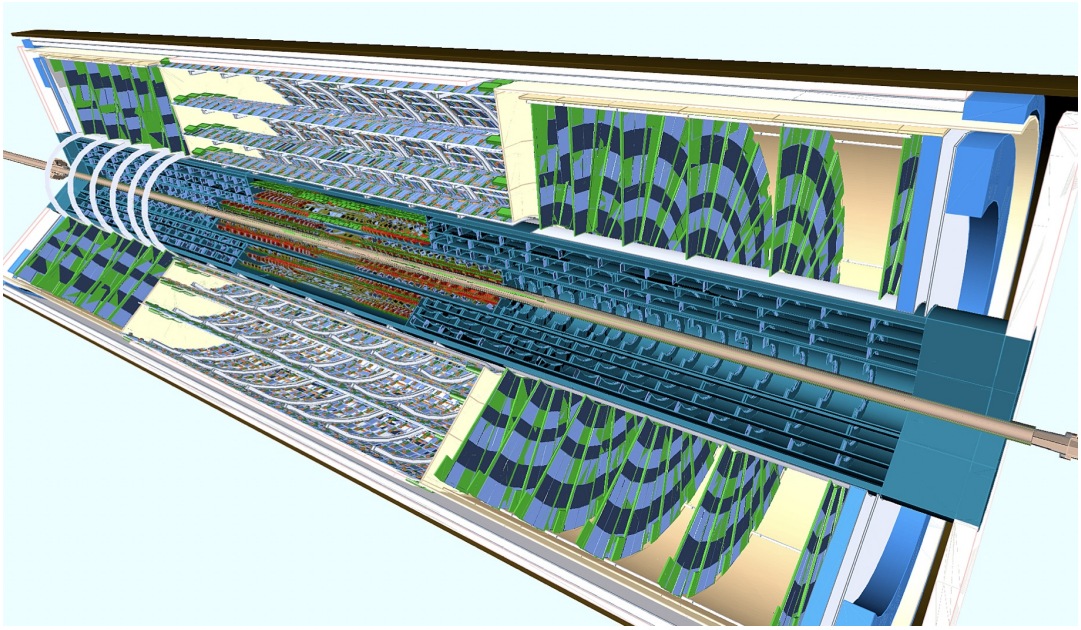


Figure 3.10: A visualization of the Inner Tracker (ITk). The ITk is composed of the Pixel Detector located closest to the beam line and the Strips Detector surrounding it. Figure taken from Ref. [63].

The ITk will be composed of the Pixel Detector closest to the beam line and the Strip Detector surrounding it, forming a cylinder as seen in Figure 3.10. The Pixel Detector will have roughly twice the radius and four times the length of the current ID Pixel Detector to provide increased granularity through a larger part of ATLAS. The Strip Detector surrounding this will also have a better granularity than the current SCT. The Pixel and Strip Detectors will require over three times the current amount of silicon of the current ID. The layout and coverage of the Pixel and Strip Detectors can be seen in Figure 3.11. The 2 T solenoid magnet currently used for the ID will continue to be used for the ITk. The ITk is designed to operate for at least 10 years with an instantaneous luminosity of $7.5 \times 10^{34}\,\mathrm{cm}^{-2}$ and a total integrated luminosity of $4000\,\mathrm{fb}^{-1}$.

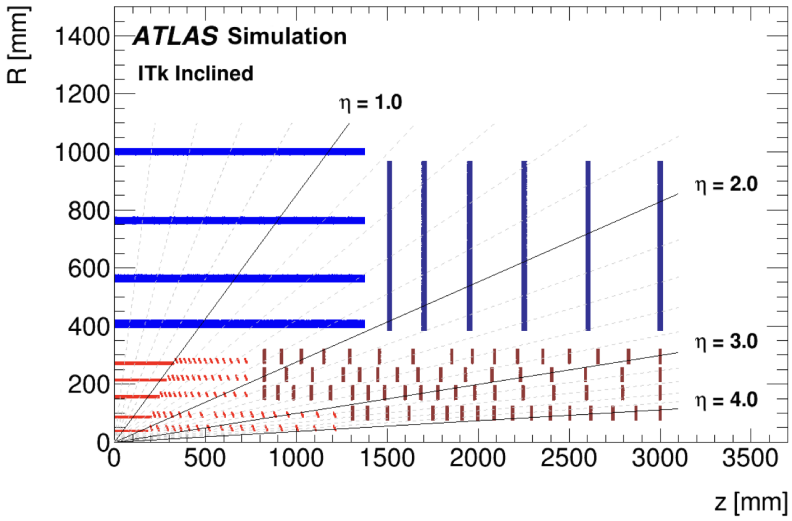


Figure 3.11: Schematic layout of one quadrant of ITk sensors. The beam line is along the horizontal axis with the center of the ATLAS detector at the origin. The active elements of the barrel Strip Detector are shown in blue and end-cap Strip Detector are shown in dark blue. For the Pixel Detector the sensors for the barrel layers are shown in red and for the end-cap rings they are shown in dark red. Figure taken from Ref. [63].

3.3.2.1 The ITk Pixel Detector

The Pixel Detector comprises the centre of the ITk with five layers of pixel sensors [64]. The two innermost layers of pixels are located in the Inner Support Tube. This will allow them to be replaced after around 2000 fb^{-1} of data collection, due to the high levels of radiation expected for these layers in the HL-LHC. Outside of this are three more layers of pixels between the Inner Support Tube and Pixel Support Tube that will remain in place for the lifetime of the ITk.

In the pixel barrel layer, the sensors in the most central region are parallel to the beam line, and sensors in the forward region are at an angle of 56° to the beam line. In layer 0, dual modules with two read-out chips will be used and in layers 2-4, quad modules with four read-out chips will be used.

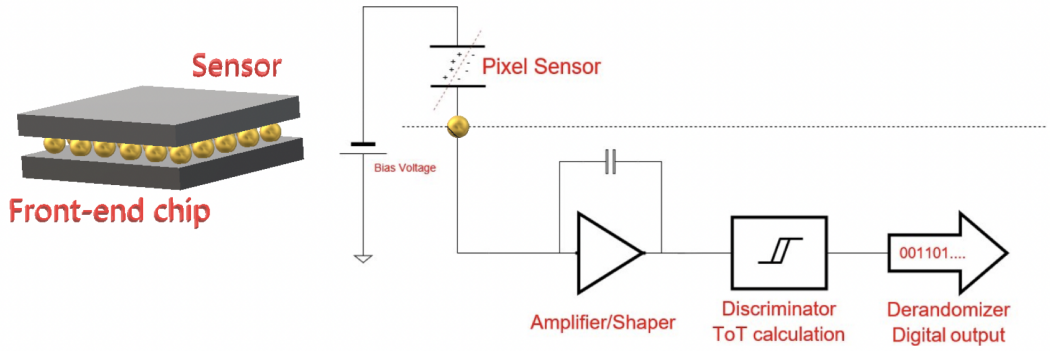


Figure 3.12: Schematic view of a pixel module without a PCB. Each pixel sensor has 1-4 front-end chips bump-bonded to it, which amplify, discriminate and readout any hits from charged particles. Figure taken from Ref. [64].

The pixel sensors have an inclined orientation in the endcaps. This is to reduce the amount of material that particles pass through and also reduces the amount of silicon required to cover then entire η range. In layer 0 of the endcap, single read-out chip modules will be used and in layers 2-4, dual modules with two read-out chips will be used.

The modules (Figure 3.12) are composed of a silicon sensor, a CMOS (complimentary metal-oxide-semiconductor) front-end readout chip and a flexible printed chip board (PCB). This is similar to the pixel sensors currently used in ATLAS, with a reduced pixel size of $50\,\mu m \times 50\,\mu m$ or $25\,\mu m \times 100\,\mu m$ to improve resolution. A single-chip module is about $2\,cm \times 2\,cm$, a double-chip module is $4\,cm \times 2\,cm$ and a quad-chip module is $4\,cm \times 4\,cm$.

3.3.2.2 The ITk Strip Detector

The ITk Strip Detector extends from $-1400\,mm$ to $1400\,mm$ on the z-axis covering $|\eta| < 2.7$ [63]. It surrounds the Pixel Detector with four barrel layers and one end-cap

on each side. The end-caps have six discs to provide coverage within 10° of the beam line. The barrel sensor is composed of units called staves, with two different layouts designated short-strip and long-strip. The building blocks of the end-cap are called petals, with six diffent layouts named R0-R5 corresponding to the 6 end-cap rings. Both staves and petals (Figure 3.13) are composed of silicon-strip modules that consist of one sensor and one or two hybrids containing the readout ASICs.

The two inner layers of the barrel sensor where track density will be the highest have a short-strip layout with strips that are 24.1 mm in length. A long-strip layout with 48.2 mm long strips is suitable for the two outer layers of the barrel. Both layouts create a nearly square module that is 97.620 mm × 97.950 mm by combining either four short-strip segments or two long-strip segments.

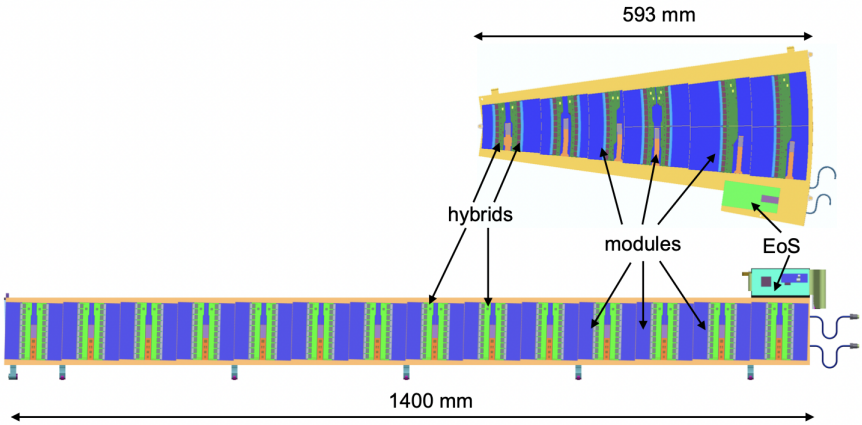


Figure 3.13: Sample petal (top) and stave (bottom) layouts. Figure taken from Ref. [63].

The six R0-R5 end-cap sensor layouts are more complex than those of the barrel sensor. The sensors are all wedge-shaped with circular arcs following the beam line forming the top and bottom edges. The three inner rings (R0-R2) have one module with one or two hybrids each. The three outer rings (R3-R5) have two modules that share a single hybrid. The complex geometry of the end-cap sensors requires six different

sensor geometries and thirteen different hybrids.

Modules are the building blocks of the ITk strip detector. The base of the module is the silicon sensor. One or two printed circuit boards (PCBs), called hybrids, are then glued on top of the silicon sensor using electronics grade epoxy. The hybrids host up to 12 ABCStar (ATLAS Binary Chip – Star Version) ASICs (described in Chapter 4) that are wire-bonded to the silicon strips for front end readout. Each hybrid also contains a single HCCStar (Hybrid Controller Chip – Star Version) ASIC that interfaces with the ABCStar chips and the stave or petal service bus. Finally, a power board with a DC-DC converter is glued to the sensor. The number of hybrids and readout ASICs differs between modules for different petal and stave layouts.

The ABCStar and HCCStar ASICs use "star architecture" (Figure 3.14), where all ABCStar chips connect directly to the HCCStar [63]. Previous versions of the chip used daisy chain signal routing, but this was upgraded to improve the readout rate of the chips. This allows an L0 readout rate of 1 MHz for the ITk Strip Detector, which may be required for some high occupancy regions of the detector.

For a short-strip barrel module, the sensor is composed of four sets of strips, each 24 mm long, resulting in a sensor that is roughly 97 mm $\times$ 97 mm. An exploded view of a short-strip barrel module can be seen in Figure 3.15.

3.3.2.3 ITk Tracking and Physics Performance

The ITk is designed to be able to deliver equal or better tracking performance to the current ID with a large increase in interactions per bunch crossing. The ITk will provide performance equal to or greater than that of the ID in the $|\eta| < 2.7$ region. It also provides extended coverage in the very forward region up to $|\eta| < 4.0$. There must

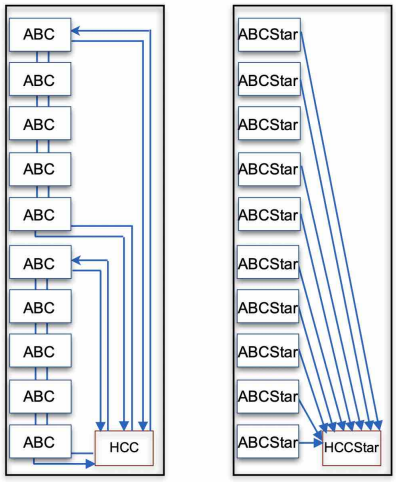


Figure 3.14: Daisy chain routing used for previous ABC and HCC chips (left) and the new star routing used to increase the readout capabilities (right). Figure taken from Ref. [63].

Table 3.1: Required efficiency for different particles in different η regions. Values obtained from Ref. [63]).

Particle	$\boldsymbol{p_T}$ **range**	**Required Efficiency**	
		$\lvert\eta\rvert \leq 1.0$	$1.0 < \lvert\eta\rvert < 2.7$
Muon	$p_T > 3$ GeV	$> 99\%$	$> 99\%$
Pion	$p_T > 1$ GeV	$> 90\%$	$> 85\%$
Electron	$p_T > 5$ GeV	$> 90\%$	$> 85\%$

be full tracking coverage in this region, and the detector response should be uniform in ϕ. Together, the Pixel and Strip Detectors will provide 13 hits for charged particles with $|\eta| < 2.7$ (excluding the barrel/end-cap transition of the Strip Detector, where the hit count is 11), and 9 hits for particles with $|\eta| > 2.7$. The tracking performance for different particle types can be seen in Table 3.1. This is the minimum required efficiency that must be met by the ITk for each particle type.

The reconstruction efficiency for muons, pions and electrons in ITk simulations can be seen in Figure 3.16. Muons show the best possible reconstruction efficiency because

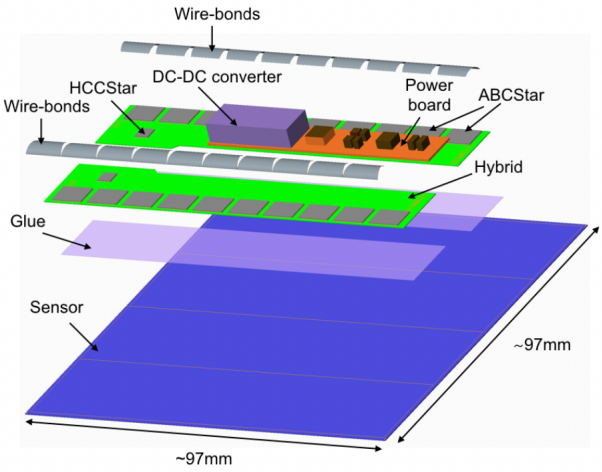


Figure 3.15: Exploded view of a short-strip barrel module. Long-strip barrel modules and end-cap modules are composed of the same components. Figure taken from Ref. [63].

they are MIPs that do not interact as strongly with the detector, or radiate as much energy as electrons. The reconstruction efficiency close to 100% for muons in the ITk. Other particles, such as hadrons and electrons, will radiate energy as they interact with detector material, which leads to a lower efficiency. The ITk has less material than the ID, so particles undergo less interaction lengths (λ), leading to a better reconstruction efficiency as seen in Figure 3.16.

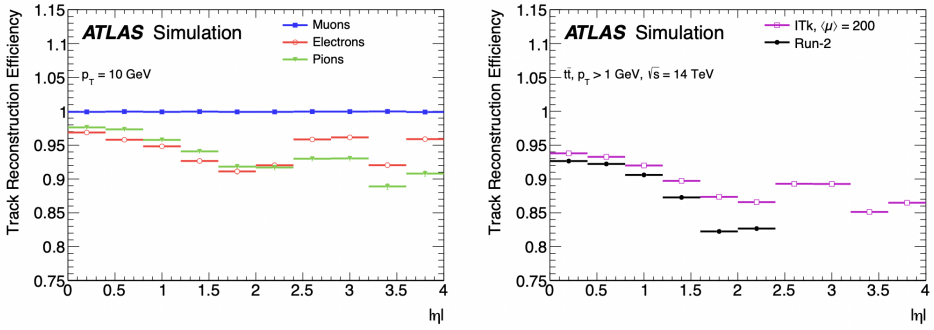


Figure 3.16: Left: The track reconstruction efficiency for muons, pions and electrons with p_{T} =10 GeV. Right: Track reconstruction efficiency for particles from $t\bar{t}$ events with an average of 200 interactions per bunch crossing for the ITk and the ID. Figure taken from Ref. [63].

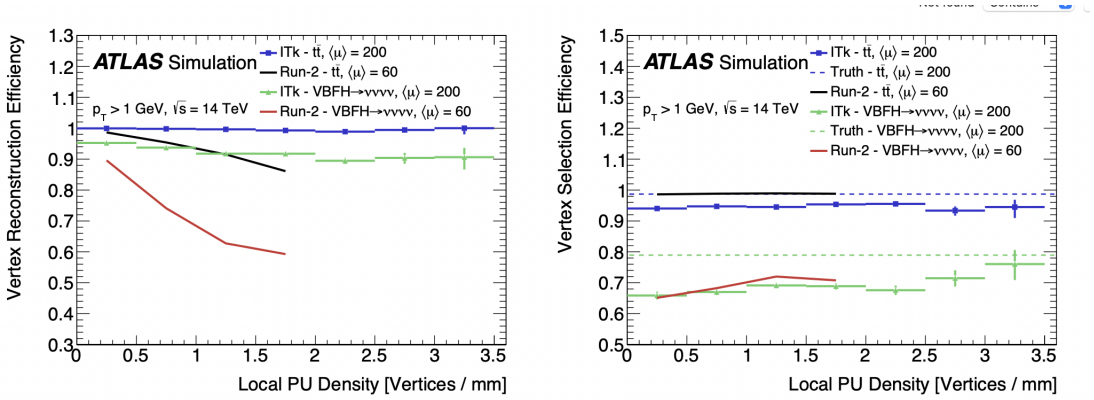


Figure 3.17: The primary vertex reconstruction efficiency (left) and identification efficiency (right) for $t\bar{t}$ events as a function of local pileup density with the ITk ($\langle\mu\rangle = 200$) and the Run-2 ID ($\langle\mu\rangle = 60$). Results for VBF $H \to \nu\nu\nu\nu$ events are also shown. PV selection is based on the vertex with the largest sum of p_{T} of the tracks, and the dotted lines indicate when the true PV actually has the largest sum of p_{T}. Figure taken from Ref. [64].

Vertex reconstruction will need to be performed for events with $< \mu >= 200$ at the HL-LHC [63]. The primary vertex of the hard scatter event will typically be within 1 mm of other vertices, meaning not all PVs will be able to be identified separately. The vertex reconstruction efficiency of the ITk is defined with respect to $t\bar{t}$ events. There must be a greater than 95% chance that the $t\bar{t}$ vertex is reconstructed, and a greater than 90% chance that the $t\bar{t}$ decay is assigned to the correct vertex. Figure 3.17 shows that the ITk can satisfy both of these requirements with $< \mu >= 200$. A more difficult case of VBF $H \to \nu\nu\nu\nu$ is also shown, where there will be four neutrinos that escape without detection and two forward jets. The ITk outperforms the ID for vertex reconstruction efficiency, despite an increase in pileup in both cases.

Measurements from the ITk are used to reconstruct the track of charged particles that pass through. The track parameters are the transverse (d_0) and longitudinal (z_0) location of the PV, the polar (θ) and azimuthal (ϕ) angles, and the transverse momentum from curvature of the track. The resolution for each parameter affects how

well leptons and jets can be identified and affects the ability to identify the flavour of quarks in jets [64]. The resolution for the ITk is shown in Figure 3.18. The resolutions are comparable between the ITk and ID for most parameters. The d_0 resolution is slightly degraded for 100 GeV muons, due to a larger radius for the first pixel layer, which is required because of the harsh radiation environment for the HL-LHC. The z_0 resolution improves with the ITk due to a decreased pixel pitch in the z direction. The resolution for track parameters also remains good in the forward region for $|\eta| > 2.5$, where new tracking information will be available outside of the current ID region.

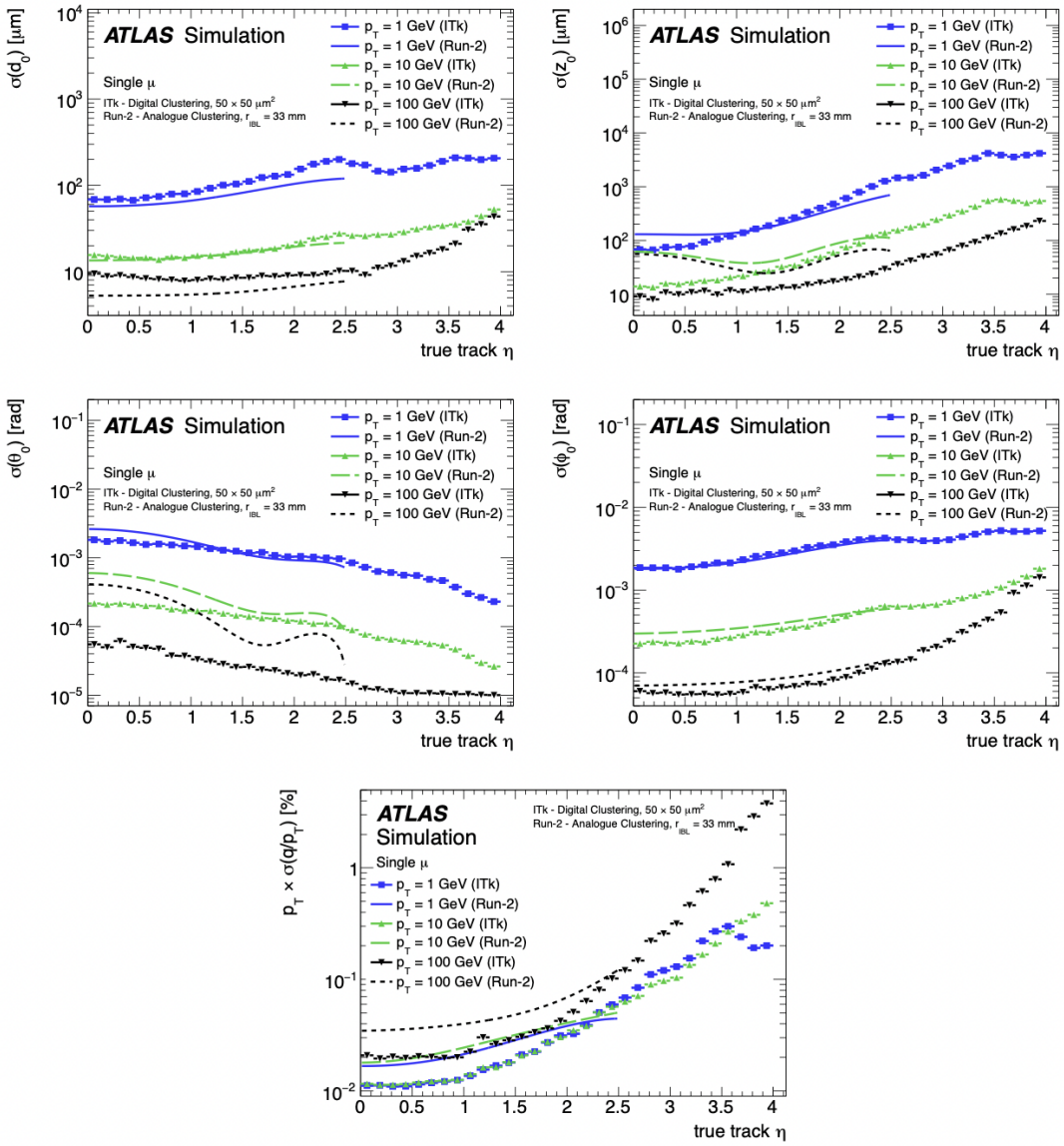


Figure 3.18: Track parameter resolution in d_0, z_0, θ, ϕ and p_{T} as a function of η for an ITk Pixel Detector with $50\,\mu\mathrm{m} \times 50\,\mu\mathrm{m}$ pixels. Results are shown for 1, 10 and 100 GeV muons using a digital clustering scheme. The resolutions for the current Run-2 Inner Detector are shown for reference. Figure taken from Ref. [64].

Chapter 4

ABCStar ASIC and Characterization

This chapter outlines the ABCStar readout chip for silicon strip sensors in the ATLAS Inner Tracker upgrade. Section 4.1 gives an overview of the ABCStar design and Sections 4.2–4.4 describe the on-wafer testing of the ABCStar, a major contribution of this thesis. Section 4.5 outlines a problem encountered with the digital logic of the ABCStar during production testing and how this problem was resolved. An overview of ABCStar on-wafer testing results is presented in Section 4.6 and handling of the ABCStar wafers and data collected after testing is covered in Sections 4.7 and 4.8 respectively.

4.1 ABCStar Front End Readout Chip

The ABCStarV1 (ATLAS Binary Chip – Star version) ASIC is the final version of the readout chip for the ITk silicon strip sensors (Section 3.3.2.2). There was a previous version called the ABCStarV0, but ABCStar in this thesis will refer to the ABCStarV1 unless stated otherwise. The ABCStar provides signal processing for 256 silicon strips on a sensor and employs binary readout architecture [63, 65, 66]. Input signals from all 256 channels are amplified, shaped, discriminated and readout to HCCStar chips upon receiving the appropriate commands. An ABCStar chip can be seen in Figure 4.1 and has dimensions of $(8.047 \pm 0.010)\,\text{mm} \times (7.110 \pm 0.010)\,\text{mm}$.

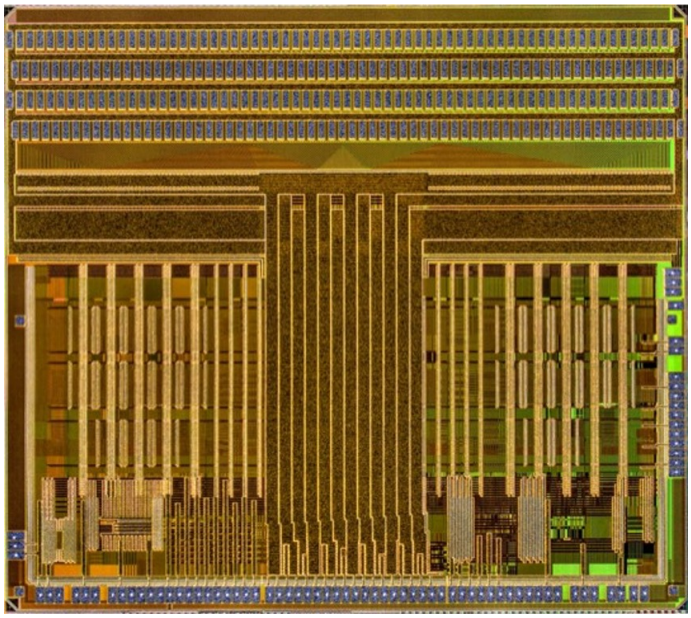


Figure 4.1: A picture of a single ABCStar chip.

ABCStar ASICs are manufactured using 130 nm CMOS technology at an ASIC foundry. The ASICs are fabricated on large silicon substrates called *wafers*. A fully fabricated ABCStar wafer can be seen in Figure 4.2. After the fabrication process, each wafer will contain multiple functional ASICs, which are later cut into individual chips using

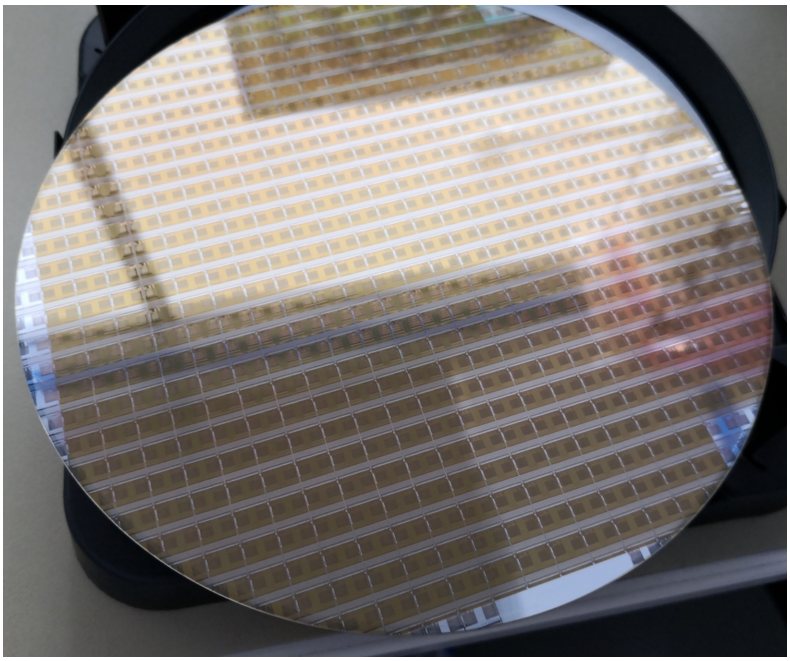


Figure 4.2: An image of an ABCStar wafer. Each wafer contains up to 470 fully functional ABCStar ASICs.

a process called *dicing* (Section 4.7). Wafers are usually manufactured in groups of 25 referred to as *batches* or *lots*. Although steps are taken to ensure consistent fabrication of ABCStar wafers at the foundry, there is expected to be variation in performance between batches of wafers, between wafers in the same batch and even between ASICs on a single wafer. One process involved in manufacturing of wafers is called spin coating [67], and can lead to radial variation in performance on a wafer.

The block diagram for the ABCStar is shown in Figure 4.3. When a charged particle passes through the silicon strip sensor, a small charge will be received by the ABCStar channel wire-bonded to that strip. The ABCStar front-end amplifies this signal into a voltage and discriminates it against a set voltage to determine if a digital "hit" is recorded for that channel. The front-end is optimized for 2.5 cm strips with a 25 ns shaping time and noise below 1000 electrons at end of life. A biasing circuit controls

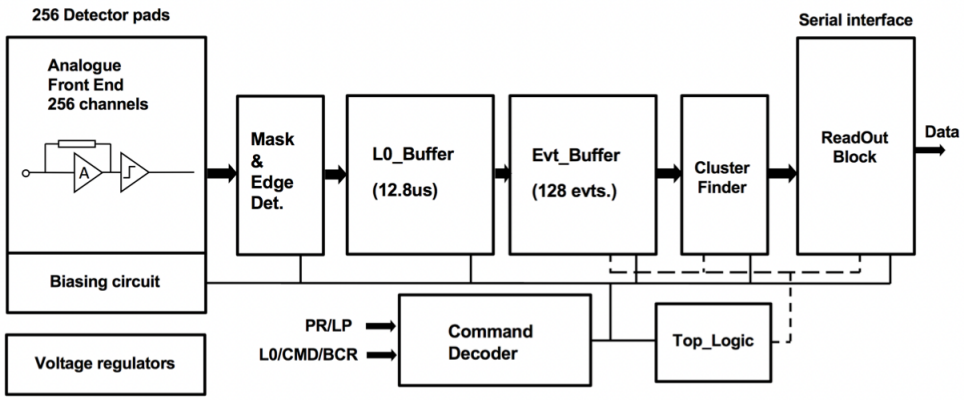


Figure 4.3: The block diagram of the ABCStar ASIC. Figure taken from Ref. [63].

the sensitivity of the 256 channels connected to the ABCStar.

At each proton bunch crossing all of the channels are sampled and any hits are filtered by a mask register to filter out any bad or noisy channels. The resulting hit pattern is then stored in the L0 buffer, where it is held for 12.8 µs. If a Level-0 Accept (L0A) trigger (see Section 3.3.1) is received, the hit counter and bunch crossing clock (BC) count corresponding to the desired latency are transferred to the event buffer. The event buffer can hold up to 128 events, which with an average L0 readout rate of 1 MHz are stored for an average of 128 µs. If a priority readout request (PR) or low priority readout request (LP) is received from an HCCStar, the event of interest is moved from the event buffer to the cluster finder. The cluster finder allows for data reduction by creating a "cluster" byte for channels found with hits. The readout block formats the cluster bytes with the event identification and transmits the data out at 160 Mbit s^{-1}.

All incoming requests, such as PR, LP and register reads are obtained from the HCCStar. These are interpreted by the Top_Logic block (Figure 4.3), and are completed based on a priority logic scheme for readout packets. Signals are sent to and from

Table 4.1: List of SLVS Pads on ABCStar the ABCStar.

Signal Name	**Description**
BC_P/BC_N	40MHz Bunch Crossing clock
PRLP_P/PRLP_N	80MHz Priority (PR) and Low Priority (LP) trigger command input
CLK_P/CLK_N	160MHz Readout Clock
LCB_IN_P/LCB_IN_N	L0A/CMD/BCR command input. The Level-0 Accept (L0A) signal is used to move events of interest to the event buffer. Command (CMD) signal is used for setup and resets. Bunch Counter Reset (BCR) resets the bunch crossing ID to zero.
DAT/DATB	Data Output

the ABCStar with the scalable low-voltage signalling (SLVS) pads listed in Table 4.1. In order to communicate with an ABCStar an *idle pattern* must be established and maintained with LCB_IN (L0A/CMD/BCR). The idle pattern is a repeating signal that is sent to the ABCStar that does not contain any encoded data. A lock is initiated by sending the idle pattern 16 times, and then it must be maintained or the locked state will be lost and must be established again before communication can resume. While the lock is maintained, other information can be sent to the ABCStar through LCB_IN. When the lock status changes, the ABCStar will send a high priority register (HPR) packet showing the lock status.

The ABCStar has a 40 MHz bunch crossing clock (BC), that matches the frequency of proton-proton collisions in the detector, and a 160 MHz readout clock (CLK). The timing of the clocks can be seen in Figure 4.4. A clock is an alternating square wave function that alternates from high (1) to low (0) with a fixed period. The ratio of high time to low time is called the duty cycle, and a 50/50 duty cycle indicates an equal amount of high time and low time in each period of the clock. A majority of the digital logic for the ABCStar is triplicated, meaning that there are three copies

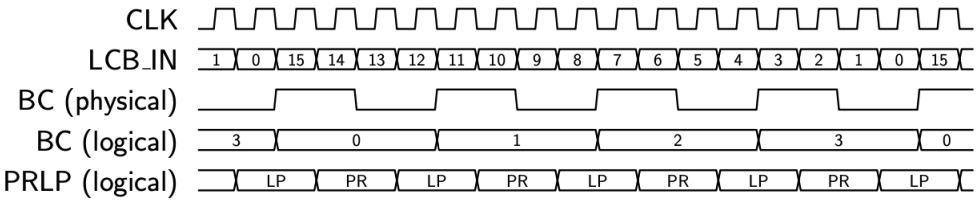


Figure 4.4: Timing of the 160 MHz CLK and 40 MHz BC on the ABCStar. Each LCB_IN frame consists of 16 bits, which takes 4 BC cycles. LP is sampled on the rising edge of the BC and PR is sampled on the falling edge of the BC, resulting in an 80 Mbit s^{-1} PRLP signal.

of all information supplied by different clock trees. Both BC and CLK have three phases (A/B/C) in the ABCStar. The results from all three clock trees are used and the majority (2/3 trees) result is used. This means that if a glitch were to happen on one of the clock trees, it will not corrupt operation of the chip as this mistake will get outvoted by the other two clock trees. If a correction does occur in the triplication for certain registers, there is a Single Event Upset (SEU) counter in the ABCStar that will record the event. The LCB_IN signal uses DC-balanced transmission, meaning that there are an equal number of 0s and 1s. This makes it easy to detect bit flips from radiation, as the DC-balanced transmission will be broken.

The ABCStar chips show a change in digital current with irradiation. This change in current is a function of the total ionizing dose (TID) received and is a known effect for 130 nm CMOS chips [68, 69]. The digital current will initially increase with TID, then eventually decrease, as seen in Figure 4.5. This bump in the digital voltage is referred to as a *TID bump*, and is a result of two competing effects. Radiation damage to the chips will start to increase the current consumption due to leakage current from positive charges that are quickly trapped in the oxide at the edge of the transistor. Negative charge trapped in interface states is slower to build up, and will begin to compete with the oxide-trapped states, bringing the current back down. There is no

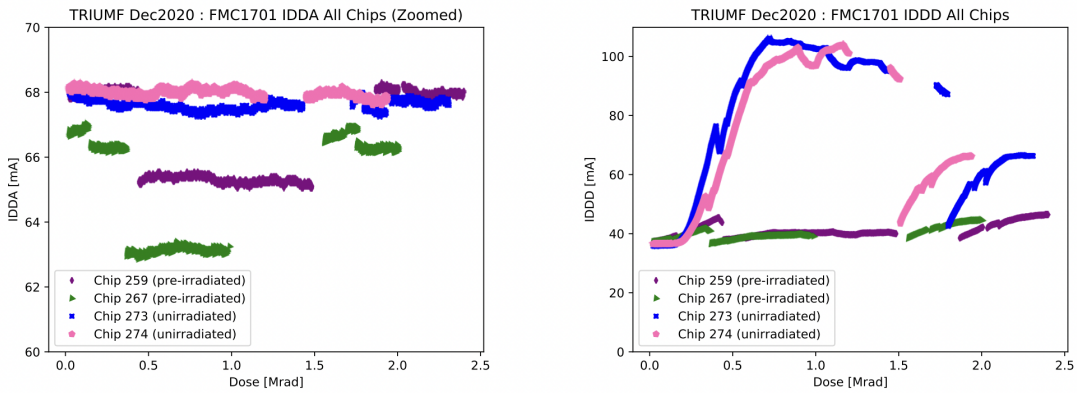


Figure 4.5: The analog (left) and digital (right) currents drawn by each ABCStar ASIC as a function of TID. There is a temporary increase in the digital current drawn as a function of dose (TID bump), but not the analog current. Two chips that had been pre-irradiated past the TID bump do not show an increase in digital current drawn. The piece-wise behaviour of the digital current curves is a result of stopping/starting data collection runs. Figure taken from Ref. [1].

increase in the analog current, due to the large size of the transistors and the current bias controls.

In order to prevent the increase in digital current of the ABCStar chips during operation of the ITk, the chips are pre-irradiated before installation. All ABCStar ASICs that pass testing to be used in the detector are taken to facilities to be irradiated and receive a TID to get them past this increase in digital current. This prevents an increase in current after installation in the ITk, as seen for the pre-irradiated chips in Figure 4.5. Irradiation of the ABCStar ASICs installed in the ITk during operation will not result in an increase in digital current once they are past this TID bump.

4.2 ABCStar On-Wafer Testing

In order to fulfill the needs of the ITk upgrade, the analog and digital circuitry of over 350,000 ABCStar ASICs need to be thoroughly tested to ensure they can accurately

process the high rate of collision data within the HL-LHC. The first round of testing is on-wafer probing, which occurs before the wafer is diced into individual chips. After the fabrication process, each ABCStar wafer contains up to 470 fully functional ASICs. Each wafer was designed to have 466 functional ASIC locations, but inspection revealed four additional ASICs outside this region that appeared to be fully processed and occasionally passed testing. These wafers are tested using setups called probe stations, so the testing process is commonly referred to as probing (probing is synonymous with testing here).

Each wafer is given a unique 7-character alphanumeric identifier after fabrication. These identifiers start with 'A' or 'V', such as V2D1GUH, and are used to distinguish ABCStar wafers during probing. In addition, a serial number (SN) is assigned to each wafer and each individual ABCStar to be used for tracking throughout the life of the ITk. The SNs have the format 20USGwwxyyyzzz, and the details of the SN for ABCStar ASICs and wafers are as follows:

- 20: ATLAS detector component.
- U: Phase-II ATLAS upgrade.
- S: Strips Detector component.
- G: General component used in both the barrel and endcap of the Strips Detector.
- ww: Unique code for each individual component. WA for ABCStar wafers and AA for ABCStar ASICs.
- x: 0 for pre-production parts, 1 for production parts.
- yyy: Unique number for each ABCStar wafer.

- zzz: Unique number for each ABCStar ASIC. 000 for ABCStar wafer SNs.

20USGWA1788000 would indicate ABCStar wafer number 788. 20USGAA1788005 would be the serial number for ABCStar ASIC 5 on wafer 788. These SNs in combination with the unique eFuse on each ABCStar allow ASIC to be identified throughout each stage of installation and operation in the ITk and previous test results can be retrieved to help with troubleshooting when necessary.

On-wafer probing of the ABCStar is performed at two sites with fundamentally different approaches. Half of the ABCStar wafers are probed at Rutherford Appleton Laboratories (RAL), a dedicated physics research facility in the UK. Probing is done using a probe station, which can probe a wafer in roughly 3.5 hours and requires human intervention to swap to new wafers. The other half of the wafers are being probed by Carleton University, in a partnership with a specialist wafer testing company, DA-Integrated[1] (DA). At DA, probing is performed with an automated wafer test system, which can be loaded with a canister of up to 25 wafers, and test each wafer automatically in just over 2 hours. Probing at RAL has the advantage of unlimited access to wafer probing, allowing for quicker debugging of issues and R&D, while probing at DA is much quicker, and has enhanced probing capabilities.

The partnership between Carleton and DA-Integrated probing wafers is a first in the particle physics community. Traditionally all testing has been done at dedicated research facilities, so this partnership has bridged the methodological and technical gap between research facilities and the semiconductor testing industry. Although the methodology of probing differs between the two sites, rigorous cross-checks were performed to ensure that probing results were consistent between the two facilities

[1] https://www.da-integrated.com

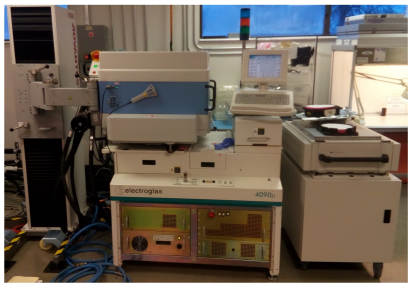
(Section 4.6.3).

4.2.1 DA-Integrated Probing Setup

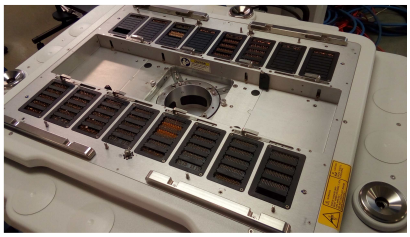
ABCStar wafers are probed at DA using either a Tokyo Electron Limited prober or an Electroglas 4090μ Automated Wafer Test System (ATE) (Figure 4.6) with a custom probe card. These systems are integrated with an Advantest V93000 Smart Scale Test Framework running on Red Hat Enterprise Linux. The system includes dedicated programmable power supplies (PSUs) with readout as well as parametric measurement units (PMUs) with per-pad capabilities. The PMUs can source/sink and read any voltage or current to/from the pad and have full control over clock frequencies and edge relationships. The ATE also allows for automated wafer swapping and alignment and can be loaded with up to 25 wafers at a time to allow for consecutive testing.

A custom probe card is attached to the ATE with an adapter ring, and acts as an interface between the ATE and the ABCStar dice during testing. The probe card (Figure 4.7) has an array of probes that touch down on the dice during testing to send and receive signals from the chip and allows measurements to be made.

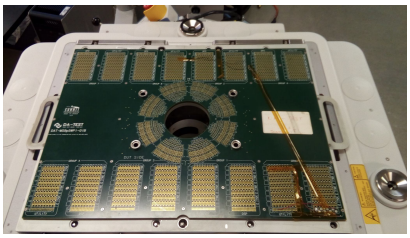
ABCStar wafers are picked up by a robot arm in the tester, placed on a chuck and held in place by vacuum during probing. The chuck is heated above room temperature (usually 30 °C) to ensure a constant temperature throughout probing of the wafer. Alignment is done using a physical notch on one side of the wafer. Machine vision is then used for fine tuning of the alignment to make sure that the pins of the probe card will touch down correctly on the pads of the die for testing. After testing each die, the prober will move on to the next die and use machine vision to confirm proper alignment. The over-travel (how hard the needles push down on the pads) is set by the



(a)



(b)



(c)

Figure 4.6: (a): The probing setup at DA-Integrated with the Electroglas 4090μ ATE and Advantest V93000 Smart Scale. (b) and (c): The ATE opens up to show the PSUs and PMUs inside the tester.

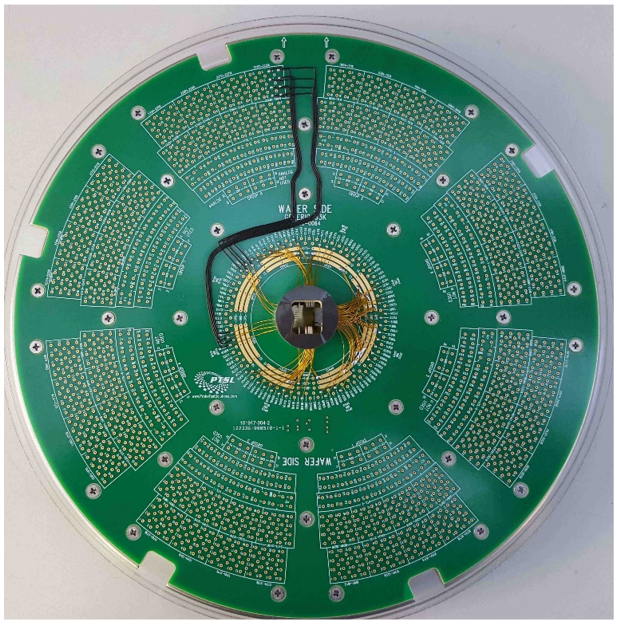


Figure 4.7: A picture of the probe card used to test ABC series ASICs at DA-Integrated.

operators during setup of the prober. The over-travel needs to be sufficiently high to get through the oxidized layer on the pads of the chip, but cannot be too high or there is risk of damaging the chip. To improve contact between the needles and the ASIC, the prober can be set to double-touch on dice before beginning testing. Touching the needles down twice helps to scratch more of the oxidized surface off of the pads to increase the chance of getting good contact. The prober is also set to retest all dice that failed during the initial round of probing. This gives a second chance to establish proper contact with the pads on the chip and pass any ASICs that may have failed due to continuity issues. Cleaning of the prober needles is done routinely before the beginning of each wafer, and after every 50 dice tested to make sure that debris does not prevent proper contact with the pads. Testing will also pause if too many dice fail in a row, and an operator will inspect the needles to ensure this is not caused by contaminated needles or travel issues.

Tests on ABCStar ASIC wafers are implemented by writing C/C++ code that interfaces directly with the proprietary hardware and software of the automated tester. This code is compiled into the test program and executed by the tester. During testing, data is saved in text files and binary test data format which is later converted into CSV format for further analysis.

One of the most important features used in ABCStar testing is *digital sourcing*, which allows test vectors to be dynamically generated in memory and sent to the tester in large chunks. This approach greatly improves efficiency by avoiding repeated small downloads and reducing time switching between the software sending vectors and hardware executing the commands. Additionally, sections of test vectors can be looped to repeat operations, which simplifies test logic. These capabilities ensure that tests are both fast and flexible, supporting the complex requirements of ABCStar chip

validation.

4.2.2 Test Outcomes and ASIC Categorization

ABCStar ASICs being tested at DA and RAL are subject to online and offline cuts to determine if they are functional. Online cuts are less rigoruous and checked while the dice are being tested. If a die fails an online cut, the test sequence is aborted and no further tests are completed on that die. Offline cuts are completed after a full wafer has been probed and more detailed statistics on the operation of each ASIC can be obtained. Each die is then placed into one of four categories based on the outcomes of the tests:

- Category A: Dice that pass all online and offline cuts. These dice are appropriate for use in the ITk.
- Category B: Any dice that have 1-2 bad channels (e.g. high noise, low gain, no response during Pedestal Test), marginal LDO tunings or a high VT50 RMS (see Section 4.3 for test details) are category B due to minor defects. These dice are used in prototype testing and could be used in the ITk if ASIC yields are too low.
- Category X: Dice that fail any other online or offline cuts are rejected.
- Category T: Two passing dice (Category A) from each lot of wafers will be designated for fast TID (x-ray) testing. These dice are used to measure the current-versus-dose curve across lots to monitor process variation.

After a wafer has been probed and offline analysis completed a wafer pick map is generated showing the category for each ABCStar ASIC. An example wafer pick map

can be seen in Figure 4.8. The expectation was to have 90% yield of category A dice for on-wafer probing.

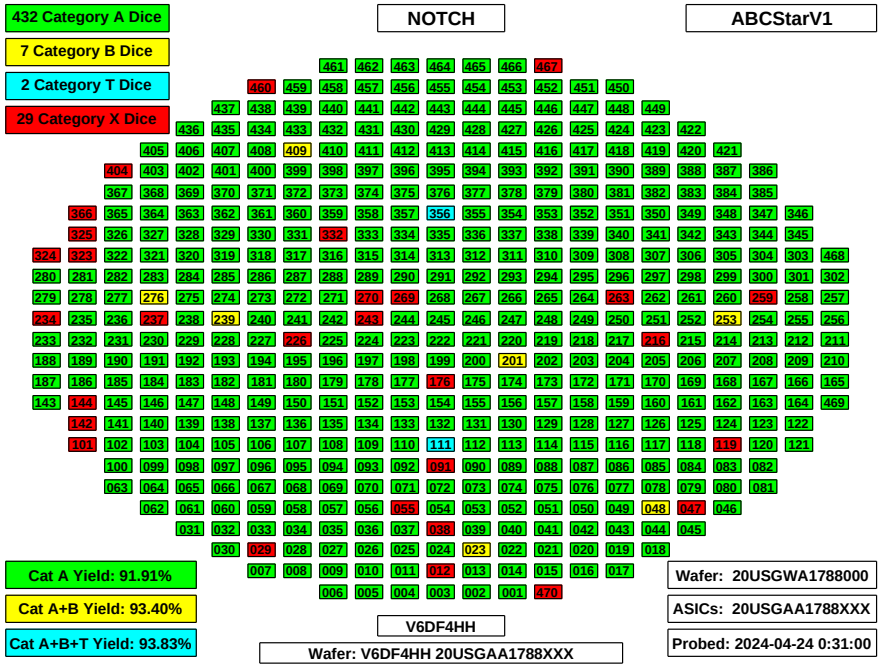


Figure 4.8: ABCStar wafer map showing the final category of all dice after probing for wafer V6DF4HH. The "NOTCH" on the wafer map refers to a physical notch on the wafer that is used to properly orient the wafer. The wafer and ASIC serial number for the ITk Production Database are included.

4.3 Analog Probing

The ABCStar is a mixed-signal chip with both digital and analogue I/O, and on-board power regulators. It has a built-in test mode used to exercise the 256 front-end analogue inputs that will be connected to the strip sensors. This functionality is exercised by a combination of analog and digital inputs and commands. A list of analog pads on the ABCStar can be seen in Table 4.2.

A list of the power related pads used in testing can be seen in Table 4.3. There are

Table 4.2: List of Analogue Pads on the ABCStar.

Pin Name	**Description**
TESTCOM	External connection for discriminator bias
TESTRES	Reference resistance, nominally 909.5 Ω to GNDA
GNDA_local	Sense line for on-chip GNDA potential
AMUXOUT	External analog multiplexer (AMUX) connection for several on-chip voltages
ADCAL	External input to allow on-chip ADC to be calibrated

Table 4.3: List of Power-related Pads on ABCStar

Pin Name	**Description**
AVDD	External/unregulated power to analog LDO
VDDA	Regulated power from analogue LDO to analog circuits on chip
DVDD	External/unregulated power to digital LDO
VDDD	Regulated power from digital LDO to digital circuits on chip
GNDA	Analog ground return
GNDD	Digital ground return
VFUSE	1.50 V power normally, 3.30 V power during eFuse programming
ShuntCtrl	Shunt regulator circuit control input
PowerLow	CMOS levels, pull to DVDD for low power mode

on-board low-dropout (LDO) regulators which power the analog and digital portions of the chip. The LDO regulators adjust the external power provided to the chip down to the nominal voltage for operation (1.20 V for analog, 1.25 V for digital). There is also a controllable shunt circuit, which can sink current from DVDD to GNDD using the ShuntCtrl signal. This was included to allow ABCStar chips to draw a constant amount of current around the TID bump, but should not be required due to pre-irradiation of the chips.

4.3.1 Test Sequence

Tests are organized in larger groups, starting with a series of initial tests to verify basic chip functionality necessary for all subsequent tests. After this, a series of analog tests are performed as outlined below, followed by digital tests outlined in Section 4.4. The sequence described here corresponds to the order tests are completed on the ABCStar at DA.

Initial Tests:

1. Continuity on all individual pads to ensure probe card contact.
2. Shorts tests on individual signal pads to ensure probe card safety.
3. Measure the built in test resistor (TESTRES) by applying voltage and measuring current.
4. Power supply shorts test: measure current on each of the 3 power supplies: external power to analog LDO (AVDD), external power to digital LDO (DVDD) and power for eFuse programming (VFUSE) to ensure there are no short circuits on the chip.
5. Perform ShuntCtrl sweep: record current on DVDD for each of the voltages applied to ShuntCtrl.
6. Measure LDO voltages to determine voltage drop across regulators.
7. Perform the initial power on tests and record currents on the power supplies. These tests are repeated as different chip functionalities are enabled.
8. Test the low power setting and log currents on the power supplies.

9. Test the initial SLVS output levels.

10. Test for presence and periodicity of HPR packets. An HPR packet should be sent 500 µsec after LCB lock is achieved or after a hardware reset and every 1 msec afterwards.

11. Test chip ID pins used by reading values in HPR packets. ID pins are used by the HCCStar to distinguish connected ABCStar ASICs.

Verify LCB_IN Communication: This tests the basic functionality of the LCB_IN pads of the chip, including locking and losing communications lock on the input, disabling the HPR packets, and performing an initial configuration of the die under test.

1. Test LCB locking by sending and maintaining a sequence of Idle Frames.

2. Test LCB Lock-Loss: When lock is lost on LCB_IN, an HPR packet should be generated indicating lock was lost.

3. Trigger Immediate HPR Output: Perform a hardware reset, establish LCB lock, perform a register write and then lose LCB lock. HPR packets should be received after 500 µsec, after LCB lock, after the register write and after LCB lock-loss.

4. Stop regular HPR output: It tests part of the HPR suppression functionality by disabling the periodic output of the HPR packets. This is important for remaining tests as the asynchronous output of an HPR packet could affect the expected pattern from a deterministic test ran on the chip.

5. Stop HPR output on LCB_IN lock status change: This tests that the output of a spontaneous HPR packet can be suppressed on changes to the lock/lock-loss status of LCB_IN.

6. Test control of single HPR at startup: This tests a special case for the HPR packet output. At least one initial HPR packet will be sent 500 µsec after a hardware reset, a Register Reset, or a Logical Reset followed by the usual sending of HPR packets. But if StopHPR is set to 1 before the 500 µsec elapses, only a single HPR packet will be sent and after that, the periodic packets will not be sent.

7. Stop HPR output: This test verifies that the HPR can be turned off completely, and then restarted.

8. Initial configuration. In this test, various registers are set to nominal values and the change in analog current is measured.

9. Confirm that the initial configuration can be read back and is correct.

Sweep and Tune LDOs: Tune the digital and analog LDO regulators on the chip to the desired settings for the remainder of the tests. the digital LDO is tuned to 1.25 V and the analog LDO is tuned to 1.20 V for all remaining tests.

Burn serial numbers into the eFuses: Program the 24-bit eFuse used to uniquely identify the wafer and chip.

Analog Circuitry: Perform analog measurements as listed below using both a direct reading with an external ADC and a reading through the on-chip ADC. For the on-chip ADC's reading to be useful, it must first be calibrated.

1. Calibrate internal ADC. Testing the ADC involves applying a voltage externally to the ADCAL pad and then reading the value of the on-chip ADC using a register read command to collect sufficient data from the ADC. In a first pass

each ADC gain range is scanned to find the range with the maximum sensitivity across a 0.9 V range. A second pass is then performed to make an accurate determination of the gain in this range, for example, any offset or gain errors. The measurements from the ADC are then calibrated in software by using the calculated offset and gain corrections.

2. Sweep and tune TESTCOM DAC. This test scans and tunes one of the ABCStar's DACs in the digital power domain.

3. Sweep and tune AMUXOUT DACs: Measure the band gap voltage of the chip. Sweep and tune the internal DAC converters referenced to this internal band gap circuit to compensate for variations in the bandgap voltage between chips.

Analog Front-End Tests: The characterization of the ABCStar front-end is primarily carried out using internal charge injection circuitry built into each chip. The calibration charge is injected to the front end amplifier input through on-chip, per channel calibration capacitors.

1. Pedestal Trimming: Reduce response variations across the chip by channel-by-channel configuration of an offset applied to the chip pedestal. This is determined by adjusting the threshold for each channel when no charge is injected. Settings are selected such that the channel-by-channel response is as close to uniform as possible. The response for three channels can be seen in Figure 4.9.

2. Strobe Delay Tuning: The default timing (strobe delay) between charge injection and readout is varied using the in-built delay which is applied to the generation of the charge injection pulse. The rising and falling edges of the window for which the pulse is seen are determined, and the strobe delay is set to a value

corresponding to a point 57% of the way between these two edges (this fraction has previously been determined to correspond to the point at which the maximum gain is observed). A sample strobe delay response can be seen in Figure 4.10.

3. 3-Point Gain Measurement: Threshold scans are performed at three different charges, 0.5, 1.0 and 1.5 fC. Each threshold scan is fitted with a complementary error function (S-curve) from which the median and width can be extracted. Analysis of the three scans enables calculation of the gain and noise of the front end circuitry. A response curve plot from a chip can be seen in Figure 4.11.

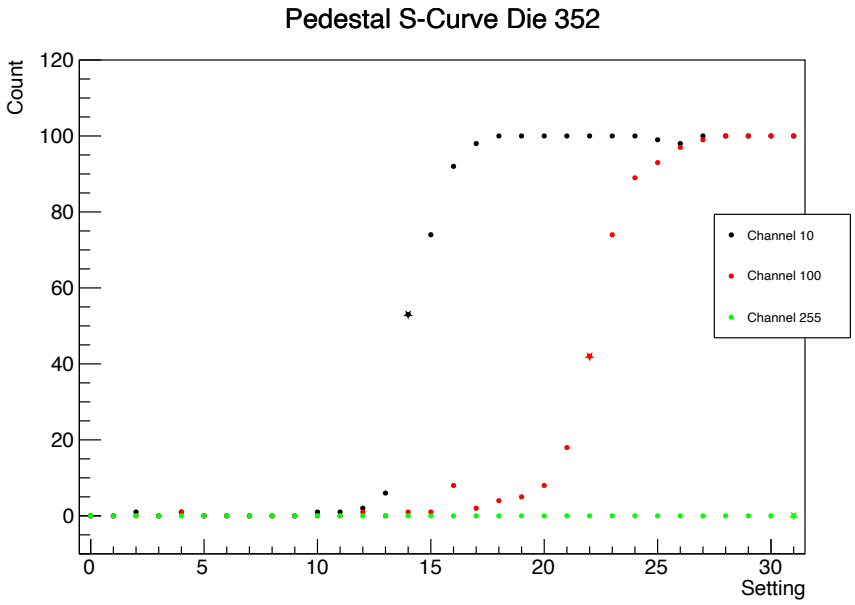


Figure 4.9: Pedestal trimming plot that shows the number of hits out of 100 on each channel (count) for each threshold setting. The settings selected for each channel where the count is closest to 50 are marked with a star. Channels 10 and 100 are good channels and channel 255 has no response for any threshold and would be classified as untrimmable.

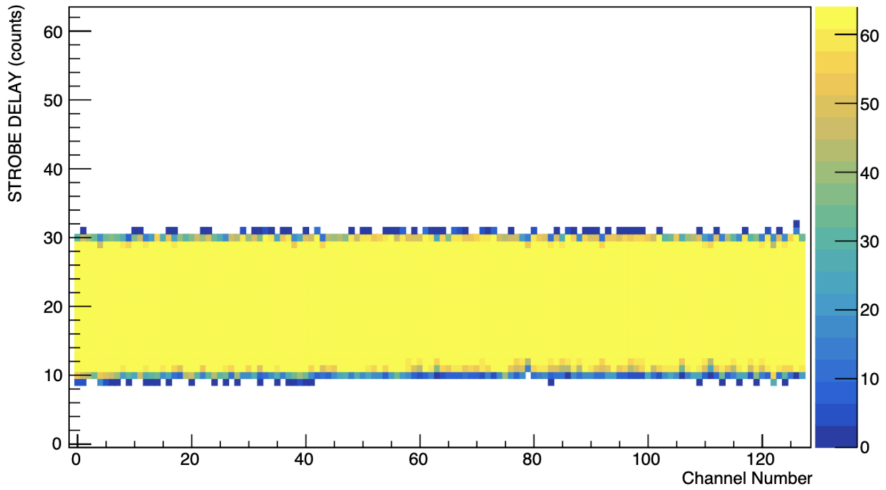


Figure 4.10: A strobe delay plot showing the number of hits for 128 channels on a die for each strobe delay setting. The setting chosen corresponds to the point 57% between the rising and falling edges of the counts where all channels are showing the maximum number of hits.

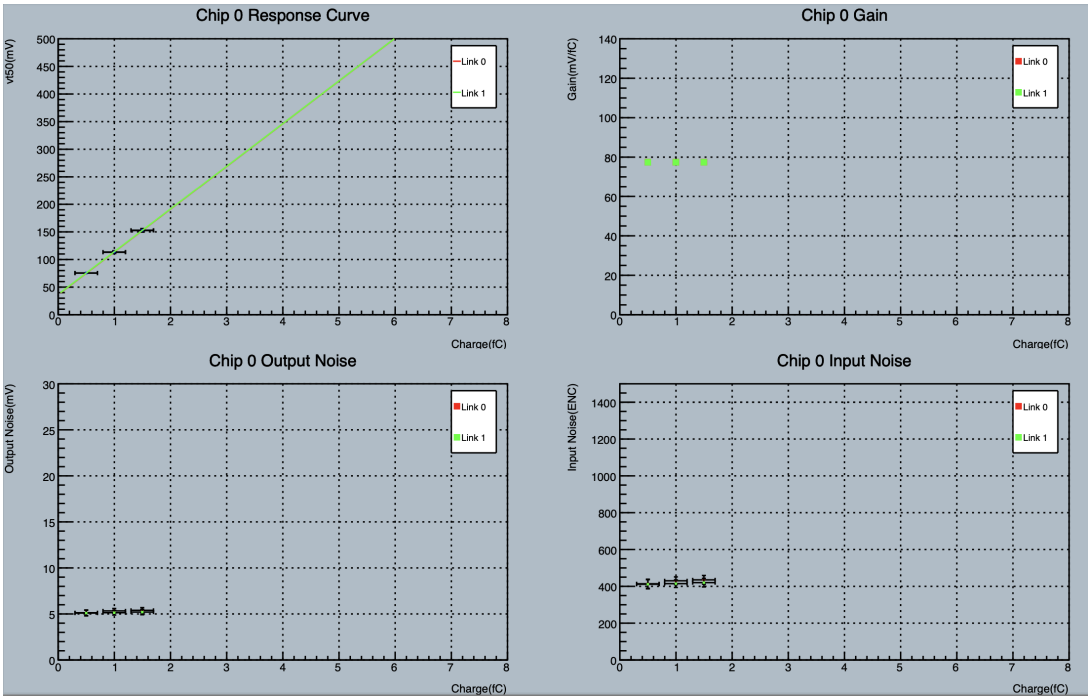


Figure 4.11: A sample 3-point gain measurement from an ABCStar chip. Three charges are injected and used to determine the gain and noise for each channel.

4.3.2 VT50 RMS Test Failures

One measurement during the 3-point gain test is the threshold voltage at which a channel has a 50% chance of registering a hit (VT50). The pedestal trimming before the 3-point gain test aims to make the VT50 as uniform as possible between all 256 channels on a chip. There is a cut on the VT50 RMS across all 256 channels (VT50 RMS $< 3.4\,\mathrm{mV}$) to ensure that the pedestal trimming test works correctly. Any chips that do not pass the VT50 RMS test are category B dice.

The VT50 RMS test was originally implemented at RAL to detect contact issues during probing. During the pedestal test no charge is injected, so the only stimulus is from noise on the channels. If there is poor contact, the response curves may not have the typical s-curve seen in Figure 4.9. This can result in channels being tuned to the wrong threshold value, meaning there will not be a uniform response across channels and the VT50 RMS in the 3-point gain test will be larger.

During production probing at DA, there were an abnormally large amount of chips failing the VT50 RMS test across different wafers from several lots. Figure 4.12 shows a wafer with many VT50 RMS failures. There also appears to be a striped pattern for the VT50 RMS failures, where it was more likely for dice to fail on every second column of dice. The prober alternates upward and downward sweeps on dice relative to the orientation seen in Figure 4.12, hinting at a directional bias on the prober.

To ensure that the directional bias was not caused by misalignment of needles on the pads when the prober was moving in a certain direction, the positioning of the needles was inspected. For example, if the prober was moving too far when switching between dice, eventually the needles would not contact the correct pads. After inspection it

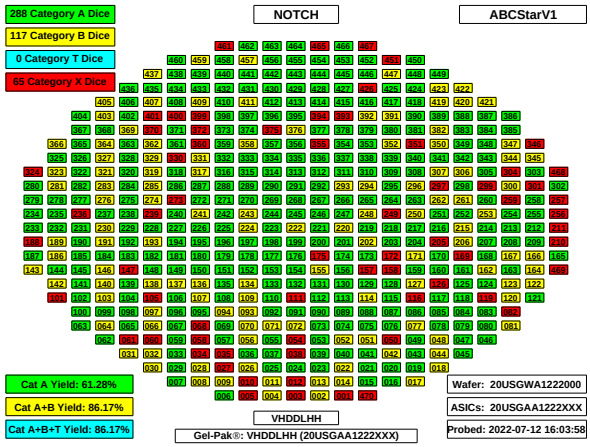


Figure 4.12: A wafer map for a wafer with many VT50 RMS failures. Failing the VT50 RMS test results in dice being classified as category B. There are more dice failing the VT50 RMS test in every second column.

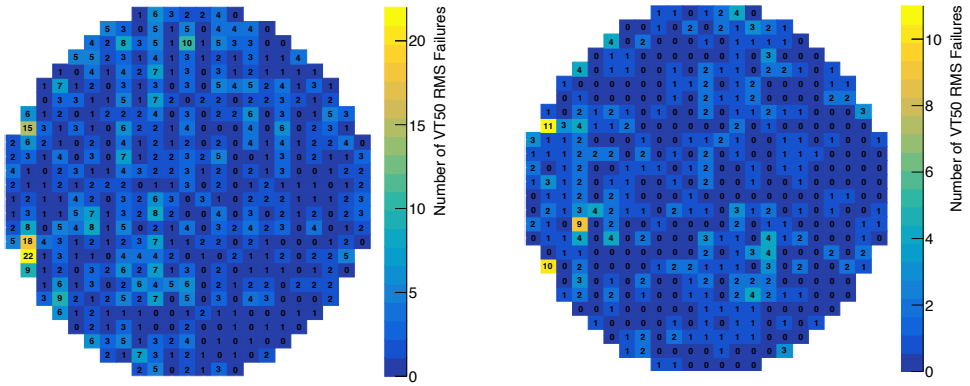
was determined that the needles were touching down properly on the pads for all of the die locations. Other analog and digital tests were working correctly and initial continuity tests must pass on each die, so misalignment is unlikely to be the issue.

In order to try and reduce the number of VT50 RMS failures at DA, steps were taken to try and improve the reliability of contact with the dice during testings. The first steps involved the touching down of the needles on the pads of the chip. The effect of over-travel of the needles on the VT50 RMS for a chip was investigated. The pedestal and 3-point gain tests were conducted and a measurement of the VT50 RMS was recorded, then the over-travel was increased and the process repeated. Unfortunately, there was no improvement in VT50 RMS measurements when increasing the over-travel. A larger over-travel will also increase the wear on the probe needles and risk damaging the chip so this was not pursued further. The tester can also be programmed to double-touch on dice before beginning the test sequence to scratch off more of the oxidized surface and increase the chance of good contact with the pads. Using the double-touch feature did not appear to reduce the amount of VT50 RMS failures

either.

Another avenue investigated, was related to the cleaning of the needles during probing. At regular intervals, the needles will touch down on a cleaning pad to remove any dirt or debris that could prevent proper contact with the pads. There are different options for cleaning pads that can be used for different probe cards and different projects. An alternative cleaning pad was tested that was coarser and could possibly create larger grooves in the needles to better penetrate through the oxide layer on pads and improve contact. Testing while cleaning with the coarser cleaning pad did not appear to improve test results, so the original cleaning pads continued to be used. Another option is to increase the frequency of needle cleaning during probing. The trade off with additional needle cleaning is that it will increase the wear and tear on the needles and shorten the lifetime of the probe card. The needles are cleaned at specific intervals during testing (unless there are issues with continuity and an operator manually cleans the needles), but no correlation was seen between the location of the needle cleaning and the location of dice failing the VT50 RMS test. Regardless, the frequency of needle cleaning was increased slightly and the needles were programmed to clean automatically before the start of a new wafer as the impact on the lifetime of the probe card is small.

Dice failing the VT50 RMS test at RAL would later be reprobed to try to improve the contact and the pedestal test would be repeated to improve the response across all channels. This is the only reliable method for improving the VT50 RMS results at DA. The needles are taken off of the ABCStar and touched back down to retest the die. The probing and wafer swapping was more automated at DA, so dice were initially not reprobed unless there were continuous strings of failures indicating an issue with the prober. Due to the high rate of dice failing the VT50 RMS test during



(a) Before retesting implemented. (b) After retesting implemented.

Figure 4.13: VT50 RMS failures across wafers probed before (left) and after (right) implementing a second testing of all failing dice during on-wafer probing. The initial nine lots had an average of 112 VT50 RMS failures per lot and the last five lots had an average of 75 failures per lot.

production probing it was decided that the prober should be programmed to retest dice upon failure. This is done for all dice failing any test (not just VT50 RMS), and after the entire wafer has been probed initially, any failing dice are reprobed a second time in an attempt to get passing results. This had a significant improvement on the yields of wafers and reduced the number of VT50 RMS failures. The improvement can be seen in Figure 4.13, where the amount of VT50 RMS failures per lot decreased by 33% after implementing retesting.

4.4 Digital Probing

A suite of digital tests are implemented to exercise the digital logic of ABCStar chips during on-wafer probing, such as LCB signalling, HPR, memory and readout capabilities. Signals are sent from the prober to the chips to perform different operations, and the output signals from the chips are recorded and compared with

expectations. The output from the chip must exactly match the expected pattern for the test to be considered a pass. Some signals are masked out that we would not expect to match between all ASICs, such as the SEU count values. All digital tests are conducted with a digital voltage of VDDD=1.25 V with a 49/51 duty cycle.

The digital tests are derived from a series of tests that were performed on a simulation of the ABCStar chip using Cadence Xcelium software. The tests are adapted for use on physical chips to ensure that all input signals can be sent from the pads connected to the prober. Cadence Xcelium is then used to run the test on a simulation of the ABCStar and save the input commands and expected output from the chip in Vector Change Dump (VCD) format. These VCD files are provided to the engineers at DA, and converted into a format that can run on the prober to inject the proper signals into each die during testing and compare the output to the expected values.

One further change is required to run the tests at the desired voltage for each ABCStar on the wafer. In order to set the digital voltage to 1.25 V for each test, the digital LDO had to be set to the proper value obtained during analog testing for that specific die. All digital tests began with a reset that resets the LDOs, and each die will have slightly different tune values to give 1.25 V. Rather than loading several versions of each digital test corresponding to the desired LDO tune values, another functionality of the probing software at DA is used to adjust the digital test as it was ran on each die. Every time that the chip is reset or an LDO value is explicitly changed in a digital test, a register write is sent to set the LDO back to the tune value corresponding to 1.25 V. This requires finding the location of all resets and LDO register writes for each test and inserting a placeholder command that is updated with the appropriate register write for that specific chip during testing.

A brief description of each digital test and what digital logic it exercises is given below:

- A02 – Readout in Test Mode 11: This is a general randomized readout test. It interleaves a variety of chip operations including triggering and readout of hits, various resets and reading and writing of chip registers. The chip readout is configured in Test Mode 11, meaning that the readout pipeline is filled with a time-varying pattern generated from the Bunch Crossing ID counter (BCID counter). As this test makes intensive use of the RAMs in the readout pipeline, it is the digital test most able to expose any timing issues during RAM operation.

- A11 – SEU counter and SEU latches: Tests the disabling of each of the three clock trees in turn, verifying that the Single Event Upset (SEU) latches are not raised and that the SEU counter does not increment. During this test, as two out of three clock trees are enabled, two out of three voters in each piece of Triple Modular Redundancy (TMR) circuitry should obtain the same results, so no errors should occur.

- A13 – Data idle: Verifies data integrity and communication reliability of the ABCStar when it is sending idle data packets.

- A16 – Readout in Test Mode 01 (static): The chip readout is configured in Test Mode 01, which is the static test mode where the contents of the mask register are used to supply test values to readout pipeline. Different patterns are injected and pass through the readout pipeline to ensure proper functionality.

- A17 – Readout in Test Mode 10 (pulse): The chip readout is configured in Test Mode 10 with option 1 selected. This is the pulsed test mode, where the contents of the mask register are used to supply the test value to the readout pipeline

for a single clock period, otherwise the contents are zero. Different patterns are injected and pass through the readout pipeline to ensure proper functionality.

- A22 – LCB decoder: This test verifies the LCB decoder's 8-bit to 6-bit (8b6b) decoding, including valid and invalid 8-bit symbols.

- A23 – LCB lock: This test verifies the LCB frame synchronizer for gain and loss of LCB lock.

- A24 – LCB single frame elements: This test verifies that the LCB decoder for single frame protocol elements: Idle frames, L0A/BCR frames and Fast Commands.

- A25 – LCB slow commands: This test verifies that the LCB protocol circuit handles slow command sequences correctly.

- A26 – LCB slow command IDs: This test verifies the operation of chip addressing during slow commands, varying the ABC and HCC IDs.

- A27 – LCB slow command error counter and flag: This test verifies counter saturation and clearing behaviour for the LCB slow command error counter and flag.

- A28 – LCB decoder error counter and flags: This test verifies operation of LCB Decoder error flags, error counter, error threshold, counter clear signal and self-reset.

- A33 – High Priority Register (HPR) packets: This test verifies operation of the autonomous heartbeat packets (High Priority Register packets).

- A35 – ABCStarV0 compatibility mode: This test verifies operation of the ABCStarV0 compatibility mode. This test verifies that if this backwards-compatibility mode is switched on, that the readout packets conform to the format generated by the previous version of the ABCStar, the ABCStarV0. It also verifies the operation of the Single Event Upset (SEU) counter.

- A999 – Automated test for High Priority Register packets: This test configures the chip to issue a first autonomous heartbeat (High Priority Register) packet. It verifies that specific fixed bits of the HPR packet are as expected.

- P03 – Readout with pulse patterns: Inject hit pulses with varying patterns and check cluster responses.

- P99 – Test operation of readout memories (RAMs): This test verifies the random-access memories (RAMs) used in the readout path. The test uses the mask function to fill the memories with ones, then varies the latency to read out all bits of the RAMs. As this test makes intensive use of the RAMs in the readout pipeline, after test A02, it is the test most able to expose any timing issues during RAM operation.

Each digital test is performed once on each chip with the exception of A02 and P99.

A02 is ran a total of 15 times on each chip with three different digital voltages for five different duty cycles. The three testing voltages are VDDD=1.25 V and the two voltages for one and two settings higher on LDOD, corresponding to approximately VDDD=1.225 V and VDDD=1.20 V. The five duty cycles ran are 48/52, 49/51, 50/50, 51/49 and 52/48. Each die must pass for VDDD=1.25 V with a duty cycle of 49/51 and higher to pass the A02 test. The additional test results are saved to determine

which chips could be set to lower voltages if required during their operation in the ITk.

P99 is executed four times at VDDD=1.25 V with different 40 MHz and 160 MHz clock phases disabled to ensure proper functionality of the clock triplication. P99 is also performed an additional time at VDDD=1.15 V (all clock phases enabled) to ensure functionality of the chip at lower voltages. The P99 test must pass for all five conditions in order for the chip to be considered functional.

4.5 ABCStar SRAM Issues

During the start of production probing for the ABCStar, one lot (lot 0HZJB00000) of wafers was probed that had very low category A yields. The average category A yield was only 31%, compared to the 86% average across all production lots. The main reason for failures across this lot was due to failures from digital tests A02, P03 and P99. When production probing began, all digital tests were being conducted at a digital voltage of 1.20 V (1.10 V for P99 with all clock phases enabled) and a duty cycle of 50/50, the intended operating conditions for ABCStar ASICs within the ITk at the time.

When this lot with extremely low yields was discovered, one-third of the ABCStar wafers had already been manufactured and delivered. The foundry that manufactures the wafers investigated and confirmed that the wafers were within their process variation limits before they were shipped. This meant that the root cause of the issue needed to be determined as fast as possible, and mitigation methods needed to be explored to ensure that ABCStar chips would be available in time for installation

in the ITk and that they would function properly throughout their lifetime in the detector.

4.5.1 Identification of SRAM Issue

ASICs will fail if a mismatch between the output and the expected pattern is detected for any digital test, but this does not give any information on which bits were flipped to cause the errors. The number or location of errors is not recorded during wafer probing. In order to determine what was going wrong with the digital logic of these failing chips, the output was captured and investigated to see where errors were occurring in the test. Repeating test P99 on a problematic die multiple times resulted in intermittent passing of the test, with a different number and location of bit flips. Figure 4.14 shows part of the output of a particular die for test P99 captured at DA. The region with black scribbles shows a location where the average output from the die was not stable, and intermittent failures were occurring. The A02, P03 and P99 tests all exhibited this similar behaviour, where the number and location of errors was not consistent if the tests were repeated, and they could have anywhere from a couple of bit flips to hundreds in a single test.

Investigation also revealed that increasing the digital voltage during testing would usually fix the errors and result in a pass on most dice. For most of the ABCStar test development phase, the digital tests had been conducted at the supply voltage of 1.50 V. Increasing the digital voltage from 1.20 V back to 1.50 V greatly increased the yields on these problematic wafers and resulting in consistent passes for the A02, P03 and P99 tests. It was also found that increasing the duty cycle supplied to the chip could recover failing dice. DA has the capability to make *Shmoo plots*, which are a

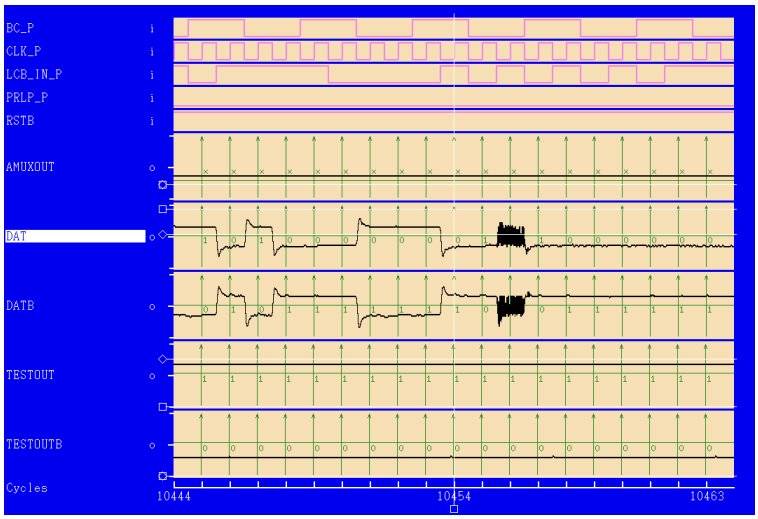


Figure 4.14: Waveform capture of an ABCStar ASIC for digital test P99 at VDDD=1.10 V. The arrows show the timing at which the output signal from the die is sampled by the wafer tester and the green numbers indicate the expected output (0/1) on the DAT/DATB pins. The black waveform overlaid indicates the average output value for a chip running P99 after executing the test multiple times. The black scribbles indicate an area where the output from the chip is not consistent between runs of P99 and therefore does not always match the expected output to pass the test.

graphical representation of test results in a multi-dimensional parameter space. The pass/fail result of the A02 test on a chip from the bad lot as a function of the digital voltage and supplied duty cycle can be seen in Figure 4.15. It can be seen that this chip would pass A02 at 1.2 V with a 50/50 duty cycle, but increasing the LDO offset by +1 (VDDD=1.175 V) would result in a failure. Similarly, running the test at 1.2 V, but decreasing the duty cycle to 48/52 would result in a failure. This indicated that many of the chips that were passing the A02 test were just marginally passing and a small decrease in voltage or duty cycle could cause hit loss to occur.

Analysis of the output from the A02 test on faulty chips revealed that when the physics data was read out, there were hits missing in the physics packets compared to what was injected into the chip. There were no errors found in any other type of readout from the chip, such as register reads or HPR readout, and the formatting of the data

Figure 4.15: A Shmoo plot showing the pass/fail result for the A02 test as a function of digital voltage and duty cycle. The x-axis shows the LDO offset from 1.20 V, where each step is a change of 25 mV and a positive value indicates a decrease in supplied voltage. The y-axis shows the absolute change in the rise time of the duty cycle. A positive value results in a higher duty cycle.

being read out from the chips was correct. The A02 test extensively exercises the static RAM (SRAM) of the ABCStar chip. These findings led to the conclusion that the bit flips causing hit losses were occurring at the on-chip SRAMs or the SRAM interfaces. Since problematic dice could be recovered at higher voltages and ASICs from other lots could pass at the lower voltages, this implied a process-related dependency during manufacturing, rather than a logic design error for the ABCStar.

The data pipeline for the A02 test involves the mask and edge detection circuit writing data to the L0_Buffer, which then copies this data onto the event buffer where it will be read out to the cluster finder if a request is received (Figure 4.3). Both the L0_Buffer and event buffer are physically implemented on the ABCStar with SRAM blocks. Figure 4.16 shows the pattern of missing hits in the SRAM blocks and their physical location on an ABCStar chip for the A02 test. Most of the SRAM blocks worked

as expected, which suggested that the blocks themselves were not faulty. Instead, it was a timing issue with the SRAM blocks that controlled read and write access to them causing the problem. It was determined that the missing hits were introduced either in the interface of the mask and edge detection circuit to the L0_Buffer, or the interface of the L0_Buffer to the Event Buffer.

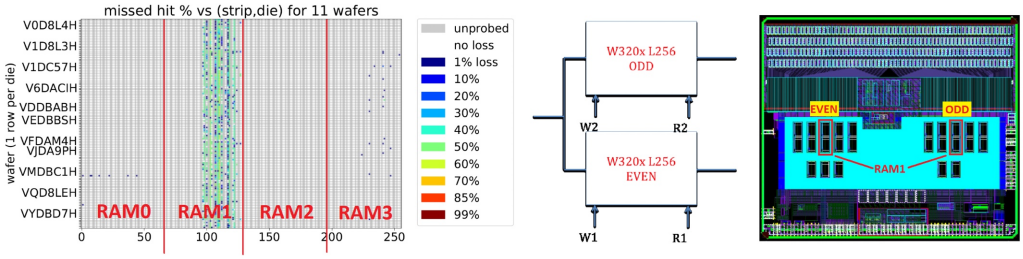


Figure 4.16: Missing hit data on each SRAM block in the ABCStar and their physical location on the chip for the A02 test. A majority of hit losses were occurring on RAM1. Figure taken from Ref. [70].

The timing issue causes a signal delay compared to the chips falling clock edge, which does not give enough time for the SRAM to read in the correct value. Increasing the digital voltage speeds up the transistors, allowing the correct value to be read into the SRAM before the falling clock edge. Similarly, an increased duty cycle would increase the available time before the falling clock edge to allow the proper value to be read in.

This timing issue was introduced due to an issue with the timing models that were used for the SRAM block when designing the chip. The glue logic[2] for the ABCStar SRAMs was synthesized to meet the requirements of this timing model, but the model was incorrect which led to marginal timing on physical chips. The SRAM design had been successfully used in other ASICs for previous projects and was considered "silicon-proven", which led to the timing models not being recomputed for the ABCStarV1.

[2]Simple digital circuitry that connects functional blocks in a chip, adapting interfaces and timing

4.5.2 Mitigation Methods

If the ABCStar chips with SRAM issues were used in the ITk, then physical hits caused by particles passing through the ITk strip sensors could be lost in the ABCStar readout chips, resulting in the loss of valuable physics data. At the time of discovery, 250 of the 750 required ABCStar wafers had already been manufactured, so it was vital to search for possible methods to prevent hit loss in the detector. It is ideal to prevent the use of any chips that could result in hit loss, while also ensuring that the wafer yield remained high enough to obtain enough working ABCStar chips to build the ITk strip detector. It was also preferable to use the wafers that had already been manufactured, to prevent the prohibitive costs of having them all replaced. A redesign of the ABCStar would be extremely costly and result in a significant delay for the progress of the ITk upgrade. Alternatives to a redesign included changes to the operating conditions of the ABCStar in the ITk or changes to the wafer manufacturing process. The first method was preferred as it would allow chips that had already been fully manufactured to be used throughout the detector.

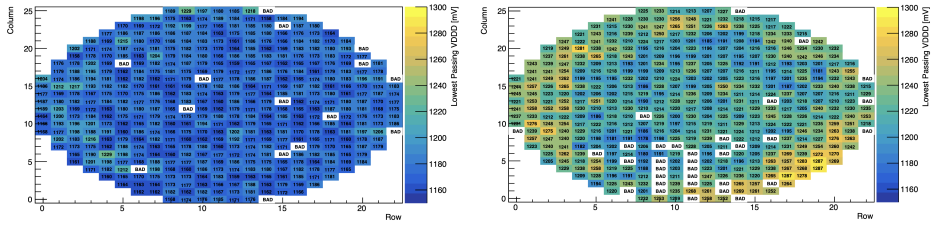
Increasing the digital voltage and increasing the duty cycle supplied to the chip both improve digital performance and were operating conditions that could be adjusted. The effect of radiation and cooling on the SRAM issue also needed to be understood before installation in the ITk.

4.5.2.1 Effect of Voltage on SRAM Issue

Initial tests had shown that operating at a higher core voltage would result in faster transistors and could remove the timing issues in the SRAM. To investigate the

relationship between the digital voltage and the SRAM performance, multiple wafers were probed at DA-Integrated, running the A02 test at a range of regulated digital voltages to see where the hit loss disappeared. The digital LDO DAC (LDOD) has 17 settings, and each setting changes the voltage by about 25 mV on average. A02 failures were being seen for lot 0HZJB00000 at a digital voltage of 1.20 V. Due to limitations for powering the ITk detector and thermo-electric modeling, it was not desired to set the digital voltage of the chips any higher than 1.30 V. Adjusting the regulated analog voltage was also investigated, but this was found to have no effect on the performance of the SRAMs.

The lowest passing digital voltage for the A02 test for two wafers can be seen in Figure 4.17. Wafer VTD8LBH was from production lot 0HZJB00000 which had poor digital performance and the lowest passing voltage is much higher on average than wafer V7DDKAH from a different lot. There is also a radial pattern visible, where chips towards the edges of the wafer required a higher digital voltage to pass A02.



(a) Wafer V7DDKAH (Good lot). (b) Wafer VTD8LBH (Bad lot).

Figure 4.17: Lowest passing digital voltage for A02 test with 50/50 duty cycle. Wafer VTD8LBH is from lot 0HZJB00000 which had poor digital performance. Dice labelled "Bad" either did not pass A02 at any voltage or failed an earlier test and A02 was not conducted.

Yields for wafers from various lots with different digital voltages for the A02 test can be seen in Table 4.4. For all production lots, increasing the digital voltage by one step on LDOD (∼25 mV) results in a much higher yield of passing chips. Increasing the

Table 4.4: Category A+B yields after applying all analog and digital test cuts for various wafers with different digital voltages for the A02 digital test with a 50/50 duty cycle.

Lot	Wafer	1.175 V	1.200 V	1.225 V	1.250 V
12YVB00000	VZDBAQH	56.38%	82.55%	87.23%	87.23%
144YB00000	VHDDLHH	79.57%	85.96%	86.17%	86.17%
144YB00000	VJDDLFH	65.11%	79.79%	81.70%	81.70%
183KB00000	VMDCEZH	88.30%	88.72%	88.72%	88.72%
183KB00000	VBDCDTH	90.43%	91.70%	91.70%	91.70%
15M2B00000	V7DDKAH	84.26%	93.62%	94.68%	94.68%
15M2B00000	VCDDJNH	79.57%	90.21%	91.49%	91.49%
2YEWB020000	V2DCCKH	57.02%	84.89%	88.72%	88.94%
2YEWB020000	V5DCFYH	67.45%	92.77%	94.04%	94.04%
0HZJB00000	VVD8HBH	2.34%	38.30%	73.83%	82.98%
0HZJB00000	VTD8LBH	2.34%	38.09%	74.26%	80.43%

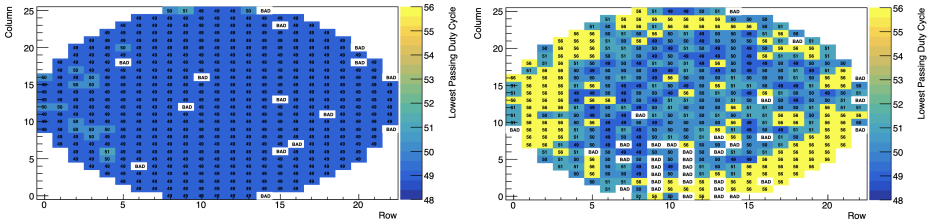
voltage any higher only marginally decreases the number of chips failing A02 for most wafers. The average passing voltage for the first nine wafers in Table 4.4 is 1.179 V and the average passing voltage for the last two wafers from the poor yielding lot is 1.225 V. Increasing the digital voltage from 1.200 V to 1.250 V would increase the average yield of category A chips from 78.78% to 88.01%, much closer to the expected wafer yield.

4.5.2.2 Effect of Duty Cycle on SRAM Issue

Another method for reducing hit loss is increasing the duty cycle that is supplied to the ABCStar chip. The duty cycle being supplied to the ABCStar by the HCCStar was intended to be a 50/50 duty cycle. This meant that the clock being supplied was high 50% and low for 50%. Increasing the high time of the duty cycle gives more time for the hits to be read in the SRAM and reduced the chance of hit loss occurring. In

order to investigate the effect of the supplied duty cycle on the digital performance of the chip, a range of duty cycles were applied while running the A02 test.

The lowest passing duty cycle for the A02 test for two wafers can be seen in Figure 4.18. Wafer VTD8LBH was from production lot 0HZJB00000, which had poor digital performance and the lowest passing duty cycle is much higher on average than wafer V7DDKAH from a different lot. There is a similar radial pattern as seen with the lowest passing voltage, where chips towards the edges of the wafer required a higher duty cycle to pass the A02 test.



(a) Wafer V7DDKAH (Good lot). (b) Wafer VTD8LBH (Bad lot).

Figure 4.18: Lowest passing duty cycle for A02 test with 1.20 V digital voltage. Wafer VTD8LBH is from lot 0HZJB00000 which had poor digital performance. Dice labeled "Bad" either did not pass A02 at any duty cycle or failed an earlier test and A02 was not conducted.

Yields for wafers from various lots with different duty cycles for the A02 test can be seen in Table 4.5. For all production lots, increasing the duty cycle increases the yield of passing chips. There is a significant increase in yield for the two wafers from lot 0HZJB00000 when increasing the duty cycle by just 1%. Increasing the duty cycle from to 49/51 to 51/49 would increase the average yield from 70.70% to 83.91%, much closer to the expected wafer yield.

It was important to understand the effect of small changes in the duty cycle as the HCCStar chips connected to the ABCStar could have variations in the duty cycle of up to 2%. The duty cycle being supplied depended on the coarse and fine delay

Table 4.5: Category A+B yields after applying all analog and digital test cuts for various wafers with different duty cycles for the A02 digital test at 1.200 V.

Lot	Wafername	49/51	50/50	51/49
12YVB00000	VZDBAQH	73.62%	82.55%	86.38%
144YV00000	VHDDLHH	84.47%	85.96%	86.17%
144YV00000	VJDDLFH	75.11%	79.79%	81.49%
183KB00000	VMDCEZH	88.72%	88.72%	88.72%
183KB00000	VBDCDTH	91.70%	91.70%	91.70%
15M2B00000	V7DDKAH	90.85%	93.62%	94.47%
15M2B00000	VCDDJNH	87.02%	90.21%	91.49%
2YEWB020000	V2DCCKH	75.32%	84.89%	87.87%
2YEWB020000	V5DCFYH	84.68%	92.77%	94.04%
0HZJB00000	VVD8HBH	12.13%	38.30%	61.06%
0HZJB00000	VTD8LBH	14.04%	38.09%	59.57%

settings set on the HCCStar as seen in Figure 4.19, meaning that the duty cycle supplied could vary during its operation in the ITk. The duty cycle was always 50/50 or lower for the HCCStar chips tested. For a wafer with poor digital performance, a change in the duty cycle from 51/49 down to 49/51 could reduce the yield by up to 60%.

4.5.2.3 Effects of Temperature and Radiation on SRAM Issue

ABCStar chips will be cooled and irradiated during operation in the ITk, so the effects of these two parameters on the SRAM performance of the ABCStar needed to be investigated. Previous testing on cooling and irradiation had already been conducted on the ABCStar, so these tests were solely focused on how these factors affected the A02 test outcome.

Radiation testing was done by irradiating ABCStar ASICs to different TIDs while applying different duty cycles and checking the lowest passing digital voltage for the

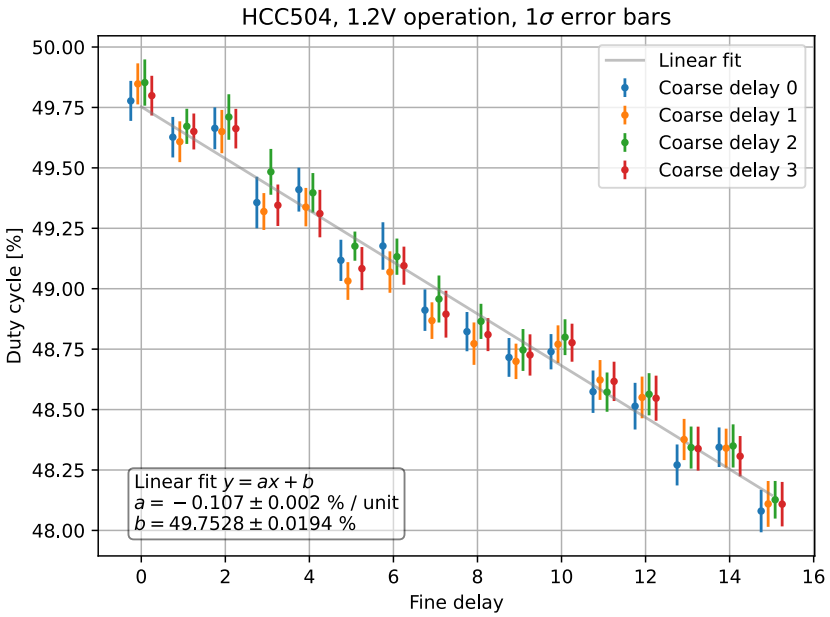


Figure 4.19: HCCStar duty cycle supplied to the ABCStar for different coarse and fine delay settings for a single chip. Figure taken from Ref. [70].

A02 test. Figure 4.20 illustrates that the A02 test can pass on chips with a 50/50 duty cycle at 1.25 V after being irradiated to 100 Mrad. On average, an increase in digital voltage of 20-30 mV was required for the A02 test to pass after irradiation.

After irradiation, the same ABCStar chips were tested in a temperature chamber to determine the failing voltage for A02 from 20 °C down to −40 °C. Cooling the ASICs results in faster transistors, which should increase the SRAM performance. Figure 4.21 shows that cooling the ABCStar chips can reduce the required digital voltage by 30-60 mV. As a result, the cooling effects should negate the decrease in performance from irradiation effects in the ITk and possibly increase the margin of operation without hit loss.

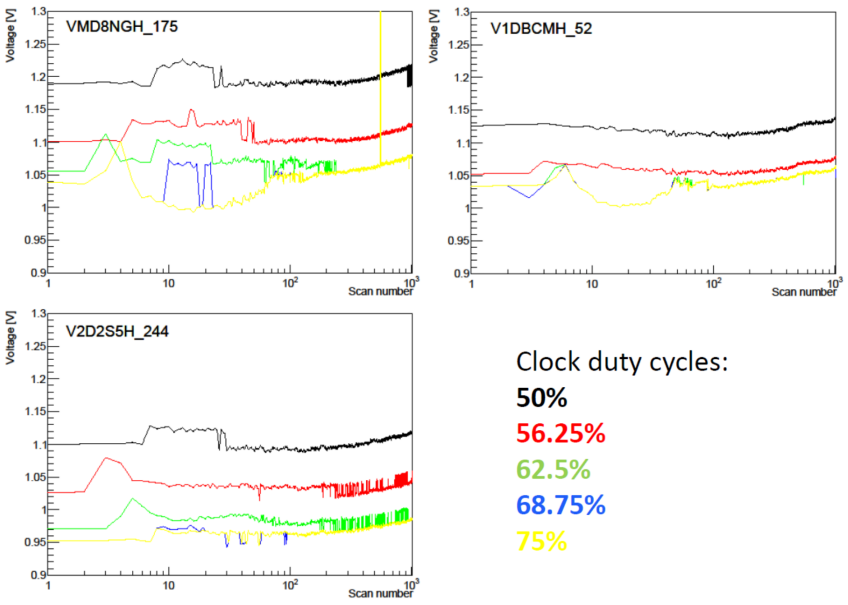


Figure 4.20: The lowest passing voltage for the A02 test for different supplied duty cycles while irradiated to 100 Mrad. The voltage was lowered until there were more than 95 bit flips, which was greater than the expected upper limit of 92 bit errors in the pattern seen due to random bit flip errors from the radiation.

4.5.2.4 Experimental Wafers

The foundry manufacturing the ABCStar wafers found that there was a correlation between the process variations during manufacturing and the digital performance of the ASICs. They found that wafers with faster transistors had improved digital performance. A set of experimental wafers were produced with modified transistors which could be tested for an improvement in SRAM performance.

Four pairs of experimental wafers were produced by the foundry with different process changes. Two pairs were manufactured with the gate size of the transistors reduced by 2 nm or 4 nm. One pair was manufactured modifying the process using a technique proprietary to the foundry which would speed up the transistors without modifying their geometry. The final pair of experimental wafers had both the 2 nm gate size

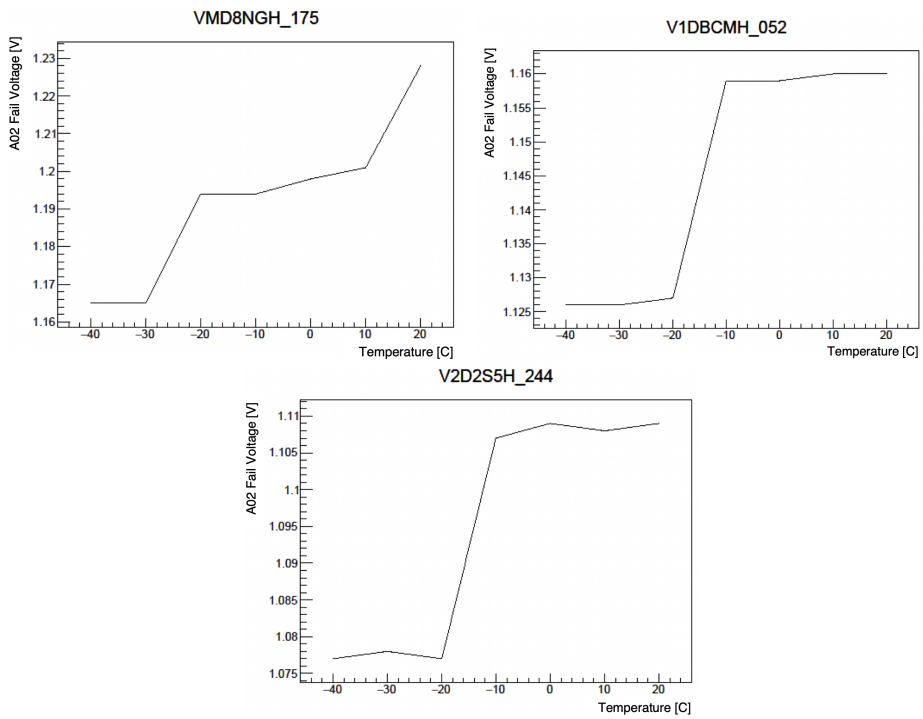


Figure 4.21: Digital voltage at which ABCStar chips failed the A02 test as a function of temperature after being irradiated to 100 Mrad with a 50/50 duty cycle.

reduction and the faster transistors. In addition, four wafers were produced in this lot using the standard processing for comparison.

The effect of the different process variations on A02 performance are summarized in Table 4.6. The different process variations all resulted in improvements of the SRAM performance and a lower passing voltage for the A02 test. The process with the biggest improvements that lead to the least amount of A02 failures and the lowest average passing voltage was the combination of the 2 nm gate size reduction with faster transistors.

Table 4.6: A02 statistics for standard wafers and experimental wafers with a 50/50 duty cycle. Decreasing (increasing) the LDO register setting increases (decreases) the digital voltage by approximately 25 mV.

Wafer	Process	LDO Register Setting vs. Fail Count							Pass Voltage (mV)	
		−3	−2	−1	+0	+1	+2	+3	Mean	Sigma
V2DCCKH	standard	10	10	10	38	209	417	457	1183	22
V3DCCJH	standard	5	5	7	22	134	376	456	1174	21
V5DCFYH	standard	8	8	9	23	151	403	463	1177	19
V9DCCDH	standard	2	2	2	22	153	375	460	1176	21
VADCFTH	gate −2 nm	6	6	6	25	175	419	462	1180	19
VDDCC9H	gate −2 nm	6	6	6	7	54	286	442	1162	19
VHDCFLH	gate −4 nm	4	4	4	4	21	195	390	1152	20
VIDCFKH	gate −4 nm	8	8	7	7	24	168	382	1149	20
VIDCK0H	faster transistors	4	4	4	4	9	72	285	1138	19
VJDCC3H	faster transistors	1	2	2	2	2	5	67	1122	14
VJDCFJH	−2 nm+faster	7	7	7	7	7	16	120	1124	14
VUDCF9H	−2 nm+faster	8	8	8	9	9	21	107	1122	14

Although the digital performance of ASICs was improved with the modified processing, there was also a change in analog performance compared to standard processing. On average, the experimental wafers had higher voltages for a given LDO setting. The wafers with faster transistor processing also had a lower average bandgap voltage and higher gain in the front-end analog circuitry. Wafers with 4 nm gate size reduction had the best combination of improved digital performance with the closest analog performance to standard process wafers.

4.5.3 Final Decision to Resolve SRAM Issue

As shown above, the yield of wafers passing A02 at the initial operating conditions (1.2 V, 50/50 duty cycle) could be quite low. To ensure that enough ABCStar ASICs would be available for the ITk and that they would not experience hit loss during their

lifetime, steps needed to be taken to improve performance. Even though experimental wafers were shown to have higher yields for digital tests of the ABCStar, the lower analog performance resulted in a net neutral gain in overall yield. In addition, 250 wafers with standard processing had already been manufactured, so the decision was made to keep the original manufacturing process for the remaining wafers.

It was determined that the safest and easiest method to increase the yield and reliability of the ASICs was to increase the digital voltage and duty cycle being supplied to the chips. The digital operating voltage of the ABCStar was increased from 1.20 V to 1.25 V. An increase of 50 mV drastically increases the yield of the worst wafers probed, and also moves each individual chip away from the threshold where hit loss would occur. The increase in voltage also satisfied the thermo-electric requirements of the ITk. There is still the possibility of further increasing the voltage for each chip up to 1.30 V without compromising the thermo-electric performance of the detector. This step could be taken if hit loss started to occur for chips after being installed in the detector due to effects of radiation throughout the lifetime of the ITk.

The duty cycle being supplied to the ABCStar could be increased by switching the wire bonds that connect the ABCStar to the HCCStar supplying the duty cycle. An HCCStar without crossed wire bonds would provide a duty cycle ranging from 50/50 down to 48/52 depending on the fine and coarse delay settings used. After the wire bonds were crossed, this would mean that a duty cycle of 50/50 up to 52/48 would be supplied to each ABCStar. The wire bonds are crossed at the output of the HCCStar to invert the signal.

The new criteria for all digital tests to pass during on-wafer probing was for the tests to pass at a digital voltage of 1.25 V with a 49/51 duty cycle. In addition, the version

of the P99 test previously conducted at 1.10 V would now have to pass at 1.15 V to keep this 100 mV below the nominal voltage. This increase in voltage would reduce the number of A02, P03 and P99 test failures and increase the yield of ABCStar wafers to ensure that there would be enough readout ASICs for the ITk. Each ABCStar should also receive a duty cycle of 50/50 or greater while installed in the detector, so probing at 49/51 allowed some additional operating margin. The digital performance should also improve once the ASICs are cooled during operation in the ITk and these improvements will surpass any decrease in SRAM performance from irradiation effects. After making these changes to the testing criteria, the average category A yield went from 66% to 86% in production probing. These changes greatly improved the yield of ABCStar ASICs, while ensuring confidence that the chips will function properly in the ITk.

4.5.4 Lessons Learned for future experiments

The ABCStar SRAM issue arose because the timing models used were old and based on previous best practices. The SRAMs worked properly during prototyping and previous versions of the ABCStar (ABC130 and ABCStarV0) and thus were considered to be silicon proven. This provided strong reasoning for the chip designers to be confident in the SRAM building blocks and not recompute the timing models. Only one lot of wafers was produced during ABCStar prototyping, which was not representative of full range of variation of process space of manufacturing, so the ASICs with slower transistors were not encountered till production probing.

For future experiments it is important to keep in mind that issues can appear with components that have been used in previous experiments and are regarded as silicon

proven. Effort should be taken to try and update all parts of the simulation during the design of ASICs for future experiments. Unfortunately models can have errors and a process called *striping* should be used on prototypes [71]. Striping involves foundries manufacturing experimental wafers to cover a representative range of process outcomes. If the time is taken to produced striped wafers and properly characterize their behaviour, a better understanding of the range of results to expect during production probing can be obtained. This may expose potential design issues and provide a more accurate estimate of the yields to expect.

4.6 Probing Results

The overall yield of category A ABCStar ASICs through production probing at DA and RAL was 85.88%. This section summarizes the probing results for wafers tested at DA and some of the most common failure modes during on-wafer probing.

The following results for probing at DA exclude lot 0HZJB00000 that had poor digital performance (Section 4.5) and lot 144YB00000 that had poor analog performance. Wafers in lot 144YB00000 had a wide range of category A yields, ranging from 3.83% to 84.68%. A majority of the failures occurred during initial power-on tests and since these chips would not power up properly, it is thought that there was an issue during manufacturing of this lot. Since the yields of the wafers vary dramatically across the lot, there is no confidence that the passing chips will not also exhibit problems in the future. For this reason, this lot needs to be replaced and will not be used in the ITk. Excluding the two problematic lots, the results from 333 wafers (156,510 ASICs) probed at DA are summarized below. Across these wafers there is an average category A yield of 85.07%, category B yield of 4.92% and category X yield of 10.01%.

The locations of category B and X failures across all 333 wafers can be seen in Figure 4.22. There is a slight radial pattern with more failures towards the edges and centre of the wafer, which is expected. There is also a stripe on the second column from left with an increased amount of category X dice. Note that die number 467 and 470 were not probed at DA and are category X for all 333 wafers. Die number 468 and 469 were added partway through production probing, so many are category X and were not tested. These are the four locations where additional dice were found to be fully processed and occasionally passed testing as mentioned in Section 4.2.

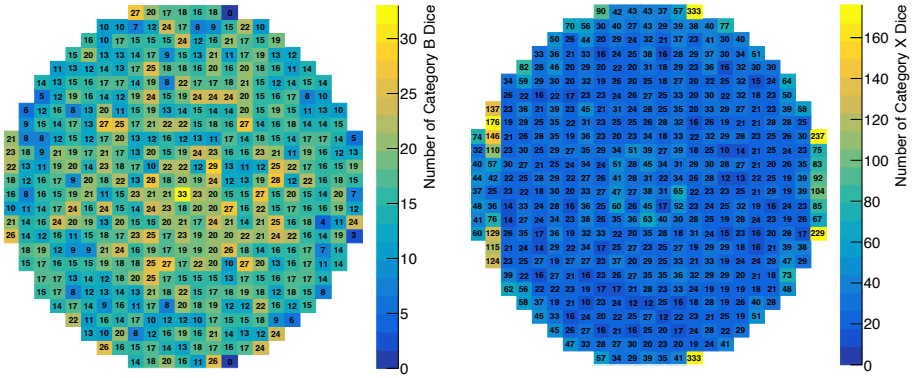


Figure 4.22: The number of category B dice (left) and category X dice (right) at each die location across 333 production ABCStar wafers probed at DA-Integrated.

The breakdown of analog and digital failures can be seen in Figure 4.23. Analog failures occur on 8.50% of dice tested and digital failures occur on 2.53% of dice tested at DA. Further breakdown of the analog and digital tests that occur on ASICs is explained in the following sections.

4.6.1 Analog Test Results

Failures from initial power-on tests and the bandgap voltage measurement can be seen in Figure 4.24. The initial power-on tests include ensuring there are no shorted

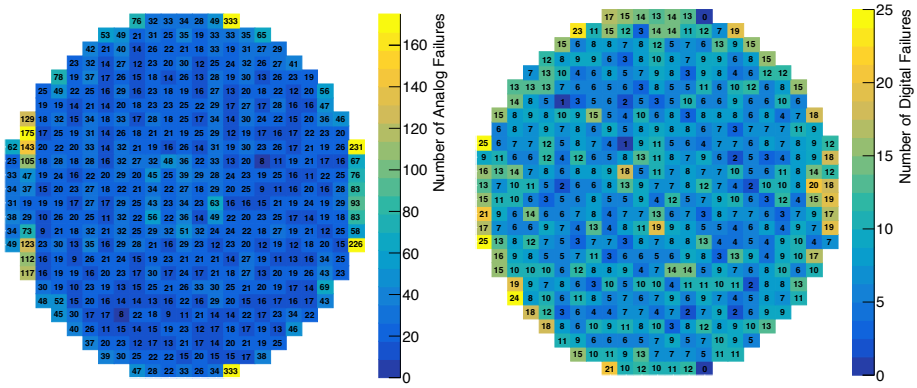


Figure 4.23: The number of analog test failures (left) and digital test failures (right) at each die location across 333 production ABCStar wafers probed at DA-Integrated.

pads and checking analog and digital voltages and currents as the power is brought up in stages to the chip. These power-on tests fail on 2.67% of the dice tested and the bandgap voltage measurement fails on just 0.58% of dice tested at DA. A majority of analog tests have similar failure patterns to this, where there is some radial dependence but no other visible patterns or pathological behaviour is observed.

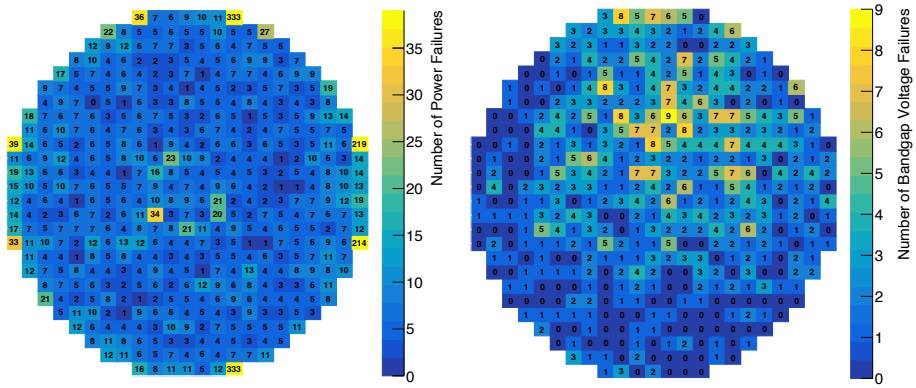


Figure 4.24: The number of dice failing initial power tests (left) and bandgap voltage test (right) at each die location across 333 production ABCStar wafers probed at DA-Integrated.

Two tests related to the front-end analog portion of the chip are mean noise and VT50 RMS. The location of failures across the production wafers probed at DA can be seen in Figure 4.25. There is an obvious pattern of stripes where some columns of dice

are more likely to fail due to high mean noise. The same dice are more likely to fail due to a high VT50 RMS, although the pattern is not as visible. It is suspected that this is related to the probing setup at DA. This striped pattern is also discussed in Section 4.3.2 and different investigations into needle cleaning or over-travel of the needles did not show improvement.

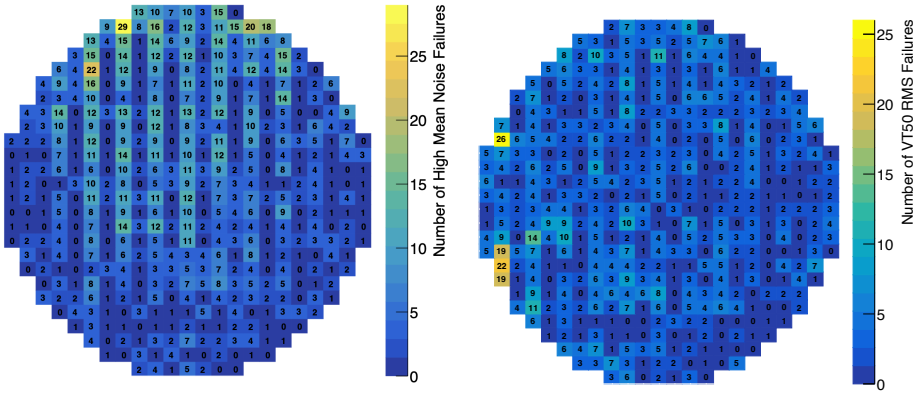


Figure 4.25: The number of dice with high mean noise (left) and high VT50 RMS (right) at each die location across 333 production ABCStar wafers probed at DA-Integrated. There is a pattern of stripes where dice in every second column are more likely to fail these tests.

Another front-end analog test that has a distinct failure pattern is related to the slope of the gain across all 256 input channels of the chip. If the gain increases or decreases too much across the channels the die is labelled as category X. Even if the gain for individual channels and mean gain across all channels is within the desired range, a "slopey" chip results in a large variation in response across the channels. It is desired to have a uniform response so that all channels will have the same response to the same stimulus, so chips with a high gain slope are removed. There are locations on the wafers that are more likely to fail this test, including a stripe of failures in the second column from the left and in the rightmost column, as seen in Figure 4.26. This same striped pattern is seen on wafers that are probed at RAL, suggesting that

this is related to manufacturing and independent of the testing site. This pattern suggests that this is not caused by normal process variation across the wafer, as the first column of dice do not show a similarly high failure rate. One possible explanation is a handling issue with the wafers during or after fabrication. It is possible that the handling process is causing damage to specific locations on the wafer which is leading to the difference in gain across the chip. If the handling is consistent across all batches of wafers, the same pattern of failures would be seen across all batches. 0.95% of ASICs probed at DA fail the gain slope test.

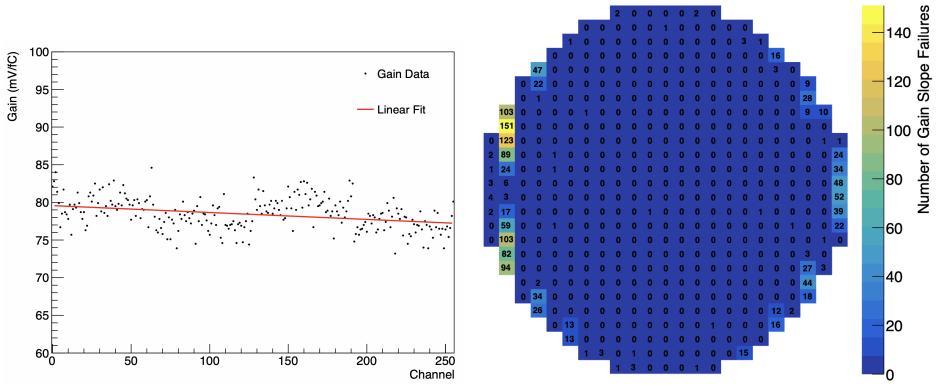


Figure 4.26: A sample plot showing the linear fit to the gain across all 256 chips for a die failing the gain slope test (left). The number of dice failing the gain slope test at each die location across 333 production ABCStar wafers probed at DA-Integrated (right).

4.6.2 Digital Test Results

The digital tests were among the last tests implemented for on-wafer probing of the ABCStar. It was expected that there would be a limited number of dice failing due to digital test failures compared to failures from analog testing. The first reason for this, is that a majority of the chip functionality has already been tested before running the digital tests, so most of the bad chips have already been detected. The second

reason is that the manufacturers of the chip perform rigorous testing to ensure that the digital performance of chips is acceptable across any process variation that may occur. Any wafers that did not fall within the acceptable limits of process parameters would have been rejected during testing at the factory. Analog circuitry, on the other hand, is deeply affected by process variation and careful design techniques have to be employed to ensure that a circuit will work on all areas of the wafer. This makes analog failures due to process variation much more likely to happen.

The A02, P03 and P99 digital tests that exercised the RAMs have the highest rates of failure. These test the triggering and readout of hits, which moves data through the RAMs, which are parts of the digital logic that are not exercised in the analog test flow. There is also a high rate of failures for these tests due to an issue with the chip SRAMs, which is described in Section 4.5. A02 is particularly prone to failure because it exercised a large portion of the digital logic, including register reads and writes and triggering and readout of hits. A02 injects a time-varying test pattern into the readout pipeline, which helps expose any timing issues with the digital circuitry. The P99 digital test exercised the RAM of the chip, and is also conducted multiple times with clock phases disabled or at a lower digital voltage than other chips. This was done for the P99 test because it tests critical registers in the ABCStar, checks the bunch crossing ID counter (BCID) and performs read of the RAMs to test the data taking ability of the chip. It is also shorter to run than some of the other robust tests, such as A02, which helps keep a lower total probing time.

A11 tests the SEU counter while disabling various clock phases of the chip. This test is particularly important as it can detect damage to the triplication of the ABCStar. If one or more of the clock trees has physical flaws, errors happen when the other clock trees are disabled. Many of the chips that fail A11, also have bad or untrimmable

channels, indicating serious issues with the metal connectivity. If a production lot has many A11 failures, this could indicate a major issue during the fabrication process.

The A16 and A17 digital tests exercise the readout pipeline in two different test modes. Both digital tests exercise similar logic and these tests will commonly fail together.

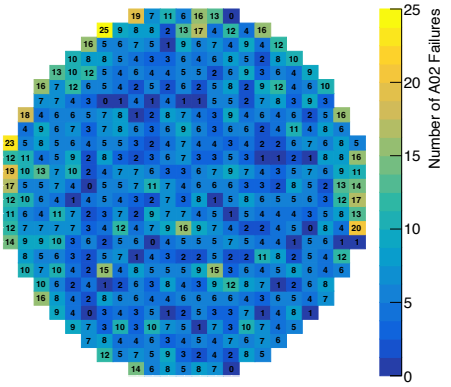
Maps showing the failures for digital tests across production lots tested at DA can be seen in Figure 4.27. A02 has the highest failure rate, at 1.80% of production chips failing the test at DA.

Any of the digital tests that are not mentioned above have a low rate of failures during on-wafer probing.

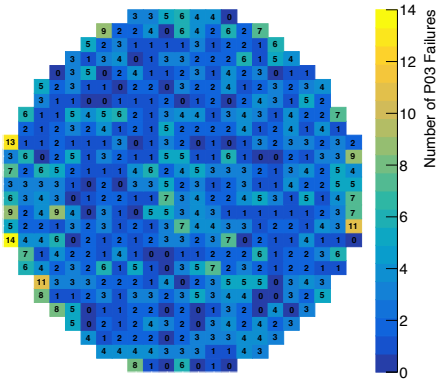
4.6.3 Wafer Comparison

To ensure that test results are consistent between the two ABCStar probing sites, two cross-checks were performed by probing the same wafer at both sites. The first wafer probing cross-check was done before production probing, while the testing routine was still being finalized at both sites. The second cross-check occurred about 5% into production probing to make sure that testing was still consistent between sites and compare results from tests that had not been finalized before the first wafer cross-check.

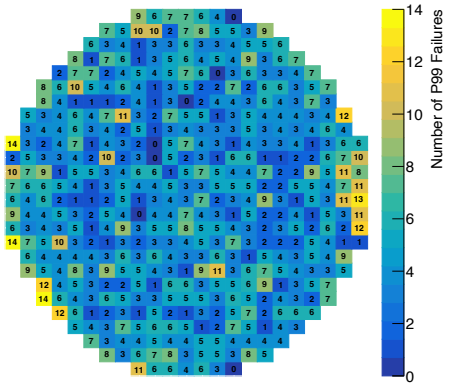
The probing cross-checks involved testing two wafers at RAL and then shipping these wafers to DA-Integrated to be reprobed. The categories for the dice on both wafers were compared between sites and any discrepancies with categorization were investigated.



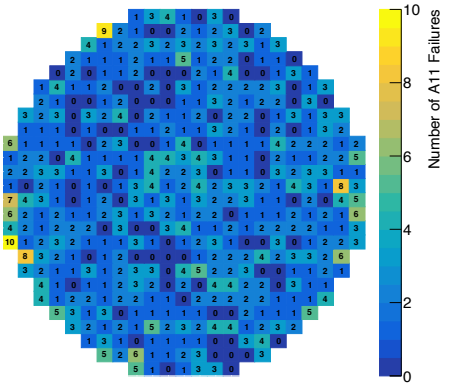
(a) A02 Failures.



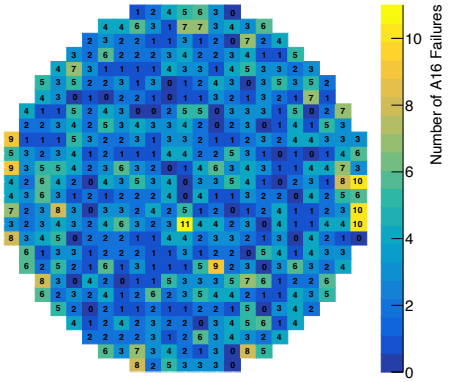
(b) P03 Failures.



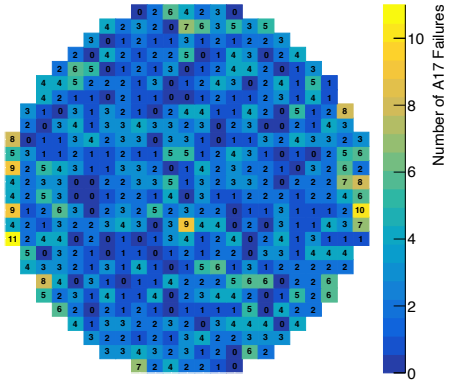
(c) P99 (clocks enabled) Failures.



(d) A11 Failures.



(e) A16 Failures.



(f) A17 Fails.

Figure 4.27: Digital test failures for six different digital tests across 333 production lots tested at DA-Integrated. Tests are executed with a 49/51 duty cycle and a digital voltage of 1.25 V, except P99 which was conducted at 1.15 V.

4.6.3.1 Initial Cross-Check

Two wafers from two separate manufacturing lots were chosen to be cross-check wafers. The wafers chosen both had relatively high yields of Category A dice.

After initial probing of the wafers at both sites 908/940 (96.60%) dice were placed in the same categories by RAL and DA. Dice that failed at both RAL and DA did not always fail the same test due to differences in test order, so only tests with different outcomes were investigated further. Causes for discrepancies in categorization are listed below:

- A02 Digital Test Vector: Some dice were category X at DA because of failing digital test A02, which had not yet been implemented at RAL. It is expected that if the dice were re-tested at RAL after the A02 digital test had been implemented, both probing sites would classify them as category X.

- ShuntCtrl: There was an extra cut on slope of the shunt device at DA causing some dice to fail. This cut has since been removed so these dice would no longer be category X.

- VT50 RMS: There is a cut on the RMS of the VT50 for all channels on a die causing some dice to be category B at RAL that was not implemented at DA. This cut was introduced during production probing at DA.

- Dice that failed initial continuity or short circuit tests at DA, but pass without problems at RAL are thought to be caused by poor connection by the probe needles. These dice are expected to pass at Carleton if re-probed. The automated tester at DA-Integrated does regularly clean the probe needles and will stop if

too many continuity failures occur in a row in order to prevent dice from being lost due to this issue.

- Mean gain and DAC measurements: Two dice failed at analog tests at RAL but passed at DA. These tests were conducted multiple times at RAL for these tests and passed for some of the measurements. It is thought that there are no issues with these dice.

- Bad Channels: Some dice had extra bad channels at RAL that were not detected at Carleton. This was due to a bug which was subsequently fixed.

If changes were made to reconcile any differences between RAL and DA based on the tests listed above, and the wafers were reprobed it is thought that these discrepancies would be resolved. This leaves only two additional dice that were categorized differently at RAL and DA that were not caused by any of the reasons listed above. This would bring the level of agreement for the wafer cross-check up to 99.8% between the two probing sites.

These two dice failed a test at one probing site but passed at the other. In both cases the measurements were right on the edge of the test limit. It is expected that this disagreement could happen for dice at the edge of a limit due to differences in probing equipment and measurement error between probing sites.

All of the discrepancies between tests at RAL and DA were addressed after the cross-check. Test results for these tests were compared between sites to ensure test coverage was uniform between both probing sites before starting production probing.

4.6.3.2 Production Cross-Check

A second probing site cross-check was conducted during the beginning of production probing. This was done to ensure that consistent results were still being obtained between probing sites during production probing and to compare results for tests that had not been implemented during the previous cross-check. Two wafers were chosen for comparison after probing at RAL. For this second cross-check two wafers with lower than average yields were selected to compare results for dice that failed various analogue and digital tests. The two wafers were reprobed at DA to compare categorization.

After initial probing 853/940 (90.74%) of dice were placed in the same categories at both probing sites. The main causes for discrepancies are listed below:

- DAC tuning values: Some dice had DAC tune values that differed by one setting between sites, resulting in an ASIC passing a test limit at one institute but failing at the other. This occurred in ASICs where the operating voltage for a DAC was right in between settings at the edge of a test limit, so slight variations in measurements between sites resulted in different tune values (and pass/fail criteria for a test).

- Strobe Delay: This lot of wafers had higher strobe delay settings on average compared to other ABCStar lots. This resulted in many dice having a strobe delay tune value right around the accepted limit. There appears to be a slight systematic difference in strobe delay values between production sites where DA selects a tune value at one lower setting than RAL. This resulted in many dice passing the strobe delay test at DA that had failed the strobe delay test at RAL.

Looking closely at the results for the strobe delay tests at both sites we see that these chips still apply an even strobe delay across all channels and there is plenty of margin in the strobe delay settings.

- VT50 RMS: When the VT50 RMS is too high it is usually a sign of poor contact with the prober. There were problems with a high number of dice failing VT50 RMS at DA. After reprobing the dice the VT50 RMS reduced below the limit of 3.4 mV and their categorization agreed with RAL.

- Bad/Untrimmable Channels: There were channels that were flagged as bad/untrimmable at one probing site but not at the other. Bad channels that occurred during initial probing at DA, but not at RAL were not flagged as bad after reprobing. It is thought that these bad/untrimmable channels could show up if there was poor contact or the channels were marginally passing test limits.

- Mean Gain/Gain Slope: Some chips marginally passed the mean gain or gain slope limit at one site but failed at the other. This discrepancies were caused by small differences in measurements between sites.

- BCAL Slope: One die failed at RAL due to a BCAL slope being marginally above the test limit, while marginally inside the limit at DA.

- Power test failures: A few dice passed initial power on tests at DA but failed at RAL. This is thought to be due to poor contact issues during initial probing at RAL.

A majority of the differences in categorization between probing sites were resolved by reprobing dice. Many other differences arose from marginal difference in measurements between sites causing a test to pass at one site but fail at another. Strobe delay

failures made up a large portion of discrepancies between categorization, but this lot was an outlier with higher than average strobe delay tune values and this was not seen across other production lots.

The production cross-check between probing sites gave confidence that neither site was passing faulty chips for use in ITk. However, discrepancies between categorization showed that there were good chips being classified as category B or X at both probing sites, reducing the yield of wafers. In order to mitigate failures caused by poor contact at DA, failing dice were reprobed upon completion of each wafer. This allowed for a second opportunity to establish good contact with each failing die and pass any tests that previously failed due to poor contact.

4.7 ABCStar Dicing

After on-wafer probing, the ABCStar wafers are sent for *dicing*, which is the process of cutting up a wafer into individual ASICs. The layout of the ASICs on an ABCStar wafer can be seen in Figure 4.28. There was intended to be 466 fully functional ASICs on each wafer, but visual inspection showed four additional ASICs outside of this region appeared to be fully processed and were found to occasionally pass on-wafer probing.

In the first step, the wafers are *thinned* – reduced in thickness by surface grinders – to $(300 \pm 5)\,\mu\text{m}$. Thinning of the wafers helps reduce the size for packaging, improves heat dissipation, and can even reduce the probability of the chips fracturing [72]. The wafer is then diced using precision saws to produce individual ASICs that will be used in the ITk. A diagram of the reticle for dicing the ABCStar wafers can

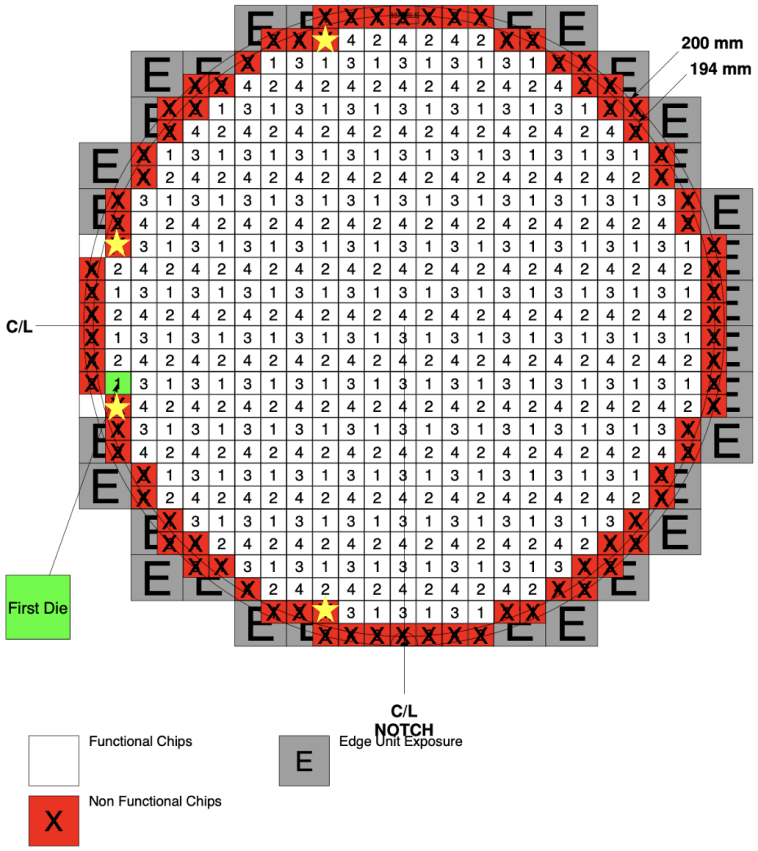


Figure 4.28: A diagram showing the wafer layout for the ABCStarV1. There are 466 white squares which indicate the location of fully processed ASICs on the wafer. Four additional ASICs indicated with stars were not intended to be functional, but were tested during production probing as they were found to occasionally be fully functional. The numbers indicate the location of dice on the dicing reticle.

be seen in Figure 4.29. The dimensions of the ABCStar ASICs after dicing are $(8047 \pm 10)\,\mu\text{m} \times (7110 \pm 10)\,\mu\text{m}$.

After dicing, the individual dice undergo visual inspection and metrology measurements to ensure they are fit for use in the ITk and did not incur any damage in the dicing process. Each wafer thickness is measured after thinning to ensure it is the correct thickness. A few dice per wafer are sampled to measure the dimensions and make sure that they are within the tolerance. Finally, all chips are inspected to make sure

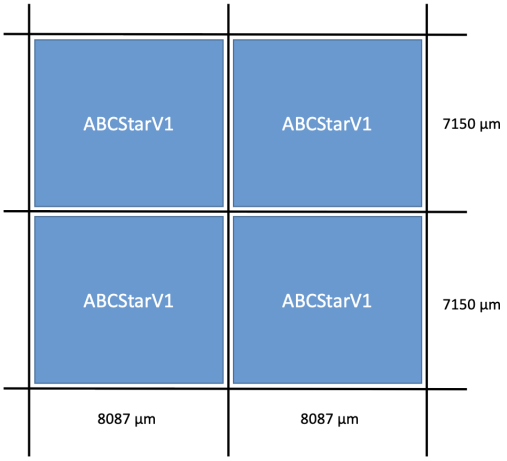


Figure 4.29: A diagram of the ABCStar dicing reticle. The blade size (kerf) is 40 µm, resulting in ASICs that are (8047 ± 10) µm $\times$ (7110 ± 10) µm after dicing.

that there is no cracking or chipping along the edges where dicing occurred. After the wafers are diced, any ASICs that are damaged in the process will be moved to category R and rejected for further use. The expected yield from dicing is over 98% passing dice. The final handling yield from dicing through production probing of the ABCStar was 99.80%.

The final step in the dicing process is placing the individual ASICs into Gel-Paks, containers that are used for shipping and storage until the chips are installed in modules. A machine picks up the diced ASICs using a vacuum pen and places them in the Gel-Pak, which can hold up to 72 chips (9×8) as seen in Figure 4.30. Each Gel-Pak will only hold one category (A, B, X, T or R) of ABCStars, based on wafer maps provided to the dicing vendor from on-wafer probing. The Gel-Paks receive a label that indicate the original ABCStar wafer name, the category of dice, the die numbers included in the Gel-Pak and a QR code linking to tracking information of the Gel-Pak on the ITk Production Database (Section 4.8).

Figure 4.30: A picture of the outside (left) and inside (right) of a Gel-Pak filled with 72 category A chips after dicing.

4.8 ITk Production Database

In order to manage all of the test data, categorization, location and shipping information on all devices used in the ITk, an online repository called the ITk Production Database (PDB) is used. All parts used in the ITk are registered in the PDB and can be tracked back to all of the testing sites they were shipped between and the results of any tests that were conducted along the way. This is extremely important for traceability, as ATLAS is an international collaboration, and parts used in the ITk can be shipped between many sites across different countries before installation. The PDB also includes all test data collected for parts before they are installed, which can be used to help troubleshoot any issues that might arise during operation of the ITk.

Work involving the on-wafer probing data for the ABCStar uploaded to the PDB and shipments regarding ABCStar wafers was completed as a part of this thesis. This included deciding which test data would be uploaded to the PDB and creating data

in the proper formats for upload. Software was also developed to interface with the PDB and upload the test results for hundreds of thousands of tested ASICs. Scripts were also used to virtually ship these ASICs in the PDB to facilities where modules were being assembled so they would have access to all probing data.

ABCStar ASICs are initially created in the database when the wafers arrive at CERN from the manufacturer. Each wafer will have 470 *child* components, which are the individual ABCStar ASICs. The wafers and ASICs have unique serial number identifiers used in the PDB as described in Section 4.2. The wafers are virtually "shipped" in the PDB from CERN to their probing location (RAL or DA) at the same time as they are physically shipped to these locations. This ensures that the probing data can be tracked back to the appropriate on-wafer probing site where the data was collected. After wafers have been probed at either location, a subset of important test results collected are uploaded to each ASIC in the database. The wafer is updated to show that probing has completed and each individual ASIC has the resulting category from probing updated. These wafers are then virtually shipped in the database to the dicing vendor, and are updated to reflect that they have been diced after dicing is completed. The wafers are then shipped back to CERN in the database (and physically), where the wafers are virtually disassembled, and ASICs are assigned to the appropriate Gel-Pak objects in the database. These Gel-Paks will then be shipped to various locations to be installed in modules, which will be placed on petals or staves to be eventually installed in the ITk. All of these installations, shipments and the results of any tests are recorded in the ITk PDB for traceability.

The on-wafer ABCStar probing results that get uploaded to the database are divided into 7 different groups based on the test type: EMUX, IMUX, DAC, POWER, GAIN, LDO and NTV tests. A breakdown of the test results saved by test type is as follows:

- POWER Tests: Analog and digital currents and voltages measured during initial power-on tests.
- LDO: Voltages and tuned values for the digital and analog LDO regulators.
- EMUX: Analog measurements made using the chips external ADC.
- IMUX: Calibration and measurements of the chips internal ADC.
- DAC: Tuned values and voltages of the chips internal DACs.
- GAIN: Results from the front-end analog tests. This includes pedestal trimming settings, strobe delay settings and gain and noise for channels on each chip.
- NTV (New Test Vector): Results from the digital tests with voltages and duty cycles that each test was performed at.

In addition to this, properties related to probing of the chip and analysis of test data are also saved for each test on the database. The information saved is as follows:

- Hexadecimal failure code which indicates which test(s) failed during probing
- Decimal eFuse ID
- Hexadecimal eFuse ID
- Probe station used to probe wafer
- User of probe station (or person who performed analysis of probing data for wafers probed at DA-Integrated)
- URL for version of analysis code

- Git tag for version of analysis code
- Commit date for last change to analysis code
- Commit message for last change to analysis code
- Parent wafer name the ABCStar originated from

All of this probing data is saved in a ROOT file and a JSON (JavaScript Object Notation) file for each test type. JSON is a standard text-based format for data and provides an easy format for uploading to the PDB. Data is uploaded to the ITk PDB using a Python script that reads the JSON files and uploads all necessary information. Each test type is added to the ABCStar ASIC as a separate *Test Run* object on the database. Each Test Run is labelled as Passed, Passed with Problems or Failed. The total size of the uploads for each chip is 24 kB.

Some of these test results are downloaded and used in later testing, such as module testing. Test results such as DAC tune values are downloaded and used to tune ABCStar chips to the appropriate settings when running tests on a module. By using information obtained from on-wafer probing that has been stored in the PDB, this ensures that the ABCStar readout chips are properly tuned to the optimal settings before running tests.

A separate Python script is used to set the *stage* for each ABCStar ASIC after the test uploads have completed. This script reads in the category from probing for each ASIC from a text file and updates the stage from "On Wafer" to the appropriate category (Category A, Category B, Category X or Category T). Once the stage has been updated from "On Wafer", probing data can no longer be uploaded to that component in the database. These checks ensure that data does not get improperly

uploaded to the wrong component, or to components in a different stage of testing.

Chapter 5

Event Reconstruction and Simulation

The ABCStar ASICs discussed in Chapter 4 will be used in the Inner Tracker upgrade that will be installed for the High Luminosity phase of the Large Hadron Collider starting in 2026. One of the highly anticipated physics analyses that will be probed during this phase of increased data taking is the Higgs to dimuon decay [73]. This thesis presents a search for the $H \to \mu\mu$ decay, focusing on data collected during Run-3 of the ATLAS detector from 2022–2024. This chapter describes the reconstruction of data collected by the ATLAS detector and the simulation of MC samples used for this analysis.

Section 5.1 outlines how different physics objects used in this analysis are reconstructed from signals collected by the ATLAS detector. Section 5.2 describes the data collected by the ATLAS detector during Run-3 and Section 5.3 discusses the Monte Carlo simulation samples used to model the data.

5.1 Event Reconstruction

An *event* refers to the full detector readout from a pp bunch crossing in the ATLAS detector. Multiple protons will collide during each bunch crossing, typically resulting in a final state with hundreds of particles, a majority of which will be hadrons and photons. When these particles interact with the ATLAS detector they will create signals that are used to *reconstruct* lists of *physics objects*. These physics objects include muons, electrons, photons, and showers of particles called jets. The goal of the ATLAS detector is to try and accurately reconstruct the particles in each event and measure their energy and momentum to perform physics analyses.

Although we are usually interested in a specific process occurring during the pp collision for each event, there will be additional particles created from pileup. In-time pileup will leave detector signals from other soft pp collisions during the same bunch crossing, which can lead to the reconstruction of additional physics objects or distortions to the physics objects of interest. Out-of-time pileup from nearby bunch crossings can also result in distortions to the physics objects for the event of interest.

The main types of signals measured by the ATLAS detector from particles are the trajectories of charged particles, called *tracks*, and the energy deposited in the calorimeters. Using this information, algorithms are used to reconstruct the particles used in this analysis. Dedicated selection criteria is then applied to the reconstructed particles, such as a minimum transverse momentum, to select particles that are well reconstructed and relevant to the $H \to \mu\mu$ analysis. A process called overlap removal is also applied to ensure multiple particles are not reconstructed from the same tracks or energy deposits.

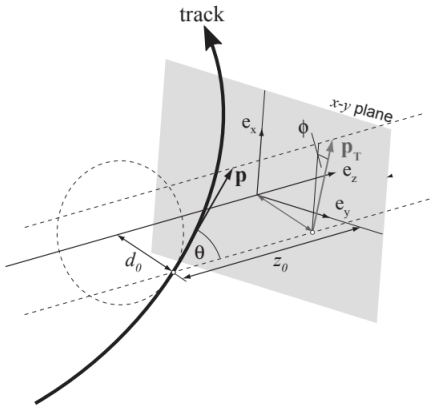


Figure 5.1: A diagram showing the track impact parameters for a particle in the ATLAS detector. Figure taken from Ref. [77].

5.1.1 Tracks and Vertex Reconstruction

Tracks from charged particles are measured using the inner detector (Section 3.2.3). The ID will record the interaction points, which are referred to as *hits*, in the layers that charged particles pass through. These hits are used to reconstruct the trajectory of the particle through the detector.

Prompt tracks near the centre of the beam line are reconstructed using a method called inside-out track building. The algorithm groups hits from adjacent layers of the ID to determine the 3D positions where charged particles pass through, forming space points [74–76]. These space points are then combined into track seeds.

For each seed, the impact parameters are calculated: transverse distance from the beam line (d_0), longitudinal distance from the beam spot (z_0), azimuthal angle (ϕ), polar angle (θ), and charge-to-momentum ratio (q/p). The impact parameters can be visualized in Figure 5.1.

Track seeds must satisfy selection criteria and be compatible with at least one additional

space point to suppress fake tracks. A Kalman filter [78] then adds additional compatible space points to each candidate. Since a space point can belong to multiple candidates, an ambiguity resolution step ranks tracks by their number of clusters, holes (layers where a hit would be expected but none are observed), χ^2, and momentum. A neural network classifier selects the best candidates based on these scores.

Finally, a global fit extends selected tracks into the TRT, again using a Kalman filter to include compatible hits. Successfully extended tracks are refit and added to the final collection.

Another method called outside-in track building is used to reconstruct tracks for charged particles that are produced further from the beam line due to secondary decays. Unused TRT hits are grouped into track segments using a Hough transform [79], which are then extended inward to the SCT and pixel detectors, following a process similar to the inside-out approach.

Tracks are used for the momentum measurement of charged particles and for vertex reconstruction. For this analysis, the primary vertex (PV) of interest is the hard scatter process where the Higgs boson is created and promptly decays into a muon-antimuon pair.

In Run-3, the PV reconstruction is done using the Adaptive Multi-Vertex Fitter (AMVF) algorithm [80]. Vertex seeds are first identified using a Gaussian track density seed finder, which models the distribution of tracks along the beam axis as a sum of Gaussian functions. Peaks in this distribution correspond to potential vertex positions. The vertex fitting is then performed with AMVF, which simultaneously fits multiple vertex candidates using all available tracks. Each track is assigned a weight based on its compatibility with each vertex, allowing tracks to contribute fractionally to

more than one candidate. An adaptive weighting scheme and deterministic annealing procedure are then used to iteratively refine the vertex positions.

Once PVs have been reconstructed, the hard scatter vertex needs to be identified. The hard scatter vertex is selected as the PV with the largest $\sum p_{\mathrm{T}}^2$. All other PVs are considered to be from pileup or interactions from other background particles passing through the detector.

5.1.2 Topological Clusters and Particle-Flow Objects

Energy deposits in the calorimeters (Section 3.2.4) are grouped into topological clusters, referred to as *topo-clusters*, using a noise suppression algorithm that identifies and combines neighbouring calorimeter cells with significant energy deposits above noise thresholds [81]. These topo-clusters represent localized energy deposits corresponding to individual particles or overlapping showers of particles within the calorimeter.

Tracks reconstructed in the Inner Detector are then matched to topo-clusters to form particle-flow objects (PFOs) [82]. This approach combines the momentum measurement of charged particles from the tracking system with the energy measurement from the calorimeter, providing improved reconstruction of physics objects. PFOs that are associated with tracks are called charged PFOs, and PFOs formed from calorimeter energy deposits without matching tracks are called neutral PFOs. For example, a charged pion is expected to produce a charged PFO, while a neutron or photon forms a neutral PFO. In the forward region ($|\eta| > 2.5$), where tracking coverage is limited, only neutral PFOs are reconstructed.

5.1.3 Muons

Muons are the most important object used in this analysis, and are reconstructed from a track in the ID or MS and energy deposits in the calorimeters. There are three different types of muons based on different reconstruction methods [83] that are considered for this analysis:

- Combined (CB): muons with tracks that are reconstructed with a combined fit where both ID and MS information are used.
- SegmentTagged (ST): muon candidates that are identified using an ID track ($p_{\mathrm{T}} > 22\,\mathrm{GeV}$) with a matching position and angle to a segment in the MS.
- CaloTagged (CT): muon candidates that are reconstructed using an ID track ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$) matching an energy deposit in the calorimeter that is isolated and consistent with a MIP.

Most of the muons that are returned from the reconstruction algorithm are produced by real muons, but there can be fake contributions from hadrons or electrons that mimic the signal of a muon. To improve the purity (the fraction of reconstructed muons produced by real muons), there are sets of quality criteria referred to as identification working points (WP). For this analysis, the "Loose" identification WP is used for muons to balance selection efficiency and background rejection. An isolation WP, "PflowLoose_VarRad", is also applied to reduce fake contributions from jets that could mimic the signal of a muon. Kinematic requirements of $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ and $|\eta| < 2.5$ are applied to muons in this analysis. Finally, to reject muons that are not from the hard scatter vertex, such as cosmic ray muons passing through the detector, there is a cut

on the impact parameters. The first cut is $|d_0^{\mathrm{BL}}/\sigma_{d_0}^{\mathrm{BL}}| < 3$, which is a significance cut on the transverse distance from the beam line over the uncertainty on the transverse distance. The second cut is $|z_0^{\mathrm{PV}} \cdot \sin\theta| < 0.5\,\mathrm{mm}$, which ensures the muon track originates from close to the PV.

5.1.4 Electrons

Electrons are reconstructed using an algorithm that matches clusters of energy deposits from the EM to a track from the ID [84]. For this analysis the "Medium LH" identification WP is used for electrons. The kinematic cuts on electrons are $p_{\mathrm{T}} > 7\,\mathrm{GeV}$ and $|\eta| < 2.47$ (excluding $1.37 < |\eta| < 1.52$, a transition region in the calorimeter with poor energy resolution). The "Loose_VarRad" isolation WP is used for electrons. The cuts on the impact parameters are $|d_0^{\mathrm{BL}}/\sigma_{d_0}^{\mathrm{BL}}| < 5$ and $|z_0^{\mathrm{PV}} \cdot \sin\theta| < 0.5\,\mathrm{mm}$.

5.1.5 Jets

Jets are collimated showers of hadrons that are produced in the pp collisions, usually produced from a high momentum parton. For example, in the VBF process (Section 2.2.2.2), two jets are expected to be produced from the outgoing quarks. Tracks and energy deposits in the calorimeters are used to produce jet constituent particles. An algorithm is then used to determine which constituents belong to which jet and calculate the properties of those jets. Jets have a radius, R, that represents the size of the jet in the $y - \phi$ plane. The four-momentum of all of the particles within a jet are added together to calculate the four-momentum of each jet.

Jets used in this analysis are reconstructed using the anti-k_t algorithm [85]. A brief

explanation of the algorithm is given below:

1. For each jet constituent particle calculate the k_t distance

$$d_{ij} = \min(p_{\mathrm{T},i}^{2k}, p_{\mathrm{T},j}^{2k}) \frac{\Delta R_{ij}^2}{R^2}, \tag{5.1}$$

   where $\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$, and $k = -1$ for the anti-k_t algorithm. The beam distance

$$d_{iB} = p_{\mathrm{T},i}^2, \tag{5.2}$$

   is also calculated.

2. Find the minimum of d_{ij} and d_{iB}. If d_{ij} is smaller, merge these two jet constituents into a single jet with four momentum equal to their sum. If d_{iB} is smaller, declare this jet to be a final jet and remove this from the list of jet constituents.

3. Repeat steps 1 and 2 until there are no jet constituents remaining.

Using the anti-k_t algorithm with $k = -1$ combines jet constituents with larger p_{T} first, resulting in circular jets clustered around the hard particles [86]. These jets are less sensitive to soft particles, like those expected from pileup, making it a common choice for ATLAS analyses.

The jets used in this thesis are built with PFOs (Section 5.1.2) as inputs, using the anti-k_t algorithm with a radius parameter of $R = 0.4$ [85]. The reconstructed jets are then calibrated to adjust their measured energies to correct for detector response and pileup effects [87].

Reconstructed jets with $|\eta| < 2.4$ must have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and jets with $2.4 < |\eta| < 4.5$ must have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ to be considered for this analysis. There is a slightly higher p_{T} threshold for *forward* jets, due to increased pileup and lack of tracking in the forward region of the detector. A jet vertex tagging (JVT) algorithm is applied to jets to reduce the number of pileup jets in the analysis. In the central region ($|\eta| < 2.5$) a neural network jet vertex tagger with the "FixedEffPt" WP is used, which has a 88% to 99% HS jet efficiency for jets with p_{T} from 20 GeV to 60 GeV. In the forward region ($2.4 < |\eta| < 4.5$) the "Loose" WP is used for the forward jet vertex tagger (FJVT).

An algorithm is also used to tag any jets passing the previous criteria as jets originating from bottom quarks, referred to as b-jets. A neural network b-tagging tool called GN2 [88] is used with the 85% WP. This means that 85% of jets originating from bottom quarks will be successfully tagged. These b-tagged jets are used in the analysis to look for decays where bottom quarks are expected, or veto events when they are not expected.

5.1.6 FSR Photons

Muons can emit final state radiation (FSR) photons through QED processes. Over half of the $H \to \mu\mu$ events will have FSR. About 10% of events will have an FSR photon with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ that can carry away a significant fraction of the energy from the Higgs decay. If we do not recover these FSR photons and add their energy back to the muons from the Higgs decay, the dimuon invariant mass, $m_{\mu\mu}$, will be lower. A dedicated tool is used to recover these FSR photons and reconstruct the dimuon invariant mass with FSR contributions, $m_{\mu\mu\gamma}$. The tool considers two types

of candidates for FSR recovery:

- 3×5 clusters seeded by topo-clusters with $\Delta R(\mu, \gamma) < 0.08$, $E_{\mathrm{T}}(\gamma) > 3.5\,\mathrm{GeV}$ and $f_1 > 0.2$, where f_1 is the fraction of energy deposited in the first layer of the EM calorimeter.
- Reconstructed photons and electrons [89] with $\Delta R(\mu, \gamma) < 0.2$, $E_{\mathrm{T}}(\gamma) > 3.5\,\mathrm{GeV}$ and $f_1 > 0.1$.

Any candidates passing the above criteria must also pass another cut,

$$E_{\mathrm{T}}(\gamma) > 25 \cdot \Delta R(\mu, \gamma) + 3\,\mathrm{GeV}. \tag{5.3}$$

The selection criteria requires that any FSR candidates are *collinear* ($\Delta R < 0.2$), which helps reduce the fake rate. Although *wide-angle* ($\Delta R > 0.2$) FSR photons can exist, they are less common (60% of FSR candidates are collinear), and including them greatly increases contamination from background processes. If more than one FSR candidate is found, the candidate with the highest E_{T} (or lowest ΔR if two candidates have the same E_{T}) is selected. Only one FSR candidate is considered for each event between the two Higgs candidate muons.

The FSR correction procedure described above was derived for events with exactly two muons. Events with busier environments, such as $t\bar{t}H$ or VH processes can result in a final state with more than two muons. FSR correction was not applied for any events where additional muons were present.

5.1.7 Missing Transverse Momentum

The protons colliding in the ATLAS detector have no initial transverse momentum, so all particles produced in each event will balance in the transverse plane due to momentum conservation. If all of the particles are reconstructed correctly, there should only be small momentum imbalances in the transverse plane caused by imperfect measurements of the particles. However, if a particle leaves no signal in the detector, there will be an imbalance. This can happen when particles, such as neutrinos, do not interact with the detector, or when particles exit through gaps in the detector.

The missing transverse momentum, $E_{\mathrm{T}}^{\mathrm{miss}}$, is calculated as the negative vector sum of transverse momenta of physics objects considered in each event[1]. For this analysis, reconstructed and calibrated muons, electrons, jets, and additional tracks from the PV not matched to reconstructed objects are used for the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. A dedicated software algorithm [90] is used to calculate the magnitude and azimuthal angle of $E_{\mathrm{T}}^{\mathrm{miss}}$ using these physics objects.

5.2 Run-3 Data

This thesis uses data from $\sqrt{s} = 13.6\,\mathrm{TeV}$ pp collisions at the LHC collected by the ATLAS detector during part of the Run-3 data-taking period from 2022–2024. The integrated luminosity for this period where the detector and triggers were functional for the the Higgs to dimuon decay is $165\,\mathrm{fb}^{-1}$. The breakdown of integrated luminosity by year can be seen in Table 5.1.

In Run-3, the expected number of inelastic pp collisions at each bunch crossing, μ,

[1] $E_{\mathrm{T}}^{\mathrm{miss}}$ has a magnitude and azimuthal angle and is sometimes denoted in vector form as $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$

Table 5.1: The integrated luminosity, $\mathcal{L}_{\text{int}}$, of the data collected by the ATLAS detector used for the $H \to \mu\mu$ analysis for Run-3 of data-taking. Values obtained from Ref. [91].

Year	$\mathcal{L}_{\text{int}}$ (fb^{-1})
2022	29.29
2023	26.66
2024	109.4
Total	165.4

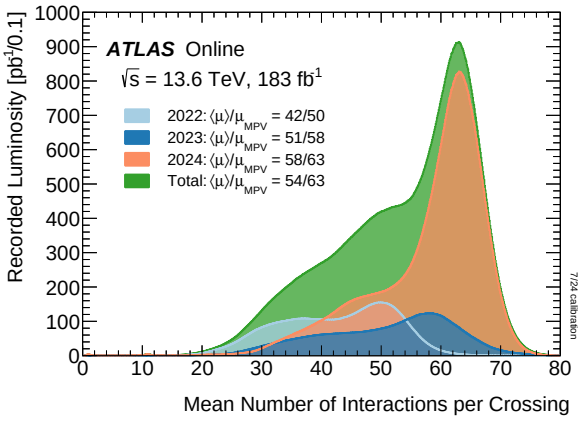


Figure 5.2: The luminosity-weighted distribution of the mean number of interactions per crossing for the 2022–2024 pp collision data at 13.6 TeV centre-of-mass energy recorded by ATLAS during stable beams. The mean pileup (μ) value, and most probable value (MPV) of μ are given for each year of data-taking. Figure taken from Ref. [91].

is approximately 50. The breakdown of μ for each year is shown in Figure 5.2. This means that every event with a $H \to \mu\mu$ hard scatter process is also expected to have about 50 inelastic pp collisions. Most of this pileup is quite soft, resulting in low momentum particles that primarily act as noise on top of the desired hard-scatter interaction. There is about a 1% chance that a pileup interaction will be hard enough to produce a $p_{\text{T}} = 20\,\text{GeV}$ jet.

5.3 Monte Carlo Simulation

Monte Carlo (MC) generators are used to simulate the $H \to \mu\mu$ signal and background processes. These samples provide predictions of the data collected by the ATLAS detector and are used for estimating backgrounds, optimizing event selection criteria and for training classifiers to discriminate between signal and background. MC simulation begins with *event generation*, which models the interactions between the quarks and gluons in the colliding protons, the subsequent parton showering, and the hadronization and decays into stable particles (Section 2.3). The end product of event generation is a list of *stable particles*, defined as those with lifetime such that $c\tau > 1\,\text{mm}$. This set of particles is stored in the MC samples and referred to as *truth particles*. These truth particles are then passed into the *detector simulation*, which calculates how each particle interacts with the detector material, how they shower into secondaries, and how much energy they deposit in each sensitive element of the ATLAS detector. The final step is *digitization* and *reconstruction*. Digitization turns the simulated energy deposits into a detector signal that mimics the raw data from the real detector. Detector signals due to pileup are also added to the MC at this step. This is done by adding simulated pileup events and superimposing their associated detector signals to those of the primary process. The number of pileup interactions overlaid matches the beam conditions of Run-3 data. The simulated output from digitization then undergoes reconstruction in exactly the same way as real data events. The resulting MC event sample can be directly compared to data.

Since detector conditions can vary between different years of data taking, MC simulation is generated in several *campaigns*. Detector components can break down or be repaired, so changes need to be made to the detector modeling. The pileup

will also vary between years as shown in Figure 5.2. The MC23a campaign matches conditions for data collected in 2022, MC23d matches 2023, and MC23e matches 2024. Reconstruction and calibration of physics objects can also differ between data from different years.

5.3.1 Signal Samples

Signal samples important for this thesis are samples where a Higgs boson is created and decays to two muons. The samples are divided based on the production mode of the Higgs boson (Section 2.2.2), which includes ggF, VBF, VH, $t\bar{t}H$, $b\bar{b}H$, and tH.

Most of the signal samples used in this thesis are generated with Powheg-Box event generator [92]. The tH and tWH samples are generated with the aMC@NLO [93, 94] program. These generators calculate the matrix elements for each process to NLO accuracy and model the particles generated in the proton hard scatter event. Outgoing partons from this hard scatter event will then radiate outwards until hadronization occurs. The parton showering for all signal samples is modeled with Pythia 8.309 [31] and the decays of b- and c- hadrons is treated by EvtGen 2.1.1 [95]. The A14 tune [96] and the NNPDF2.3LO PDF set [97] is used for all signal samples.

Most of the samples are simulated at NLO accuracy. The ggF signal sample was simulated at NNLO accuracy in QCD using the NNLOPS prescription.

5.3.2 Background Samples

There are various processes that produce two muons in the final state and must be considered as backgrounds for the Higgs to dimuon decay. The main background

process is Drell-Yan (DY) $Z \to \mu\mu$. Other significant background processes include electroweak $Z \to \mu\mu$ production and top or diboson processes that result in at least two muons.

The DY and EW $Z \to \mu\mu$ samples were generated with SHERPA 2.2.14 [98] at NLO accuracy for up to two partons, and to LO-accuracy for up to five partons. Parton showering is also modelled with SHERPA. The NNPDF3.0NNLO PDF set [99] is used.

The $t\bar{t}$ events were modelled using POWHEG-BOX v2 [92] at NLO with the decays of b- and c- hadrons treated by EVTGEN 2.1.1 [95]. The parton showering is modelled with PYTHIA 8.308 [31] with the A14 tune and the NNPDF2.3LO PDF set [97].

The single top samples and associated production of top quarks with W bosons (tW) samples were modelled using POWHEG-BOX v2 [92] at NLO in QCD using the NNPDF3.0NLO PDF set [99]. The events were showered with PYTHIA 8.308 [31] using the A14 tune and the NNPDF2.3LO PDF set [97].

The diboson samples with production and decays of two vector bosons (ZZ, ZW and WW) are simulated with SHERPA 2.2.14 or SHERPA 2.2.16 [98] at NLO accuracy in QCD for up to one additional parton and at LO accuracy for up to three additional parton emissions. The parton showering is modelled by SHERPA, using the NNPDF3.0NNLO set of PDFs [99].

The $gg \to ZZ \to 4l$ sample is modelled using POWHEG-BOX v2 [92] with the decays of b- and c- hadrons treated by EVTGEN 2.1.1 [95]. Parton showering for these events is modelled with PYTHIA 8.313 [31] with the A14 tune and the NNPDF2.3LO PDF set [97].

The $t\bar{t}W$ sample is simulated with SHERPA 2.2.14. The parton showering is modelled

by SHERPA, with the NNPDF3.0NNLO PDF set [99].

The $t\bar{t}Z$ sample is simulated with aMC@NLO [93, 94] and the decays of b- and c-hadrons are treated by EVTGEN 2.1.1 [95]. Events were showered with PYTHIA 8.308 [31] using the A14 tune and the NNPDF2.3LO PDF set [97].

5.3.3 ATLAS Detector Simulation

The interactions of the simulated events in the ATLAS detector are passed through a GEANT4 [100] simulation. Event reconstruction is then applied following the same procedure used for data. Effects due to pileup from multiple pp collisions are simulated using EPOS [101, 102] and PYTHIA with A3 tune [103]. These pileup interactions are then added to simulated samples to match the μ distribution for each MC campaign.

MC events are produced by sampling the scattering amplitude for a specific process over the full partonic phase space. The cross section will vary across this phase space, so simulated events are assigned an event weight to account for this. These assigned weights can also be negative, due to NLO corrections to processes.

5.3.4 Corrections to Simulated Samples

One key difference between data and MC simulation is that MC has a weight, w, assigned to each event. The final weight indicates how many data events each MC event corresponds to. It is desirable to have small weights, $|w| \ll 1$, for MC, so that there is a large MC sample to provide accurate predictions with. The final weight applied is based on an initial weight and several additional weights and corrections to help MC match more closely with real data and adjust for shortcomings in simulation.

The initial weight is determined by the MC generator when the sample is produced. These MC samples are generated by sampling the scattering amplitude over the full partonic phase space [104], so different event weights are provided based on the cross section for each region of the phase space. These initial weights also account for efficiencies in event filters used when generating events, and even be negative to account for interference terms from higher order Feynman diagrams [105].

A correction to this initial MC weight to account for differences between pileup in simulated samples and ATLAS data is done using *pileup reweighting* [106]. An additional weight is factored in to make the pileup profile of simulation match more closely with data.

Other additional corrections to the weight for each event are applied, called *scale factors*. These scale MC events to match data more closely based on the efficiency of different algorithms and selection criteria applied to the objects in each event. For example, the muon reconstruction efficiency represents the probability that a real muon produced in the detector successfully results in reconstructed muon candidate objects. The muon reconstruction efficiency is found to be slightly overestimated in the MC simulation [83], which leads to an overestimation of the event yield. To account for this discrepancy, a scale factor is applied to MC simulation such that the predicted event yield is reduced. Many different scale factors are applied to each event to improve a dedicated aspect of the MC simulation.

The final weight used for each MC event is the product of all the previously mentioned weights and scale factors and can range in size. Note that the only possible negative contribution to the event weights comes from the initial MC weight provided by the generator.

Chapter 6

Analysis

This chapter outlines the strategy for measuring the $H \to \mu\mu$ decay using Run-3 data collected by the ATLAS detector. This involves the selection of high quality events matching the expected signal of the $H \to \mu\mu$ decay, using the novel approach of neural networks trained with a distance correlation (DisCo) loss function to categorize these events, and modelling of the signal and background distributions in order to ultimately measure the Higgs signal.

The events that are considered for this analysis contain a series of physics objects that are reconstructed as described in Chapter 5. Each of these objects has an associated four momentum[1], $p = (E, p_x, p_y, p_z)$. These four momentum vectors can be combined to reconstruct different systems in the event. Most notably for this analysis, the four momentum of the dimuon system is given by, $p_{\mu\mu} = p_{\mu_1} + p_{\mu_2}$. The mass of the dimuon system is $m_{\mu\mu}$, which can be obtained from the four momentum using the energy-momentum relation, $E^2_{\mu\mu} = \vec{p}^{\,2}_{\mu\mu} + m^2_{\mu\mu}$. If these two muons were produced from

[1] $E_{\mathrm{T}}^{\mathrm{miss}}$ is an exception here. It only has a magnitude and azimuthal angle associated with it.

the decay of a Higgs boson, the mass should be around the Higgs mass, $m_{\mu\mu} \approx 125\,\text{GeV}$. Other systems can be reconstructed analogously in each event, such as the Z boson in a ZH decay event, and this provides additional features that can be provided to neural networks to improve their performance.

An outline of the analysis strategy used for the $H \to \mu\mu$ resonance search is given in Section 6.1. An overview of the selection applied to identify signal events and reduce background contributions is described in Section 6.2. A brief introduction to neural networks is presented in Section 6.3. This section also introduces the DisCo loss function (Section 6.3.2.1), enabling the use of neural networks for the $H \to \mu\mu$ analysis, which has traditionally been limited in previous analyses. Section 6.4 outlines how the events considered for this analysis are divided into mutually exclusive categories, and how neural networks are used to increase the significance of the $H \to \mu\mu$ measurement. Finally, Section 6.5 explains the signal and background modelling used to parametrize each category.

6.1 Analysis Strategy

The principle of this thesis is a resonance search for the $H \to \mu\mu$ decay. The dimuon invariant mass ($m_{\mu\mu}$) spectrum is expected to have a large, smoothly-falling background with a resonance peak around the Higgs mass, $m_H = 125.09\,\text{GeV}$ [13]. An illustration of the expected dimuon invariant mass spectrum in the signal region is shown in Figure 6.1a. The signal region (SR) for this analysis is defined as $m_{\mu\mu} \in [110, 160]\,\text{GeV}$, which contains the Higgs mass. To enhance the sensitivity of the $H \to \mu\mu$ search, the significance needs to be optimized. To first order, the significance is given by $Z = s/\sqrt{b}$, where s is the number of signal events in the signal window (SW) and b is the number

of background events in the SW. The SW is defined as $m_{\mu\mu} \in [120, 130]$ GeV. In order to extract the signal and background yields from the dimuon invariant mass spectrum, a fit is performed across the entire SR. If we consider a hypothetical situation where we expect 1000 background events in the SW ($b = 1000$), and we observe 1060 events in the SW of our dataset, we have an excess over the background expectation of $s = 60$. The significance of this excess becomes $Z = 60/\sqrt{1000} = 1.90\sigma$.

To increase the significance of this dataset, the events can be divided into mutually exclusive categories. With multiple categories, the total significance is given by the individual significances of each category added in quadrature [107],

$$Z = \sqrt{\sum_{i}^{N} Z_i}, \tag{6.1}$$

where Z_i is the significance of the i^{th} category.

If the hypothetical dataset is divided into two mutually exclusive categories as shown in Figure 6.1b, there will be an increase in significance. If one category has $s = 20$ and $b = 120$, and the second category has $s = 40$ and $b = 880$, the total significance will be $Z = \sqrt{\left(\frac{20}{\sqrt{120}}\right)^2 + \left(\frac{40}{\sqrt{880}}\right)^2} = 2.35\sigma$.

In order to estimate the background in each category a fit is performed to the sidebands of data events, as shown in Figure 6.2a. The sidebands are defined as $m_{\mu\mu} \in [110, 120] \cup [130, 160]$ GeV. In principle, MC simulation could be used to predict the background expectation in the SW, however these predictions come with large uncertainties. It is preferential to estimate the background directly from data and avoid any potential effects from mismodelling of the background processes.

One condition to parametrize the background from the data sidebands is to have

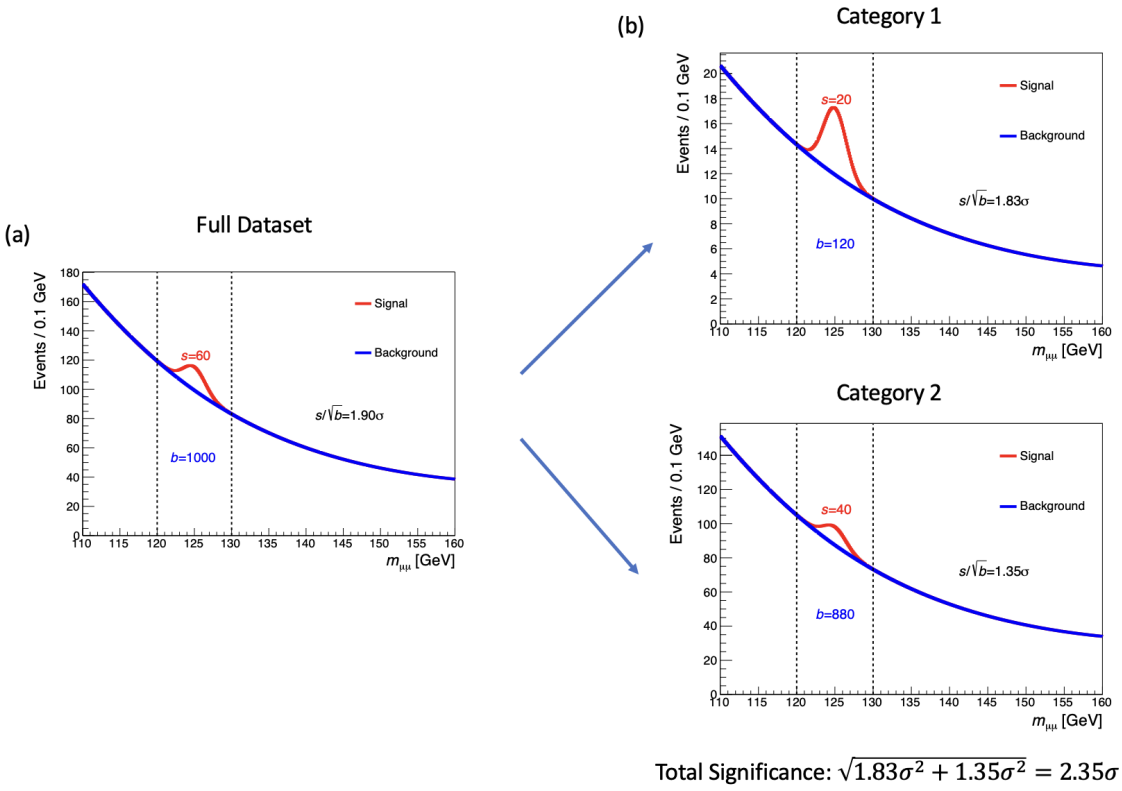


Figure 6.1: An illustration of the main approach of the $H \to \mu\mu$ analysis. (a): The $H \to \mu\mu$ signal events manifest as a resonance peak around the Higgs boson mass on top of a large continuous background, primarily from $Z \to \mu\mu$. A fit is performed across the range $m_{\mu\mu} \in [110, 160]$ GeV, and the expectation value of the background yield under the Higgs peak, b, is obtained from this fit. The significance from this dataset is 1.90σ. (b): By splitting this hypothetical dataset into mutually exclusive categories with different signal-to-background ratios, the overall significance can be improved. These two categories derived from the full dataset give a total significance of 2.35σ.

a smoothly falling background distribution that can be easily modelled using the sidebands. In this analysis, different functions are tested to model the background distribution. Figure 6.2a shows that two different commonly used functions, Epoly ($e^{(a_1 m_{\mu\mu}+\ldots+a_N m_{\mu\mu}^N)}$) and Power ($m_{\mu\mu}^{(a_0+a_1 m_{\mu\mu}^2+\ldots+a_N m_{\mu\mu}^N)}$) functions both model the background distribution well and would provide similar estimates for the background yield. Methods used to categorize $H \to \mu\mu$ events can dramatically shape the background distributions (see Section 6.3.2.1) and Figure 6.2b illustrates an example where the background distribution has a large bump around the SW. Even if more complex functions are used to model the background distribution using the sidebands, the resulting background yield estimates are very sensitive to the different functions. This is an important point to remember for this analysis, that the methods used to divide events into categories cannot introduce any significant shaping into the background distributions or else the signal cannot be extracted from the $m_{\mu\mu}$ spectrum.

6.2 Event Preselection

It is important that all events in the dataset are reconstructed properly, with physics objects satisfying the conditions outlined in Chapter 5, and that the events have the expected signature of $H \to \mu\mu$. This is achieved using a *preselection*, which includes a variety of criteria that events must pass in order to be considered for the analysis.

Triggers (Section 3.2.6) are used to collect events that look interesting for physics analyses with the ATLAS detector. For the Higgs to dimuon analysis, events that are collected by the single muon triggers, `HLT_mu24_ivarmedium_L1MU14FCH` or `HLT_mu50_L1MU14FCH` are considered. These triggers indicate the presence of a 24 GeV and 50 GeV trigger muon object respectively, built using limited information

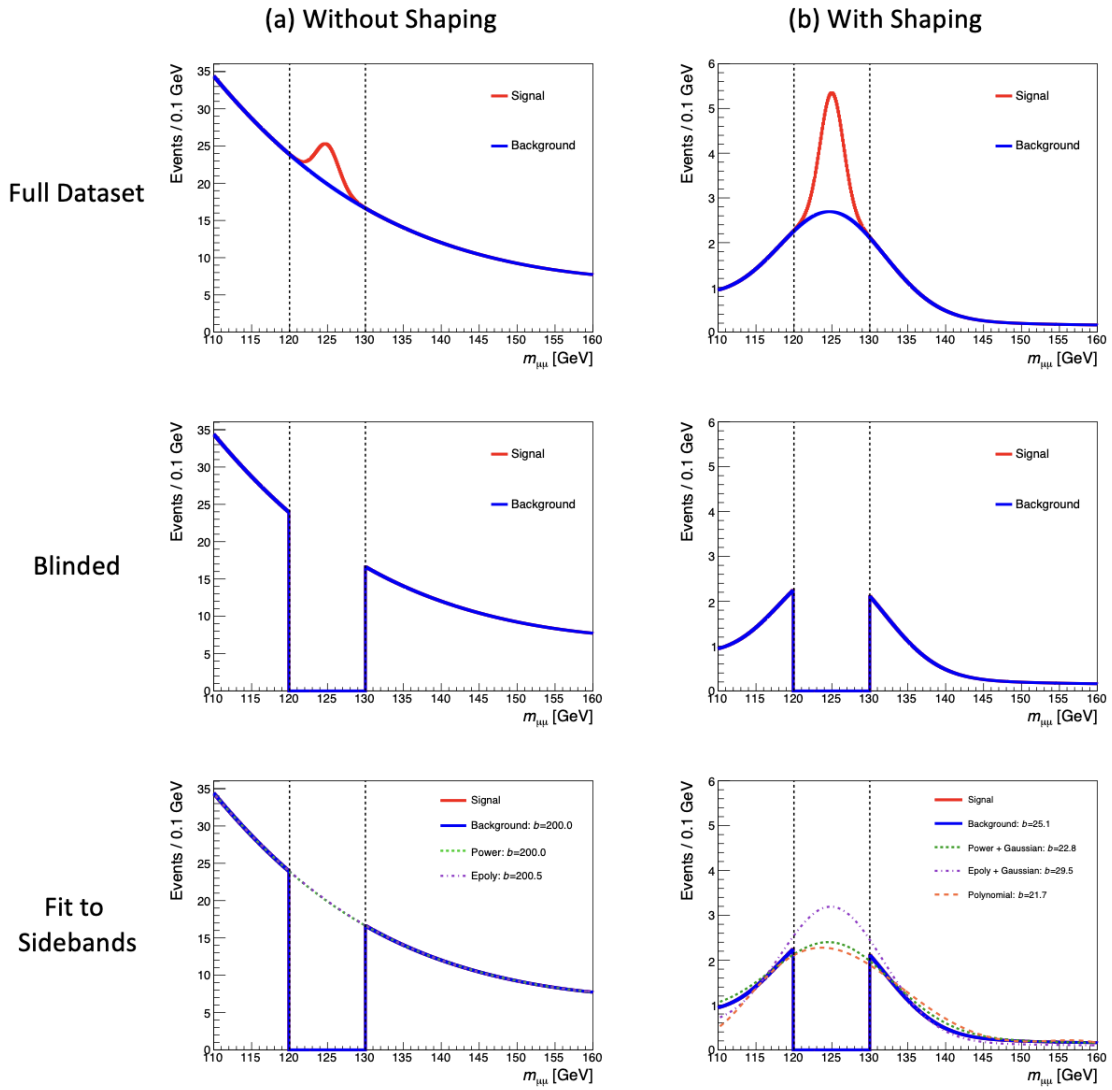


Figure 6.2: An illustration of the background estimation procedure for the $H \to \mu\mu$ categories. The numbers in the bottom two plots indicate the true background yield in $m_{\mu\mu} \in [120, 130]$ GeV and the estimated yield from fits to the sidebands. (a): Data is blinded in the window, $m_{\mu\mu} \in [120, 130]$ GeV, a fit is performed to the sidebands and the background yield under the Higgs peak is accurately estimated using two different functions. (b): If the category has significant shaping in the background, which can occur when using machine learning classifiers, the background yield under the signal peak will be very inaccurate as it becomes very sensitive to the assumed functional form used in the statistical fit.

from the detector.

All data events used in the analysis must pass the good runs list (GRL), to remove events recorded during time periods with unstable LHC beams or poor performance of the ATLAS detector. Data quality flags are also used to remove any events that have passed the GRL, but were affected by temporary detector problems. After applying the GRL selection, there is $165\,\text{fb}^{-1}$ of data available from 2022–2024 as outlined in Table 5.1.

The next key criteria for this analysis is the presence of two high-quality muons, as expected from $H \to \mu\mu$. The leading muon must have $p_{\text{T}} > 27\,\text{GeV}$ and the sub-leading muon must have $p_{\text{T}} > 15\,\text{GeV}$ to be considered for this analysis. The leading muon p_{T} threshold is motivated by the 24 GeV single muon trigger, which is fully efficient above 27 GeV. The leading reconstructed muon is required to match the muon object used by the trigger to ensure high-quality reconstructed events.

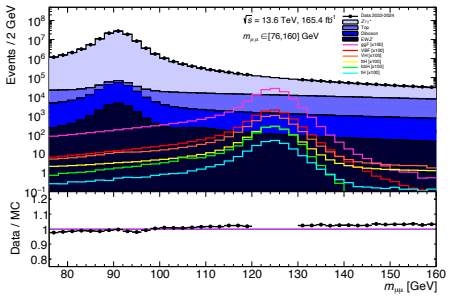
Events passing the aforementioned criteria are split into the Z-mass control region (CR) or the signal region (SR). The CR is defined as $m_{\mu\mu} \in [76, 106]\,\text{GeV}$ and is centred around the Z boson mass ($m_Z \approx 91\,\text{GeV}$). This region is used as a validation region to compare MC simulation with data after applying the selection used in this analysis. The SR region is defined as $m_{\mu\mu} \in [110, 160]\,\text{GeV}$, which is the region used to extract the Higgs signal and contains the Higgs mass (125.09 GeV). Note that until the analysis strategy was finalized, data events in $m_{\mu\mu} \in [120, 130]\,\text{GeV}$ were *blinded*. Blinding the data ensures that there is no unintended bias in the decision made when optimizing the analysis. This $m_{\mu\mu} \in [120, 130]\,\text{GeV}$ blinded region is referred to as the signal window (SW), and the region around it ($m_{\mu\mu} \in [110, 120] \cup [130, 160]\,\text{GeV}$) is referred to as the sidebands. The event yields passing preselection for data and the

main background samples are summarized in Table 6.1. After preselection there is a signal-to-background ratio of roughly 1:500 in the signal window.

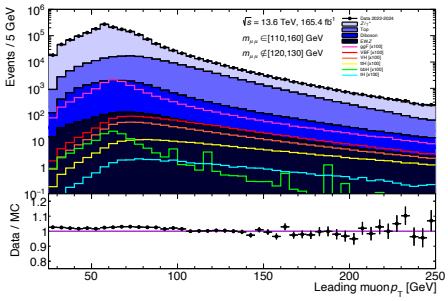
Table 6.1: Observed and MC predicted event yields in the Z-boson control region (CR), sideband (SB), and signal window (SW) for different subsets of the dataset passing preselection. The Z boson CR is defined by $m_{\mu\mu} \in [76, 106]$ GeV. The SB is defined by $m_{\mu\mu} \in [110, 160]$ GeV with all processes blinded: $m_{\mu\mu} \notin [120, 130]$ GeV. The SW is defined by $m_{\mu\mu} \in [120, 130]$.

$H \to \mu\mu$ selection	Preselection								
$H \to \mu\mu$ dataset	2022			2023			2024		
Region	CR	SB	SW	Z-CR	SB	SW	Z-CR	SB	SW
Data	19.58M	317.6k	96.4k	18.15M	294.0k	88.8k	70.41M	1.14M	347.8k
$Z \to \mu\mu$	19.21M	268.5k	81.8k	17.93M	250.9k	76.3k	71.59M	1.01M	305.7k
EW $Z \to \mu\mu$	9.56k	313	93.9	9.04k	293	88.2	41.4k	1.17k	354
$t\bar{t}$	36.9k	29.4k	8.60k	34.9k	27.8k	8.08k	139.3k	111.0k	32.3k
Single Top	3.69k	3.06k	881	3.50k	2.89k	835	13.9k	11.6k	3.33k
$t\bar{t}V$	487	62.7	17.6	462	58.7	16.7	1.85k	238	67.5
VV	37.7k	3.60k	1.07k	35.5k	3.36k	994	141.7k	13.5k	3.97k
EW VV	241	44.0	13.3	230	39.3	11.9	921	161	44.6
ggF: $H \to \mu\mu$	3.03	23.3	163	2.84	22.4	152	11.4	95.8	603
VBF: $H \to \mu\mu$	0.23	2.19	12.6	0.22	2.11	11.7	0.88	9.07	46.3
ttH: $H \to \mu\mu$	0.06	0.39	1.72	0.06	0.38	1.62	0.25	1.62	6.39
VH: $H \to \mu\mu$	0.19	1.29	6.52	0.18	1.25	6.07	0.71	5.30	24.1
bbH: $H \to \mu\mu$	0.03	0.20	1.57	0.03	0.22	1.61	0.11	1.00	6.41
tH: $H \to \mu\mu$	0.01	0.08	0.38	0.01	0.08	0.36	0.03	0.26	1.05
MC bkg total	19.29M	305.0k	92.4k	18.02M	285.3k	86.3k	71.93M	1.14M	345.7k
Data/MC bkg	1.015	1.041	1.043	1.007	1.030	1.029	0.979	0.999	1.006
MC signal total	-	-	185	-	-	173	-	-	687
Sig : bkg	-	-	1 : 499	-	-	1 : 498	-	-	1 : 504

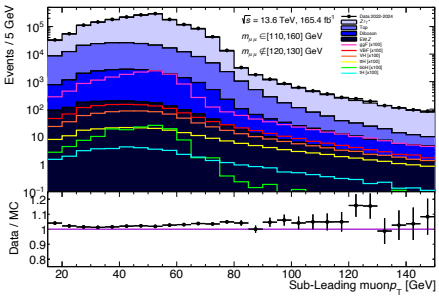
Comparison between Run-3 data and MC simulation for key variables can be seen in Figure 6.3. These plots show good agreement between the MC simulation and 165 fb^{-1} of Run-3 data collected from 2022–2024. It is important to note that MC simulation is only used in this analysis for the modelling of the $H \to \mu\mu$ signal (Section 6.5.1). The background is determined directly from data (Section 6.5.2) and background MC is not used. It is still important to have accurate MC simulation, as this is used to train machine learning classifiers to identify signal events in this thesis.



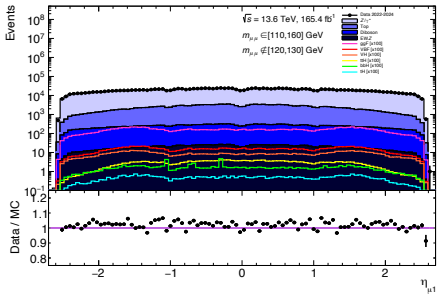
(a) Dimuon invariant mass across the CR and SR. Note that the data is blinded in the SW.



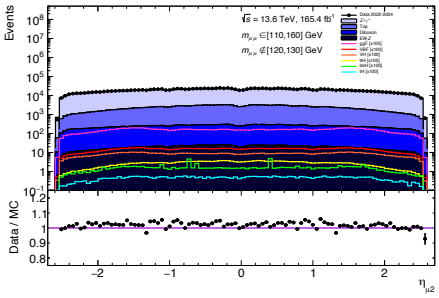
(b) Leading muon p_{T} in the SR.



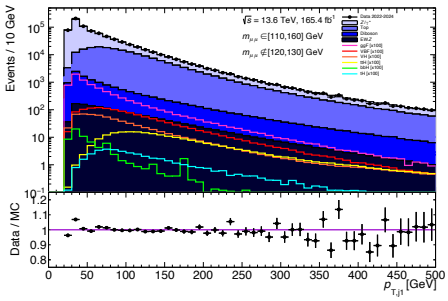
(c) Subleading muon p_{T} in the SR.



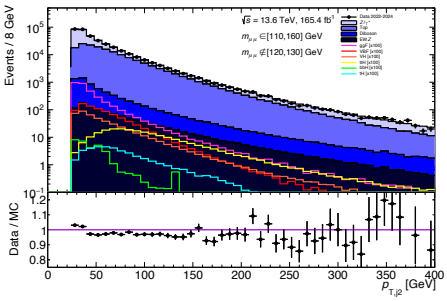
(d) Leading muon η in the SR.



(e) Subleading muon η in the SR.



(f) Leading jet p_{T} in the SR.



(g) Subleading jet p_{T} in the SR.

Figure 6.3: Comparison between Run-3 data and MC simulation for key variables in the signal region (SR) and control region (CR). MC simulation models the data well for different muon and jet variables.

6.3 Neural Networks

This section provides a brief introduction to neural networks (NNs) that are used to define categories and increase the significance of the $H \to \mu\mu$ analysis. This section draws on material from Ref. [108].

Machine learning (ML) is a branch of artificial intelligence where statistical algorithms learn patterns in data and apply this knowledge to predict features in other datasets without explicitly being programmed. This analysis uses a branch of ML called neural networks (NNs) to classify events as signal or background. The NN is trained using *supervised learning*, where it is given a series of input features, $\boldsymbol{x}$, with a series of known labels, $\boldsymbol{y}$, to provide an output, $f_{\text{NN}}(\mathbf{x}) = \hat{\boldsymbol{y}}$, that predicts the label.

NNs get their name because they are made up of artificial *neurons*, which are inspired by the mechanism of learning in biological neurons. These artificial neurons are provided a vector of input signals, $\boldsymbol{x}$, which are multiplied by a vector of weights, $\boldsymbol{w}$, and possibly shifted by a bias term, b, to calculate an output, $\hat{y}$. The architecture of a single neuron is shown in Figure 6.4, and its output is given by,

$$\hat{y} = \boldsymbol{w}^{\text{T}}\boldsymbol{x} + b. \tag{6.2}$$

A NN is composed of many neurons, which are composed of interconnected *layers*. Each layer has an input vector, $\boldsymbol{x}$, with a weight matrix, $\boldsymbol{W}$, and possibly a bias vector, $\boldsymbol{b}$, which calculates an output vector, $\hat{\boldsymbol{y}}$. In equation form,

$$\hat{\boldsymbol{y}} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}. \tag{6.3}$$

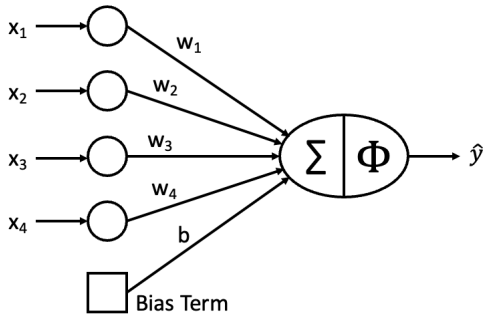


Figure 6.4: Architecture of a single neuron with inputs $\boldsymbol{x} = [x_1, x_2, x_3, x_4]$, weights $\boldsymbol{w} = [w_1, w_2, w_3, w_4]$, bias function b, activation function Φ, and output $\hat{y}$.

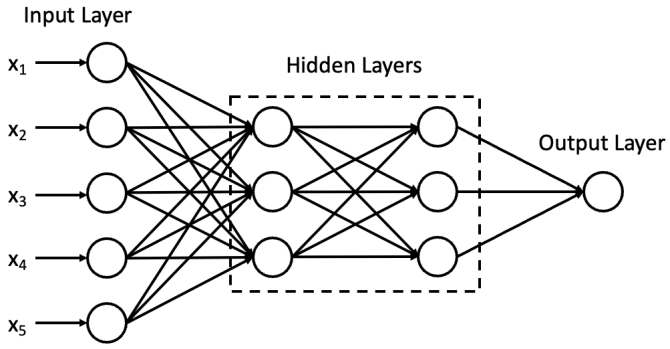


Figure 6.5: The basic architecture of a NN with two hidden layers and one output layer.

The output vector, $\hat{\boldsymbol{y}}$, from one layer of the NN can be provided as the input, $\boldsymbol{x}$, for another layer of the NN to create a *deep* neural network (DNN). A DNN consists of multiple layers of neurons that are interconnected, as seen in Figure 6.5. The individual neurons in these DNNs are commonly referred to as *nodes*. The layers in between the input layer and output layer are referred to as *hidden layers*.

For the $H \to \mu\mu$ analysis, the inputs provided to the NN are features from the physics objects in each event. These inputs are used to train a NN to distinguish between $H \to \mu\mu$ signal ($y = 1$) and background ($y = 0$) processes. The output from the NN is a prediction, $\hat{y} \in [0, 1]$, where a prediction close to 1 indicates a signal-like event and a prediction close to 0 indicates a background-like event. Categories can then be

defined based on the NN predictions assigned to events. For example, a category with events with $\hat{y} > 0.5$ can be defined. This category should have a higher s/b than a category defined as $\hat{y} < 0.5$.

6.3.1 Activation Functions

The operations in Equation 6.3 are linear and as a result the entire network behaves like a single linear transformation. Stacking multiple such layers would be equivalent to just one layer, since the linear functions could be mathematically combined without changing the output. To model more complex, non-linear relationships in the data, non-linear *activation functions* are essential between layers. These apply a non-linear transformation, Φ, to each layer,

$$\hat{\boldsymbol{y}} = \Phi(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}). \tag{6.4}$$

There are a variety of activation functions that can be used when training NNs. There are two activation functions that are used in the training of NNs in this thesis, the rectified linear unit (ReLU) activation function and the sigmoid activation function. The ReLU function has the form,

$$\Phi(x) = \max(0, x). \tag{6.5}$$

This activation function is useful for optimization of the NN and is used in the hidden layers. The sigmoid activation function has the form,

$$\Phi(x) = \frac{1}{1 + e^{-x}}. \tag{6.6}$$

The sigmoid activation function has a range of $(0, 1)$ which makes it useful for NNs used for classification. A value close to 0 represents data that is background-like, and a value close to 1 represents data that is signal-like.

6.3.2 Loss Functions

In order to train a NN to perform a task, there must be a metric to evaluate its performance. The performance of a NN is calculated using a *loss function*, $L(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are the parameters of the NN. During training the NN will work to minimize the loss, so this function usually measures how close the outputs from the NN are to the true values of what the NN is trying to predict. A commonly used loss function for NNs training to predict a binary outcome is the binary cross-entropy (BCE) function,

$$L(\boldsymbol{\theta}) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(f(\boldsymbol{x}_i;\boldsymbol{\theta})) + (1-y_i)\log(1-f(\boldsymbol{x}_i;\boldsymbol{\theta}))\right]. \tag{6.7}$$

The loss is evaluated over all events in the dataset and the closer the predictions of the NN ($f(\boldsymbol{x}_i;\boldsymbol{\theta})$) are to the true labels, y_i, the lower this BCE will be. This means that when the NN lowers the loss, the predictions should get closer to the true values and the NN will become more accurate.

6.3.2.1 DisCo Loss Function

As discussed in Section 6.1, the extraction of the $H \to \mu\mu$ signal relies on a precise estimate of the background event yield under the signal peak in the $m_{\mu\mu}$ distribution. The background is constrained by smoothly falling data events in the sidebands and when splitting the data into categories it is crucial to avoid any shaping of the

background.

The dimuon invariant mass has very good separating power between signal and background events due to the distinct Higgs peak around 125 GeV. If a NN is provided $m_{\mu\mu}$, events around the Higgs mass will have a higher score, which will introduced shaping into the mass distribution (Figure 6.2). Even if $m_{\mu\mu}$ is not provided directly to the NN as an input during training, modern NNs are very powerful and can learn the mass if they have access to other input features that are correlated with the mass. In order to use NNs for the $H \to \mu\mu$ analysis there cannot be correlations between $m_{\mu\mu}$ and the NN predictions.

In the previous ATLAS $H \to \mu\mu$ analysis [3], sculpting of the mass was prevented by carefully selecting the variables used to train their classifiers and avoiding variables correlated with the mass. Avoiding these variables limits the information that can be provided to the classifier, which should negatively impact its performance.

An alternative method for NNs is to include a new term in the loss function that will penalize any correlations between the NN predictions and the dimuon invariant mass. This can be done using a distance correlation (DisCo) loss term,

$$L_{\text{DisCo}} = \text{dCorr}^2_{\boldsymbol{y}=0}(\hat{\boldsymbol{y}}, \boldsymbol{m_{\mu\mu}}). \tag{6.8}$$

This calculates the distance correlation for only background events ($\boldsymbol{y} = 0$), but can be expanded to prevent correlation with signal events. Using this DisCo loss term allows the full set of variables to be provided to the NN during training, as the loss function will prevent any issues with shaping. DisCo [109–112] is a statistical measure that is sensitive to non-linear correlations between two samples, which makes it more powerful than other measures of linear correlations, such as Pearson's correlation. The

use of a DisCo loss term is relatively new, but has been used for a few applications in particle physics, such as identifying W bosons [113] and searching for the Higgs to dimuon decay [114].

Consider the two vectors $(\boldsymbol{x}, \boldsymbol{y})$ of length N, compute pairwise distance matrices, $a_{ij} = |x_i - x_j|$ and $b_{ij} = |y_i - y_j|$, and double-centre them, $A_{ij} = a_{ij} - \bar{a}_{i\cdot} - \bar{a}_{\cdot j} + \bar{a}_{\cdot\cdot}$ and $B_{ij} = b_{ij} - \bar{b}_{i\cdot} - \bar{b}_{\cdot j} + \bar{b}_{\cdot\cdot}$. Here $\bar{a}_{i\cdot}$ is the mean over row i, $\bar{a}_{\cdot j}$ is the mean over column j and $\bar{a}_{\cdot\cdot}$ is the mean over all entries. The distance correlation is then given by,

$$\mathrm{dCorr}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\mathrm{dCov}(\boldsymbol{x}, \boldsymbol{y})}{\sqrt{\mathrm{dCov}(\boldsymbol{x}, \boldsymbol{x})\mathrm{dCov}(\boldsymbol{y}, \boldsymbol{y})}}, \tag{6.9}$$

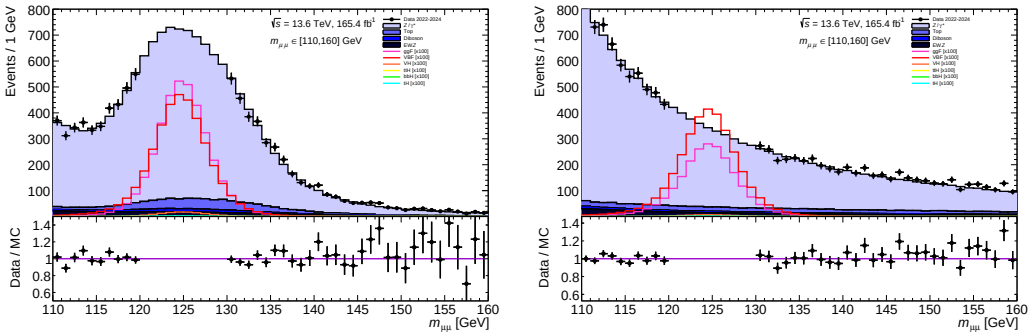
where dCov is the distance covariance,

$$\mathrm{dCov}^2(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{N^2} \sum_{i,j} A_{ij} B_{ij}. \tag{6.10}$$

This analysis uses a combination of BCE and a DisCo loss term when training NNs,

$$L_{\mathrm{Total}} = L_{\mathrm{BCE}}(\hat{\boldsymbol{y}}, \boldsymbol{y}) + \lambda \cdot \mathrm{dCorr}^2_{\boldsymbol{y}=0}(\hat{\boldsymbol{y}}, \boldsymbol{m_{\mu\mu}}). \tag{6.11}$$

Here $= L_{\mathrm{BCE}}$ is the BCE term from Equation 6.7 and λ is a *hyperparameter* to control the contribution of the DisCo loss term relative to the BCE loss term. The λ hyperparameter is adjusted for each NN trained in this analysis. When a NN is trained with the loss function in Equation 6.11, the BCE term separates signal and background events, while the DisCo loss term ensures there is no correlation between the NN predictions and the dimuon invariant mass and prevents any shaping of the background distributions. A comparison of the $m_{\mu\mu}$ distributions for two categories defined using a NN trained with and without a DisCo loss term can be seen in Figure 6.6.



(a) $\lambda = 0.0$ (b) $\lambda = 100.0$

Figure 6.6: The $m_{\mu\mu}$ distributions for events categorized using a NN trained with a DisCo hyperparameter of (a): $\lambda = 0$ (no DisCo loss term) and (b): $\lambda = 100$. Neither NN was explicitly provided $m_{\mu\mu}$ as an input training variable. Introducing a DisCo loss term to penalize correlations with $m_{\mu\mu}$ prevents shaping of the background distributions.

6.3.3 NN Training

The process that NNs generally use to learn is called the backpropagation algorithm. This is how the parameters of a NN training on a dataset are optimized to minimize the loss. An overview of the backpropagation method is described below:

1. Forward phase: For a fixed input, x_i, compute the output of the NN layer by layer, using the current values of the parameters (weights and biases). This results in a predicted output, $\hat{y}_i$. The loss is then computed by comparing this to the true target value, y_i using the loss function, $L(\boldsymbol{\theta})$.

2. Backward phase: Compute the gradient of the loss with respect to the chosen parameters, $\nabla_{\boldsymbol{\theta}} L(\theta)$. This allows the network to determine how each parameter contributes to the loss. It is called the backward phase because the derivatives of the loss with respect to weights near the output are easier to compute. These derivatives are computed first and the chain rule is used to help compute the

more complex derivatives as we move away from the output.

3. Parameter update: Update the weights in the negative direction of the gradient, $\boldsymbol{\theta}' = \boldsymbol{\theta} - \gamma \nabla_{\boldsymbol{\theta}} L(\theta)$, which will be the steepest direction of descent. Here γ is the learning rate, which determines how fast the parameters are updated towards the minimum of the loss function.

The example above describes backpropagation and computing the gradient of the loss for a single fixed input and output, which is referred to as *stochastic gradient descent*. To lower the loss for the NN you would need to compute the gradient of the loss across all points, which is computationally very expensive. Since calculating the loss across all points is impractical, a method called *mini-batch gradient descent* is used for training NNs. Instead of using a single point, a *batch* of points is used in each update following the process described above. The number of points used in each update can be adjusted and is referred to as the *batch size*. Mini-batch gradient descent works well because the data used for training NNs usually has a high level of redundancy and the gradient obtained from a sample of points is usually very accurate. This allows a good trade off for increasing the speed of calculations, while still acquiring an accurate gradient to lower the loss during training.

The backpropagation method is repeated in batches until all of the data provided to the NN for training has been passed through. Each time the full training dataset is passed through the NN is referred to as an *epoch*. NNs are usually trained over several epochs to continue decreasing the loss.

One problem with NNs is *overtraining*, where a model learns statistical features specific to the training data. The model will perform very well at predicting the outcome of data it is training on, but does not generalize well to additional data from the same

distribution. In order to prevent overtraining, the dataset is randomly divided into two subsets, which are provided to the NN. The first set of data is the *training data*, which the NN can see and trains on using the backpropagation method. The second set of data is *validation data*, which is not directly used in training. The loss of the validation data will be calculated after each epoch and if the validation loss stops decreasing, training will stop. When this occurs and NN training is halted before the set number of epochs have elapsed, it is referred to as *early stopping*.

Another method used to help reduce overtraining is called *dropout*. When dropout is enabled during training, certain nodes are turned off during the backpropagation process preventing them from contributing. This occurs for a random fraction of nodes (usually 20% – 50%) during each epoch. In order to compensate for the nodes that have been turned off, the remaining nodes are scaled up by a factor of $1/(1-p)$, where p is the probability of a node being dropped out. Dropout has been shown to decrease overtraining and make NNs generalize better to data not used in training [115].

6.4 Categorization Scheme

After preselection, events are divided into distinct categories that each target a specific $H \to \mu\mu$ final state signature, which differ depending on the production mode. For each signature, NNs are trained to discriminate between signal and background events. The predicted labels, $\hat{y}$ from the NNs are used to define categories that will increase the overall significance of the $H \to \mu\mu$ signal.

Table 6.2: The 4-fold training method used. Each NN is trained on two folds of data, validated on one fold and tested on one fold.

Event Number % 4	0	1	2	3
Fold 0	Testing Set	Training Set	Training Set	Validation Set
Fold 1	Validation Set	Testing Set	Training Set	Training Set
Fold 2	Training Set	Validation Set	Testing Set	Training Set
Fold 3	Training Set	Training Set	Validation Set	Testing Set

6.4.1 Neural Network Setup

NNs with a DisCo loss function are used to discriminate between signal and background to enhance the analysis sensitivity. For each Higgs production mode, a dedicated binary neural network is trained on simulated MC events, with the corresponding signal samples and the dominant background processes provided for training.

The NNs are trained using k-fold cross-validation with $k = 4$. The events are divided into 4 folds of similar sizes. This is done using the unique *event number* (Event Number % 4) assigned to each event during data collection. The event numbers have no correlation with the event kinematics and this method produces statistically independent datasets of equal size. Four separate NNs are then trained using two of the folds as training data, one fold as validation data and one fold as testing data as seen in Table 6.2. The NN is trained directly on the training sets and validation accuracy and loss are calculated using the validation set. Each NN then assigns scores to all events in their testing set. This ensures that all events used in the analysis are assigned a score by a NN that did not train on that event, which prevents bias from overtraining.

The NNs are trained using Keras [116] and Tensorflow [117]. Each NN is a sequential model with a stack of layers, where each layer has one input tensor and one output

tensor. The NNs have three hidden layers with ReLU activation units and a dropout of 25%. The number of nodes for each hidden layer is based on the number of training variables used (N_{Train}), with the first layer having $3 \cdot N_{\text{Train}}$ nodes, the second layer having $2 \cdot N_{\text{Train}}$ nodes and the third layer having $1 \cdot N_{\text{Train}}$ nodes. The output layer has a sigmoid activation unit.

Training occurs for up to 400 epochs with early stopping enabled based on the validation loss. A batch size of 4096 is used for all NNs unless stated otherwise. A relatively large batch size needs to be used for correlations between the dimuon invariant mass and NN predictions to be accurately predicted by the NN during training. The DisCo loss hyperparameter, λ, that controls the contribution of the DisCo loss term relative to the BCE loss term varies for each NN.

6.4.2 $t\bar{t}H$ Categories

The $t\bar{t}H$ categories are particularly relevant to this thesis as NNs with a DisCo loss function were used to enhance the sensitivity for these categories in the most recent ATLAS $H \to \mu\mu$ analysis [2]. The approach detailed in this thesis was used for the $t\bar{t}H$ categories and produced a better expected significance compared to other machine learning techniques such as boosted decision trees that were proposed by researchers from other institutes.

When the $t\bar{t}H$ process occurs there are many final state particles that produce signals in many parts of the detector. In addition to the two muons arising from the Higgs boson decay, each top quark will produce three final state quarks or leptons. This results in many features that can be used for training a NN, giving it many handles for discriminating between signal and background events.

To be considered for the $t\bar{t}H$ categories, events must pass the preselection outlined in Section 6.2 and have at least one b-jet. Additional muons can be created from the top quark decay chain, resulting in ambiguity for the Higgs candidate muons. When events have more than two muons, the two leading oppositely charged muons are assigned to the Higgs boson candidate. This method assigns the correct two muons to the Higgs candidate 91% of the time in $t\bar{t}H$ events. The observed and MC predicted yields passing $t\bar{t}H$ selection can be seen in Table 6.3.

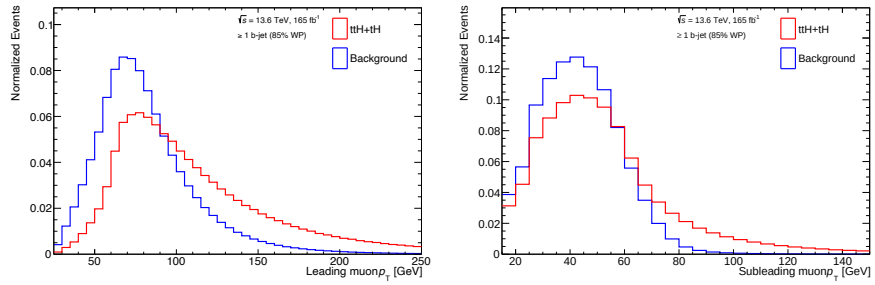
Using the simulated events fulfilling the selection, NNs are trained to discriminate between signal and background events. The NNs are trained on $t\bar{t}H$ and tH signal events against DY, $t\bar{t}$, single-top, $t\bar{t}V$ and diboson backgrounds. The variables used for training the NN are shown in Table 6.4. The MC signal and background distributions for the training variables are shown in Figures 6.7, 6.8, 6.9, 6.10 and 6.11. A DisCo hyperparameter of $\lambda = 75.0$ and batch size of 20,000 were used during training.

Table 6.3: Observed and MC predicted event yields in the Z-boson control region (CR), sideband (SB), and signal window (SW) for events passing $t\bar{t}H$ selection. The Z boson CR is defined by $m_{\mu\mu} \in [76, 106]$ GeV. The SB is defined by $m_{\mu\mu} \in [110, 160]$ GeV with all processes blinded: $m_{\mu\mu} \notin [120, 130]$ GeV. The SW is defined by $m_{\mu\mu} \in [120, 130]$.

$H \to \mu\mu$ selection	$t\bar{t}H$: At least one b-jet		
$H \to \mu\mu$ dataset	2022-2024		
Region	CR	SB	SW
Data	2.60M	188.5k	55.2k
$Z \to \mu\mu$	2.65M	42.8k	12.9k
EW $Z \to \mu\mu$	3.15k	113	33.3
$t\bar{t}$	183.3k	145.8k	42.5k
Single Top	14.8k	12.4k	3.60k
$t\bar{t}V$	2.50k	318	89.8
VV	15.0k	699	202
EW VV	119	18.1	5.26
ggF: $H \to \mu\mu$	0.80	5.24	27.1
VBF: $H \to \mu\mu$	0.08	0.76	3.68
ttH: $H \to \mu\mu$	0.33	2.16	8.76
VH: $H \to \mu\mu$	0.11	0.98	4.51
bbH: $H \to \mu\mu$	0.03	0.34	2.18
tH: $H \to \mu\mu$	0.04	0.32	1.31
MC bkg total	2.87M	202.2k	59.4k
Data/MC bkg	0.905	0.933	0.930
MC signal total	-	-	46.2
Sig : bkg	-	-	1 : 1.28k

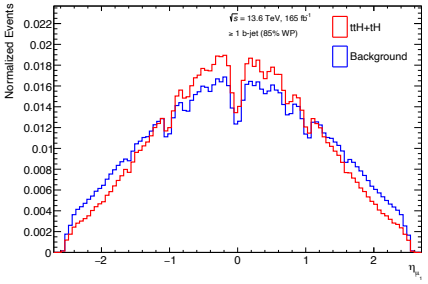
Table 6.4: The 30 variables used in training the NN for the $t\bar{t}H$ category.

Variable	Description
$p_{\mathrm{T}}^{\mu_1}$	Transverse momentum of the leading muon
$p_{\mathrm{T}}^{\mu_2}$	Transverse momentum of the subleading muon
η_{μ_1}	Pseudorapidity of the leading muon
η_{μ_2}	Pseudorapidity of the subleading muon
ϕ_{μ_1}	Azimuthal angle of the leading muon
ϕ_{μ_2}	Azimuthal angle of the subleading muon
N_{muons}	The number of muons
$\cos\theta^*$	The lepton decay angle $\cos\theta^*$ in the Collins–Soper frame [118]
$p_{\mathrm{T}}^{j_1}$	Transverse momentum of the leading jet
$p_{\mathrm{T}}^{j_2}$	Transverse momentum of the subleading jet
η_{j_1}	Pseudorapidity of the leading jet
η_{j_2}	Pseudorapidity of the subleading jet
$N_{\mathrm{track}}^{j_1}$	Number of tracks in the leading jet with $p_{\mathrm{T}} > 0.5$ GeV
$N_{\mathrm{track}}^{j_2}$	Number of tracks in the subleading jet with $p_{\mathrm{T}} > 0.5$ GeV
m_{jj}	Invariant mass of the dijet system
H_{T}	Scalar sum of the transverse momentum of all jets
$N_{j\text{-central}}$	Number of central jets ($\lvert\eta\rvert < 2.5$)
$N_{b\text{-jet}}$	Number of b-jets (85% WP)
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum
$N_{\mathrm{electrons}}$	Number of electrons
$p_{\mathrm{T}}^{\ell_3}$	Transverse momentum of the third lepton
$p_{\mathrm{T}}^{\ell_4}$	Transverse momentum of the fourth lepton
m_{sub}	Subleading mass of the third-leading muon and the leading or subleading muon of opposite charge
$m_{\ell_3\ell_4}$	Dilepton mass of the third- and fourth- leading leptons
$\Delta\phi(\ell_3, \ell_4)$	Azimuthal angle between third and fourth leptons
$\Delta\eta(H, \ell_3)$	Pseudorapidity separation between the dimuon system and the third lepton
m_{lep}^t	Transverse mass of leptonic t decay reconstructed from the third leading lepton, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, and a b-jet
m_{had}^t	Mass of hadronic t decay reconstructed from three jets, one of which is a b-jet
m_{lep}^W	Transverse mass of leptonic W decay reconstructed using Equation 6.13
m_{had}^W	Mass of hadronic W decay reconstructed from the two non b-tagged jets used to reconstruct the hadronic t mass

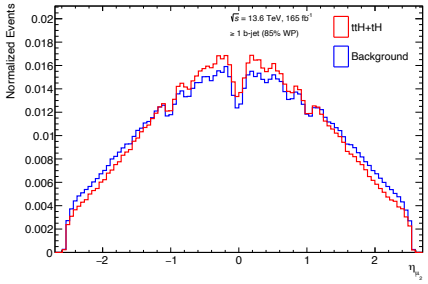


(a) Leading muon p_{T}.

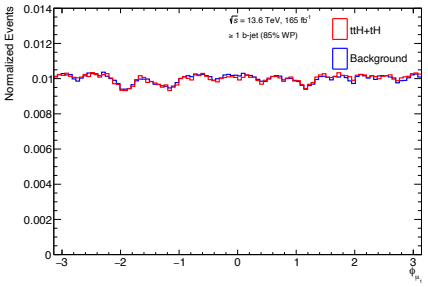
(b) Subleading muon p_{T}.



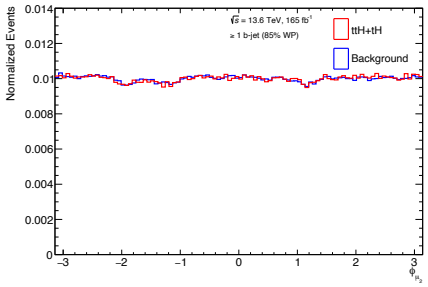
(c) Leading muon η.



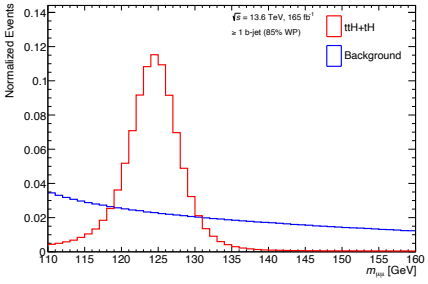
(d) Subleading muon η.



(e) Leading muon ϕ.

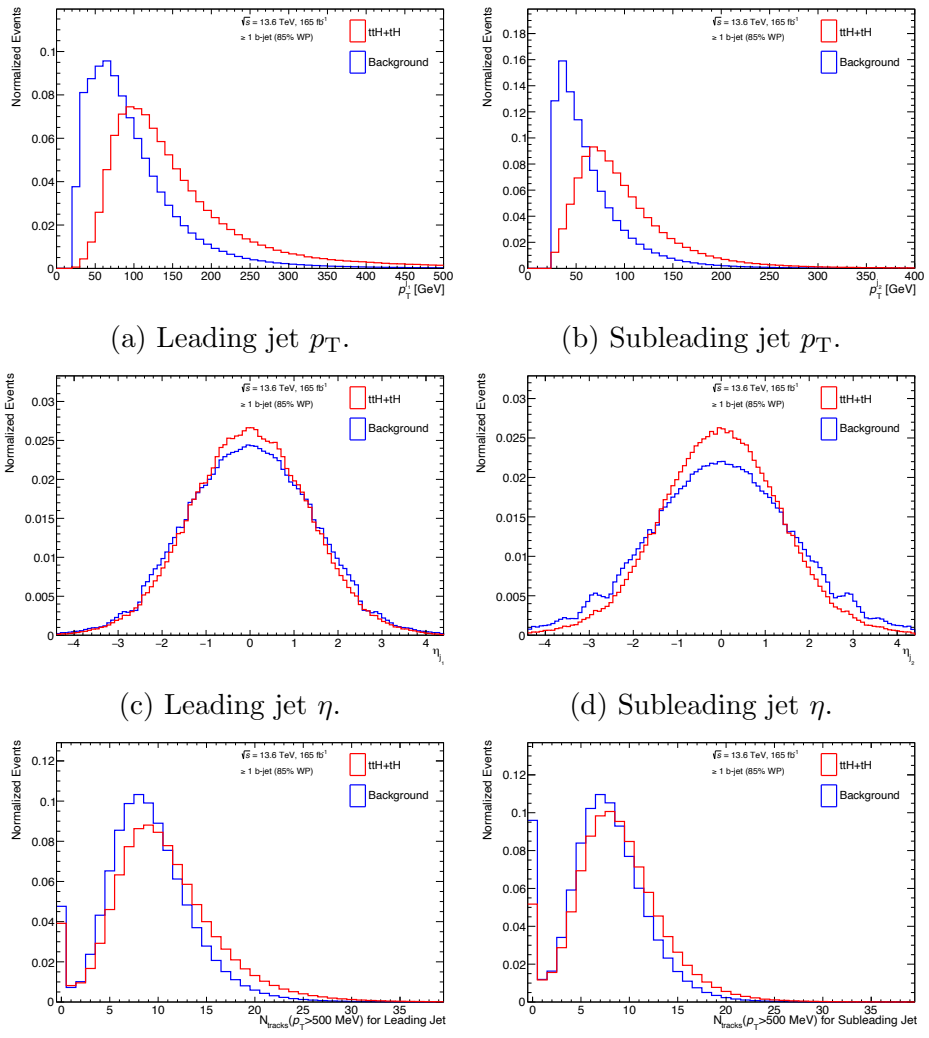


(f) Subleading muon ϕ.



(g) Higgs candidate $m_{\mu\mu}$.

Figure 6.7: Distributions of features obtained from muons that are used to train the NN $t\bar{t}H$ classifier. Separate distributions are shown for the MC simulated signal and background samples. The muons from $t\bar{t}H$ and tH signal events tend to have a higher transverse momentum than background processes. The NN classifier is provided $m_{\mu\mu}$ for the DisCo loss function, but it does not use the mass as a training variable.

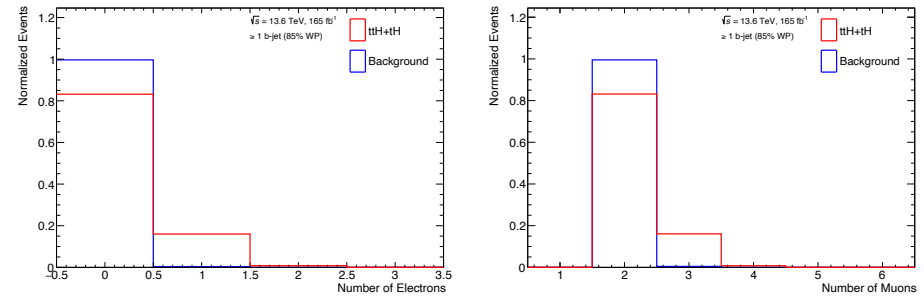


(a) Leading jet p_{T}. (b) Subleading jet p_{T}.

(c) Leading jet η. (d) Subleading jet η.

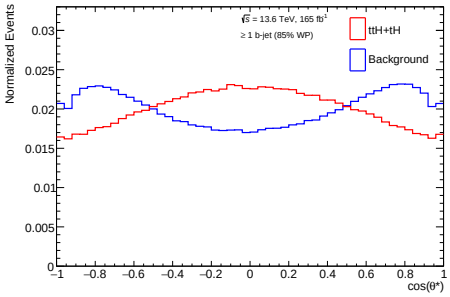
(e) Leading jet track multiplicity. (f) Sub-leading jet track multiplicity.

Figure 6.8: Distributions of features obtained from jets that are used to train the NN $t\bar{t}H$ classifier. Separate distributions are shown for the MC simulated signal and background samples. The $t\bar{t}H$ and tH signal events tend to have high p_{T} jets that are more central and have a slightly higher track multiplicity.

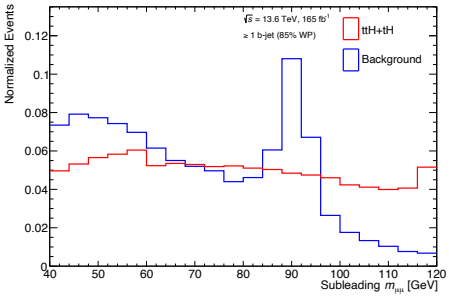


(a) Electron multiplicity.

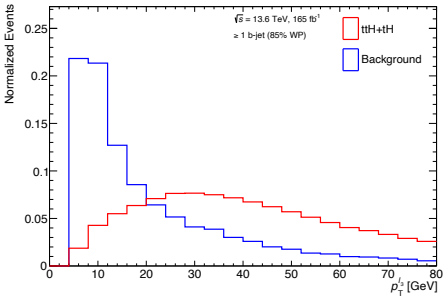
(b) Muon multiplicity.



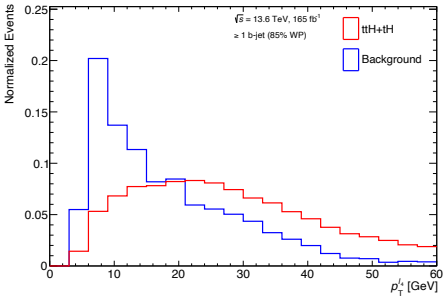
(c) Dimuon system $\cos\theta^*$.



(d) Subleading $m_{\mu\mu}$.

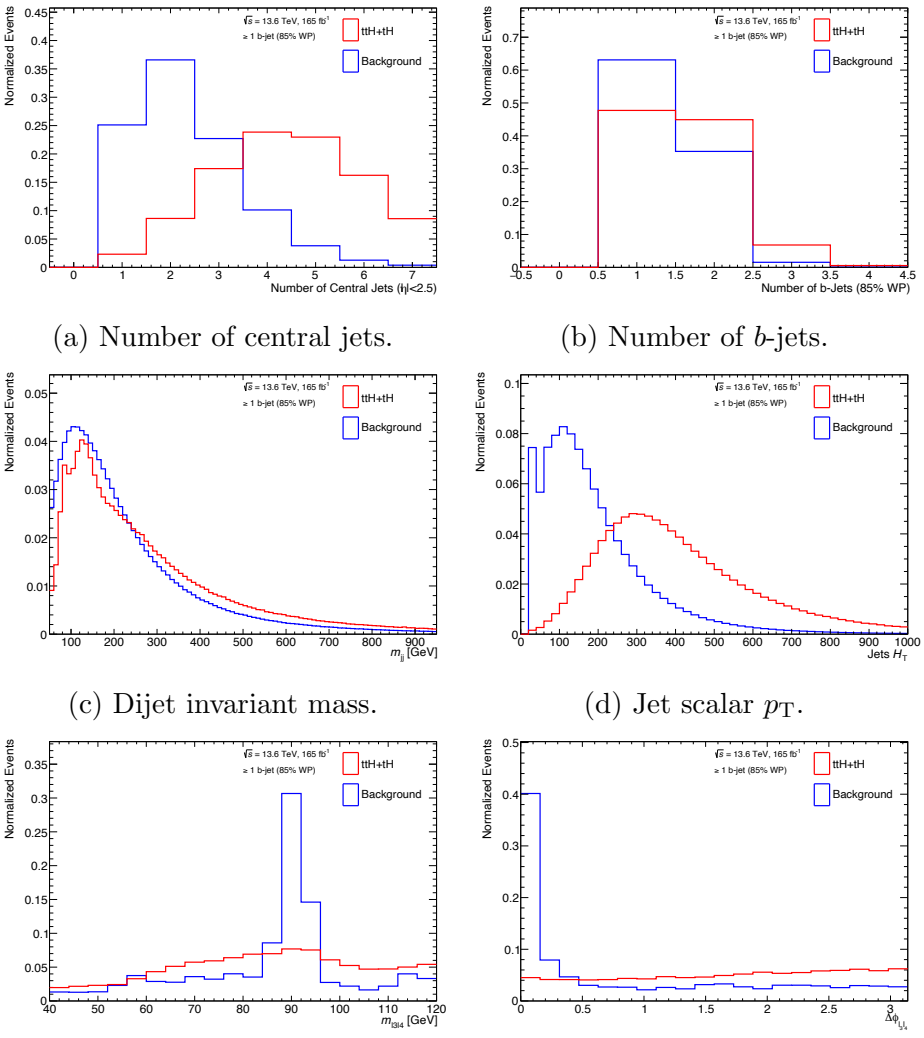


(e) Third lepton p_{T}.



(f) Fourth lepton p_{T}.

Figure 6.9: Distributions of additional leptonic features that are used to train the NN $t\bar{t}H$ classifier. Separate distributions are shown for the MC simulated signal and background samples. Additional leptons from $t\bar{t}H$ and tH signal events tend to have a higher p_{T} than background events, which is expected due to the large centre-of-mass for $t\bar{t}H$. The subleading invariant mass has a distinct peak around $m_Z \approx 91\,\mathrm{GeV}$ in background events due to $ZZ \to \mu\mu\ell\ell$ and $t\bar{t}V$ with $Z \to \ell\ell$. The cosine of the lepton decay angle in the Collins-Soper frame has a distinct difference in distribution between signal and background events.



(a) Number of central jets.

(b) Number of b-jets.

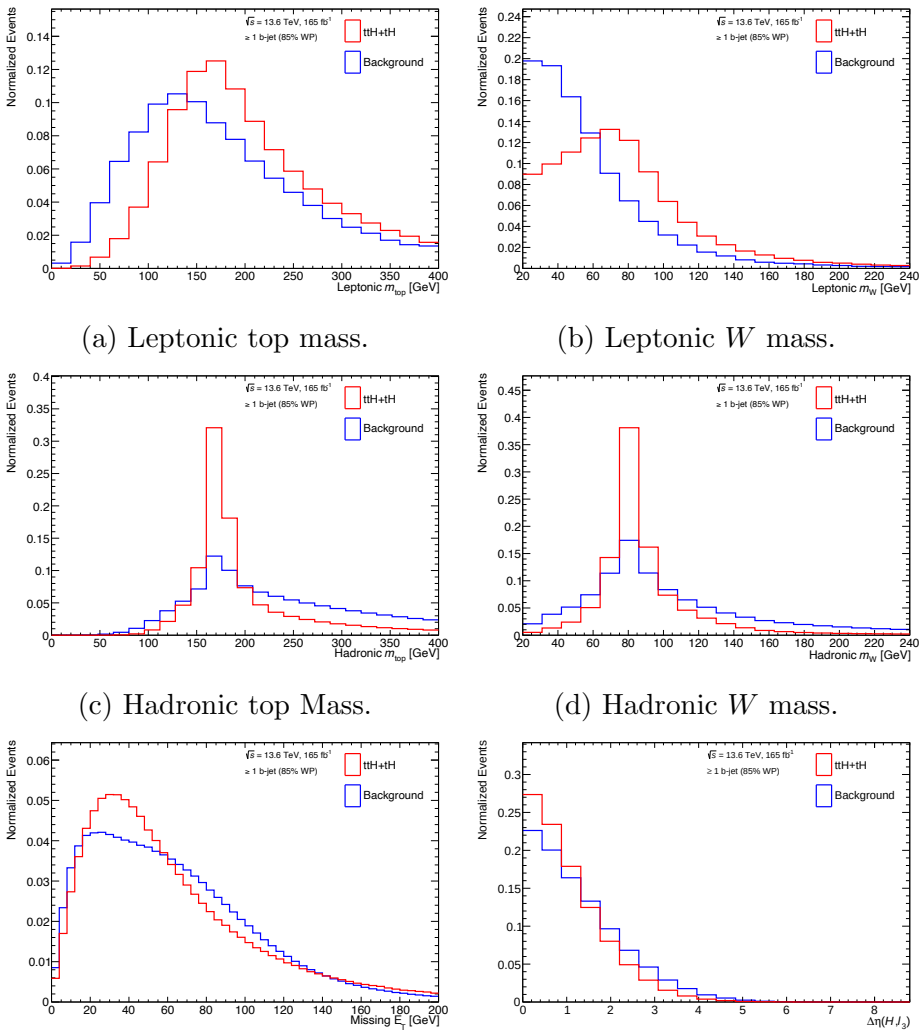
(c) Dijet invariant mass.

(d) Jet scalar p_{T}.

(e) Sub-leading lepton mass.

(f) Third and fourth lepton $\Delta\phi$.

Figure 6.10: Distributions of additional features obtained from leptons and jets that are used to train the NN $t\bar{t}H$ classifier. Separate distributions are shown for the MC simulated signal and background samples. $t\bar{t}H$ and tH signal events tend to have more b-jets, more central jets, and a larger jet p_{T} sum compared to background processes. The invariant mass of the third- and fourth-leading leptons has a distinct peak around $m_Z \approx 91\,\mathrm{GeV}$ due to $ZZ \to \mu\mu\ell\ell$ and $t\bar{t}V$ with $Z \to \ell\ell$. The azimuthal separation of the third and fourth leptons tends to to be smaller for signal events compared to background events.



(a) Leptonic top mass.

(b) Leptonic W mass.

(c) Hadronic top Mass.

(d) Hadronic W mass.

(e) Missing transverse momentum.

(f) $\Delta\eta$ between H candidate and ℓ_3.

Figure 6.11: Additional features used for training NNs passing $t\bar{t}H$ category selection. The leptonic and hadronic top quark and W boson masses for $t\bar{t}H$ and tH peak around the known top quark and W boson masses. Signal events tend to have lower $E_{\mathrm{T}}^{\mathrm{miss}}$ and a smaller η separation between the Higgs candidate and the third lepton.

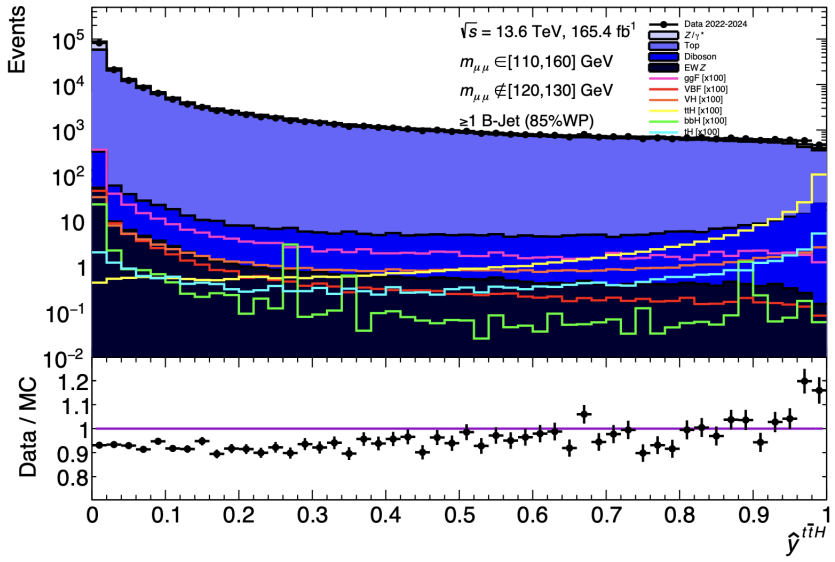


Figure 6.12: Distribution of the $t\bar{t}H$ NN predictions, $\hat{y}^{t\bar{t}H}$, for signal and background MC compared with data in the sidebands for events passing $t\bar{t}H$ selection. The NN is used as a classifier by dividing events into categories based on these predictions.

A comparison of the NN scores, $\hat{y}^{t\bar{t}H}$, between data and MC simulation after training on $t\bar{t}H$ events can be seen in Figure 6.12.

6.4.3 VH4L Category

The VH4L category targets ZH Higgs-strahlung production where the Z boson decays to a pair of leptons. Events considered for this category must have an additional pair of oppositely charged muons or electrons. To be considered for the VH4L category, there must be a pair of oppositely charged leptons with mass, $m_{\ell\ell} \in [75.5, 102]$ GeV, for the Z boson candidate. The second pair of oppositely charged leptons assigned to the Higgs candidate must not have a mass in this range. A b-jet veto is also applied for VH4L events. Additional selection requiring well isolated leptons is applied to reduce backgrounds:

- Lepton isolation: Leptons are required to have a minimum angular separation of 0.2, $\Delta R(\ell_i, \ell_j) > 0.2$.

- Lepton pair isolation: The Higgs candidate and Z candidate must have a minimum angular separation of 0.05, $\Delta R(H, Z) > 0.05$

- J/ψ-veto: No lepton pairings can have a mass in the range, $m_{\ell_i\ell_j} \in [3, 4]\,\text{GeV}$.

The predicted and observed event yields after applying this selection are presented in Table 6.5.

Table 6.5: Observed and MC predicted event yields in the Z-boson control region (CR), sideband (SB), and signal window (SW) for events passing VH4L selection. The Z boson CR is defined by $m_{\mu\mu} \in [76, 106]\,\text{GeV}$. The SB is defined by $m_{\mu\mu} \in [110, 160]\,\text{GeV}$ with all processes blinded: $m_{\mu\mu} \notin [120, 130]\,\text{GeV}$. The SW is defined by $m_{\mu\mu} \in [120, 130]$.

$H \to \mu\mu$ selection	VH4L Selection		
$H \to \mu\mu$ dataset	2022-2024		
Region	CR	SB	SW
Data	563	96.0	15.0
$Z \to \mu\mu$	1.58	0.04	0.00
EW $Z \to \mu\mu$	0.00	0.000	0.000
$t\bar{t}$	0.29	0.09	0.000
Single Top	0.000	0.02	0.01
$t\bar{t}V$	1.92	0.25	0.07
VV	598	94.0	24.7
EW VV	12.0	0.97	0.27
ggF: $H \to \mu\mu$	0.000	0.000	0.000
VBF: $H \to \mu\mu$	0.000	0.000	0.000
ttH: $H \to \mu\mu$	0.00	0.00	0.00
VH: $H \to \mu\mu$	0.04	0.09	0.25
bbH: $H \to \mu\mu$	0.000	0.000	0.000
tH: $H \to \mu\mu$	0.00	0.00	0.00
MC bkg total	614	95.3	25.0
Data/MC bkg	0.917	1.007	0.600
MC signal total	-	-	0.25
Sig : bkg	-	-	1 : 99.9

For events with four muons, there is an ambiguity for the muon pair assigned to the Higgs boson. All possible muon pairings are computed and the muon pair with the

Table 6.6: The 14 variables used in training the NN for the VH4L category.

Variable	Description
$p_{\mathrm{T}}^{\mu_1}$	Transverse momentum of the leading muon
$p_{\mathrm{T}}^{\mu_2}$	Transverse momentum of the subleading muon
η_{μ_1}	Pseudorapidity of the leading muon
η_{μ_2}	Pseudorapidity of the subleading muon
ϕ_{μ_1}	Azimuthal angle of the leading muon
ϕ_{μ_2}	Azimuthal angle of the subleading muon
N_{muons}	The number of muons
$N_{\mathrm{electrons}}$	The number of electrons
$p_{\mathrm{T}}^{\ell_3}$	Transverse momentum of the third lepton
$p_{\mathrm{T}}^{\ell_4}$	Transverse momentum of the fourth lepton
$\Delta\phi(\ell_1^Z, \ell_2^Z)$	Azimuthal separation between the Z boson leptons
$\Delta\zeta$	Projection of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ on the bisector of the Higgs candidate dimuon system in the transverse plane [119]
$\Delta R(\mu_1^H, \mu_2^H)$	Angular separation of the Higgs candidate muons
$\Delta R(\ell_1, \ell_2)$	Angular separation of the two leading leptons

mass closest to the Z boson is labelled as the Z candidate pair. The two remaining muons are assigned to the Higgs boson. This results in the correct muon assignment for 99% of events where $m_{\mu\mu} \in [120, 130]\,\mathrm{GeV}$.

After VH4L selection and muon assignment, a NN is trained to discriminate between signal and background events. The NNs are trained on ZH signal events against $VV \to \ell\ell\ell\ell$ and $gg \to \ell\ell\ell\ell$ backgrounds. The variables used for training can be seen in Table 6.6. The MC signal and background distributions for the training variables are shown in Appendix A.1. A DisCo hyperparameter of $\lambda = 1.0$ is used during training. The distribution of scores after training, $\hat{y}^{\mathrm{VH4L}}$, can be seen in Figure 6.13.

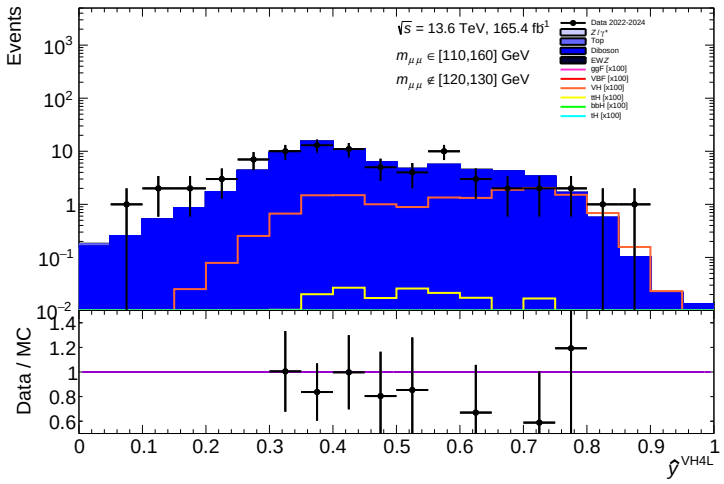


Figure 6.13: Distribution of the VH4L NN predictions, $\hat{y}^{\mathrm{VH4L}}$, for signal and background MC compared with data in the sidebands for events passing VH4L selection. The NN is used as a classifier by dividing events into categories based on these predictions.

6.4.4 VH3L Categories

The VH3L category targets WH Higgs-strahlung production where the W boson decays leptonically. Events considered for this category must have an one additional charged lepton. A Z-veto is applied to reject events that contain a muon pair with dimuon mass in the window $m_{\mu\mu} \in [80, 105]$ GeV. A b-jet veto is also applied for VH3L events. The predicted and observed event yields after applying this selection are presented in Table 6.7.

For events with three muons, there is an ambiguity for the muon pair assigned to the Higgs boson. To decide the muon assignment, a χ^2 minimization test is performed to determine the muon compatibility with the H decay and W decay. All possible muon

Table 6.7: Observed and MC predicted event yields in the Z-boson control region (CR), sideband (SB), and signal window (SW) for events passing VH3L selection. The Z boson CR is defined by $m_{\mu\mu} \in [76, 106]$ GeV. The SB is defined by $m_{\mu\mu} \in [110, 160]$ GeV with all processes blinded: $m_{\mu\mu} \notin [120, 130]$ GeV. The SW is defined by $m_{\mu\mu} \in [120, 130]$.

$H \to \mu\mu$ selection	VH3L: No b-jets, 3 leptons, no Z candidate		
$H \to \mu\mu$ dataset	2022-2024		
Region	CR	SB	SW
Data	190.4k	102.8k	100.1k
$Z \to \mu\mu$	129.6k	66.5k	65.8k
EW $Z \to \mu\mu$	91.3	67.6	66.5
$t\bar{t}$	4.88k	4.75k	4.55k
Single Top	460	438	406
$t\bar{t}V$	568	534	528
VV	25.7k	15.8k	14.6k
EW VV	445	282	261
ggF: $H \to \mu\mu$	1.15	1.21	1.58
VBF: $H \to \mu\mu$	0.14	0.15	0.18
ttH: $H \to \mu\mu$	1.91	1.94	2.03
VH: $H \to \mu\mu$	3.09	3.52	4.85
bbH: $H \to \mu\mu$	0.03	0.04	0.07
tH: $H \to \mu\mu$	0.24	0.25	0.28
MC bkg total	161.7k	88.4k	86.2k
Data/MC bkg	1.178	1.163	1.161
MC signal total	-	-	8.72
Sig : bkg	-	-	1 : 9.89k

pairings are created and the following is computed,

$$\chi^2 = \left(\frac{m_{\mu_1\mu_2} - m_H}{\sigma_H}\right)^2 + \left(\frac{m_{\mathrm{T}}^{\mu_3} - m_{\mathrm{T}}^W}{\sigma_W}\right), \tag{6.12}$$

where,

$$m_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{\mu} E_{\mathrm{T}}^{\mathrm{miss}} \left[1 - \cos\left(\Delta\phi\left(\vec{p}_{\mathrm{T}}^{\mu}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}}\right)\right)\right]}, \tag{6.13}$$

and with $m_H = 125.09\,\mathrm{GeV}$, $\sigma_H = 3\,\mathrm{GeV}$, $m_{\mathrm{T}}^W = 70\,\mathrm{GeV}$ and $\sigma_W = 20\,\mathrm{GeV}$. The pairing minimizing Equation 6.12 is used. This is shown to properly assign the muons in 95.95% of events.

After VH3L selection and muon assignment, a NN is trained to discriminate between

Table 6.8: The 20 variables used in training the NN for the VH3L category.

Variable	Description
$p_{\mathrm{T}}^{\mu_1}$	Transverse momentum of the leading muon
$p_{\mathrm{T}}^{\mu_2}$	Transverse momentum of the subleading muon
η_{μ_1}	Pseudorapidity of the leading muon
η_{μ_2}	Pseudorapidity of the subleading muon
ϕ_{μ_1}	Azimuthal angle of the leading muon
ϕ_{μ_2}	Azimuthal angle of the subleading muon
N_{muons}	The number of muons
$N_{\mathrm{electrons}}$	The number of electrons
$\cos\theta^*$	The lepton decay angle $\cos\theta^*$ in the Collins–Soper frame [118]
$p_{\mathrm{T}}^{\ell_3}$	Transverse momentum of the third lepton
ℓ_3^{iso}	Tight isolation WP flag for the W boson lepton
m_{lep}^{W}	Transverse mass of the W boson
$\Delta\eta(H,\ell_3)$	Pseudorapidity separation of the Higgs boson and W lepton
$\Delta\phi(H,\ell_3)$	Azimuthal separation of the Higgs boson and W lepton
$\Delta R(\mu_1^H,\mu_2^H)$	Angular separation of the Higgs candidate muons
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum
$\Delta\phi(H,\vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal separation of the Higgs boson and missing transverse momentum
$\Delta\zeta$	Projection of the $\vec{p}_T^{\mathrm{miss}}$ on the bisector of the dimuon system in the transverse plane [119]
$p_{\mathrm{T}}^{j_1}$	Transverse momentum of the leading jet
N_{jets}	The number of jets

signal and background events. The NNs are trained on WH signal events against $Z \to \mu\mu$, $t\bar{t}$, $VV \to \ell\ell\ell\ell$ and $VV \to \ell\ell\ell\nu$ backgrounds. The variables used for training can be seen in Table 6.8. The MC signal and background distributions for the training variables are included in Appendix A.2. A DisCo hyperparameter of $\lambda = 25.0$ is used during training. The distribution of scores after training, $\hat{y}^{\mathrm{VH3L}}$, can be seen in Figure 6.14.

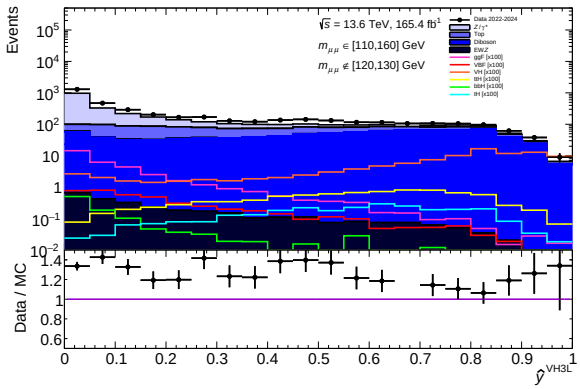


Figure 6.14: Distribution of the VH3L NN predictions, $\hat{y}^{\mathrm{VH3L}}$, for signal and background MC compared with data in the sidebands for events passing VH3L selection. The NN is used as a classifier by dividing events into categories based on these predictions.

6.4.5 VH2L Categories

The VH2L categories target ZH Higgs-strahlung production where the Z boson decays to two neutrinos. Events considered for this category must have exactly two muons and no additional charged leptons. There must be a missing transverse momentum of at least 120 GeV, motivated by the presence of final state neutrinos. Events with a b-jet are not considered for the VH2L category. The predicted and observed event yields after applying this selection are presented in Table 6.9.

After VH2L selection, a NN is trained to discriminate between signal and background events. The NNs are trained on ZH signal events against $Z \to \mu\mu$, $t\bar{t}$ and diboson backgrounds. The variables used for training can be seen in Table 6.10. The MC signal and background distributions for the training variables are shown in Appendix A.3. A DisCo hyperparameter of $\lambda = 50.0$ is used during training. The distribution of scores after training, $\hat{y}^{\mathrm{VH2L}}$, can be seen in Figure 6.15.

Table 6.9: Observed and MC predicted event yields in the Z-boson control region (CR), sideband (SB), and signal window (SW) for events passing VH2L selection. The Z boson CR is defined by $m_{\mu\mu} \in [76, 106]\,\text{GeV}$. The SB is defined by $m_{\mu\mu} \in [110, 160]\,\text{GeV}$ with all processes blinded: $m_{\mu\mu} \notin [120, 130]\,\text{GeV}$. The SW is defined by $m_{\mu\mu} \in [120, 130]$.

$H \to \mu\mu$ selection	VH2L: No b-jets, 2 muons, $E_{\text{T}}^{\text{miss}} > 120\,\text{GeV}$		
$H \to \mu\mu$ dataset	2022-2024		
Region	CR	SB	SW
Data	54.2k	5.08k	1.35k
$Z \to \mu\mu$	96.0k	1.79k	522
EW $Z \to \mu\mu$	103	9.11	2.53
$t\bar{t}$	3.10k	2.69k	774
Single Top	413	414	113
$t\bar{t}V$	24.4	6.37	1.90
VV	4.59k	887	256
EW VV	163	41.1	11.9
ggF: $H \to \mu\mu$	0.04	0.25	0.99
VBF: $H \to \mu\mu$	0.01	0.05	0.10
ttH: $H \to \mu\mu$	0.00	0.02	0.07
VH: $H \to \mu\mu$	0.03	0.36	1.24
bbH: $H \to \mu\mu$	0.00	0.00	0.01
tH: $H \to \mu\mu$	0.00	0.00	0.01
MC bkg total	104.4k	5.83k	1.68k
Data/MC bkg	0.519	0.871	0.801
MC signal total	-	-	2.41
Sig : bkg	-	-	1 : 697

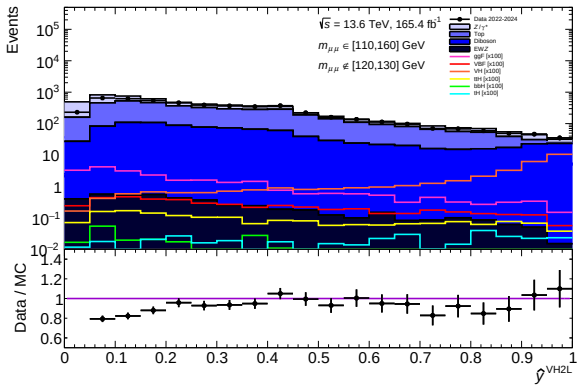


Figure 6.15: Distribution of the VH2L NN predictions, $\hat{y}^{\text{VH2L}}$, for signal and background MC compared with data in the sidebands for events passing VH2L selection. The NN is used as a classifier by dividing events into categories based on these predictions.

Table 6.10: The 15 variables used in training the NN for the VH2L category.

Variable	Description
$p_{\mathrm{T}}^{\mu_1}$	Transverse momentum of the leading muon
$p_{\mathrm{T}}^{\mu_2}$	Transverse momentum of the subleading muon
η_{μ_1}	Pseudorapidity of the leading muon
η_{μ_2}	Pseudorapidity of the subleading muon
ϕ_{μ_1}	Azimuthal angle of the leading muon
ϕ_{μ_2}	Azimuthal angle of the subleading muon
$\cos\theta^*$	The lepton decay angle $\cos\theta^*$ in the Collins–Soper frame [118]
$\Delta R(\mu_1^H, \mu_2^H)$	Angular separation of the Higgs candidate muons
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum
$\sigma(E_{\mathrm{T}}^{\mathrm{miss}})$	Significance of the missing transverse momentum given by the ratio of $E_{\mathrm{T}}^{\mathrm{miss}}$ to the scalar sum of p_{T} of muons and jets in the event
$\Delta\phi(H, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal separation of the Higgs boson and missing transverse momentum
$\Delta\zeta$	Projection of the $\vec{p}_T^{\mathrm{miss}}$ on the bisector of the dimuon system in the transverse plane [119]
$p_{\mathrm{T}}^{j_1}$	Transverse momentum of the leading jet
$p_{\mathrm{T}}^{\mathrm{balance}}$	Transverse momentum balance ratio between the dimuon and dijet system [8]
N_{jets}	The number of jets

6.4.6 VBF and Higgs 2-Jet Categories

These categories target VBF and ggF Higgs production events. Events considered for this category must have exactly two muons and no additional charged leptons. There must be a missing transverse momentum of less than 120 GeV to avoid overlap with the VH2L categories. There must be at least two jets in each event. Events with a b-jet are removed. The predicted and observed event yields after applying this selection are presented in Table 6.11.

After selection, two different NNs are trained targeting different Higgs production modes. The first NN is trained to only target VBF signal events. A second NN is trained targeting both VBF and ggF Higgs signal events. The scores from the first

Table 6.11: Observed and MC predicted event yields in the Z-boson control region (CR), sideband (SB), and signal window (SW) for events passing VBF selection. The Z boson CR is defined by $m_{\mu\mu} \in [76, 106]$ GeV. The SB is defined by $m_{\mu\mu} \in [110, 160]$ GeV with all processes blinded: $m_{\mu\mu} \notin [120, 130]$ GeV. The SW is defined by $m_{\mu\mu} \in [120, 130]$.

$H \to \mu\mu$ selection	No b-jets, $E_{\mathrm{T}}^{\mathrm{miss}} < 120$ GeV, 2 or more jets		
$H \to \mu\mu$ dataset	2022-2024		
Region	CR	SB	SW
Data	10.70M	197.8k	60.4k
$Z \to \mu\mu$	10.14M	173.0k	52.5k
EW $Z \to \mu\mu$	31.7k	1.00k	300
$t\bar{t}$	14.8k	11.6k	3.36k
Single Top	1.90k	1.55k	441
$t\bar{t}V$	170	18.5	5.45
VV	86.8k	3.58k	1.05k
EW VV	533	112	32.9
ggF: $H \to \mu\mu$	3.53	30.1	165
VBF: $H \to \mu\mu$	0.76	7.90	39.6
ttH: $H \to \mu\mu$	0.01	0.12	0.55
VH: $H \to \mu\mu$	0.33	3.40	16.1
bbH: $H \to \mu\mu$	0.02	0.11	0.73
tH: $H \to \mu\mu$	0.01	0.07	0.33
MC bkg total	10.28M	190.9k	57.7k
Data/MC bkg	1.040	1.036	1.048
MC signal total	-	-	222
Sig : bkg	-	-	1 : 260

NN trained on VBF events are then used to define four VBF targeting categories. Any events that are not included in these four categories are then considered for a second set of categories targeting the remaining VBF and ggF signal. All remaining events are divided into four extra categories called "Higgs 2-Jet" categories using this second NN.

The VBF NN is trained with only VBF signal events and the Higgs 2-Jet NN is trained with VBF and ggF signal events. Both NNs are trained against $Z \to \mu\mu$, $t\bar{t}$ and diboson backgrounds. The NNs are trained using the same set of variables, which are shown in Table 6.12. The MC signal and background distributions for the training variables are shown in Appendix A.4 and Appendix A.5. A DisCo hyperparameter

of $\lambda = 100.0$ is used during training for both NNs. The distribution of scores for the VBF NN, $\hat{y}^{\text{VBF}}$, can be seen in Figure 6.16. The scores for the Higgs 2-Jet NN, $\hat{y}^{\text{Higgs 2-Jet}}$, can be seen in Figure 6.17.

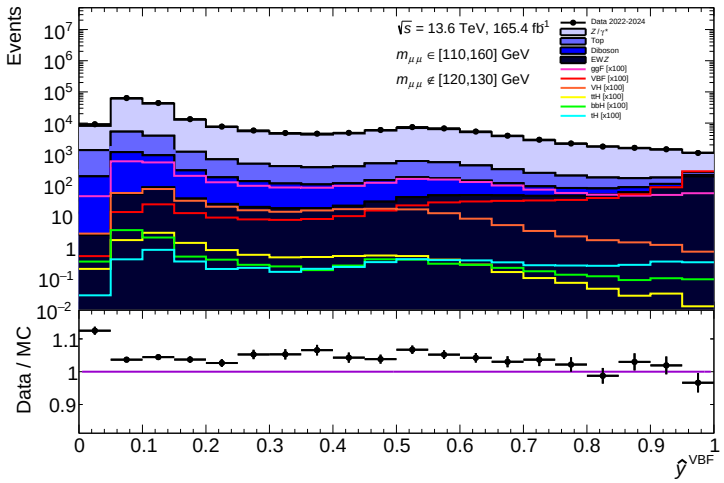


Figure 6.16: Distribution of the VBF NN predictions, $\hat{y}^{\text{VBF}}$, for signal and background MC compared with data in the sidebands for events passing VBF selection. The NN is used as a classifier by dividing events into categories based on these predictions.

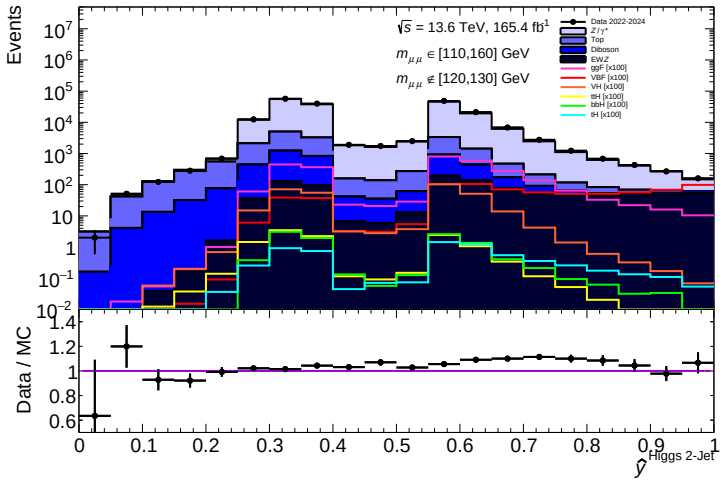


Figure 6.17: Distribution of the Higgs 2-Jet NN predictions, $\hat{y}^{\text{Higgs 2-Jet}}$, for signal and background MC compared with data in the sidebands for events passing Higgs 2-Jet selection. The NN is used as a classifier by dividing events into categories based on these predictions.

Table 6.12: The 23 variables used in training the NNs for the VBF and Higgs 2-Jet categories.

Variable	Description
$p_{\mathrm{T}}^{\mu_1}$	Transverse momentum of the leading muon
$p_{\mathrm{T}}^{\mu_2}$	Transverse momentum of the subleading muon
η_{μ_1}	Pseudorapidity of the leading muon
η_{μ_2}	Pseudorapidity of the subleading muon
ϕ_{μ_1}	Azimuthal angle of the leading muon
ϕ_{μ_2}	Azimuthal angle of the subleading muon
$\cos\theta^*$	The lepton decay angle $\cos\theta^*$ in the Collins–Soper frame [118]
$\Delta R(\ell_1^H, \ell_2^H)$	Angular separation of the Higgs candidate muons
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum
$\Delta\zeta$	Projection of the $\vec{p}_T^{\mathrm{miss}}$ on the bisector of the dimuon system in the transverse plane [119]
$\Delta\phi(H, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal separation of the Higgs boson and the missing transverse momentum
$p_{\mathrm{T}}^{j_1}$	Transverse momentum of the leading jet
$p_{\mathrm{T}}^{j_2}$	Transverse momentum of the subleading jet
η_{j_1}	Pseudorapidity of the leading jet
η_{j_2}	Pseudorapidity of the subleading jet
$N_{\mathrm{track}}^{j_1}$	Number of tracks in the leading jet with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$
$N_{\mathrm{track}}^{j_2}$	Number of tracks in the subleading jet with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$
$p_{\mathrm{T}}^{\mathrm{balance}}$	Transverse momentum balance ratio between the dimuon and dijet system [8]
N_{jets}	The number of jets
m_{jj}	Dijet invariant mass
$\Delta R(j_1, j_2)$	Angular separation of the two leading jets
$H_{\mathrm{T}}^{\mathrm{jets}}$	Scalar sum of the transverse momentum of jets
$\Delta\phi_{\mathrm{min}}(H, j)$	Minimum azimuthal separation between the Higgs candidate and leading or subleading jet

6.4.7 Higgs 1-Jet Categories

This category targets VBF and ggF Higgs production events with one final state jet. These are referred to as "Higgs 1-Jet" categories. Events considered for this category must have exactly two muons and no additional charged leptons. There must be a missing transverse momentum of less than 120 GeV to avoid overlap with the VH2L category. Events with a b-jet are also removed. The predicted and observed event yields after applying this selection are presented in Table 6.13.

Table 6.13: Observed and MC predicted event yields in the Z-boson control region (CR), sideband (SB), and signal window (SW) for events passing Higgs 1-Jet selection. The Z boson CR is defined by $m_{\mu\mu} \in [76, 106]$ GeV. The SB is defined by $m_{\mu\mu} \in [110, 160]$ GeV with all processes blinded: $m_{\mu\mu} \notin [120, 130]$ GeV. The SW is defined by $m_{\mu\mu} \in [120, 130]$.

$H \to \mu\mu$ selection	No b-jets, 2 muons, $E_{\mathrm{T}}^{\mathrm{miss}} < 120$ GeV, 1 jet		
$H \to \mu\mu$ dataset	2022-2024		
Region	CR	SB	SW
Data	22.98M	371.2k	113.2k
$Z \to \mu\mu$	22.45M	346.1k	105.2k
EW $Z \to \mu\mu$	23.7k	463	141
$t\bar{t}$	6.61k	5.44k	1.58k
Single Top	2.42k	1.84k	538
$t\bar{t}V$	3.04	1.63	0.52
VV	48.2k	4.63k	1.37k
EW VV	111	29.9	7.17
ggF: $H \to \mu\mu$	5.16	42.0	271
VBF: $H \to \mu\mu$	0.36	3.93	22.4
ttH: $H \to \mu\mu$	0.00	0.00	0.00
VH: $H \to \mu\mu$	0.15	1.31	7.24
bbH: $H \to \mu\mu$	0.03	0.23	1.56
tH: $H \to \mu\mu$	0.00	0.01	0.03
MC bkg total	22.53M	358.5k	108.9k
Data/MC bkg	1.020	1.035	1.040
MC signal total	-	-	302
Sig : bkg	-	-	1 : 361

After selection a NN is trained targeting both VBF and ggF Higgs signal events. The NN is trained against $Z \to \mu\mu$, $t\bar{t}$ and diboson backgrounds. The variables used for

Table 6.14: The 16 variables used in training the NN for the Higgs 1-Jet category.

Variable	Description
$p_{\mathrm{T}}^{\mu_1}$	Transverse momentum of the leading muon
$p_{\mathrm{T}}^{\mu_2}$	Transverse momentum of the subleading muon
η_{μ_1}	Pseudorapidity of the leading muon
η_{μ_2}	Pseudorapidity of the subleading muon
ϕ_{μ_1}	Azimuthal angle of the leading muon
ϕ_{μ_2}	Azimuthal angle of the subleading muon
$\cos\theta^*$	The lepton decay angle $\cos\theta^*$ in the Collins–Soper frame [118]
$\Delta R(\ell_1^H, \ell_2^H)$	Angular separation of the Higgs candidate muons
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum
$\Delta\zeta$	Projection of the $\vec{p}_T^{\mathrm{miss}}$ on the bisector of the dimuon system in the transverse plane [119]
$\Delta\phi(H, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal separation of the Higgs boson and missing transverse momentum
$p_{\mathrm{T}}^{j_1}$	Transverse momentum of the leading jet
η_{j_1}	Pseudorapidity of the leading jet
$N_{\mathrm{track}}^{j_1}$	Number of tracks in the leading jet with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$
$\Delta\phi(H, j_1)$	Azimuthal separation between the Higgs candidate and the leasing jet

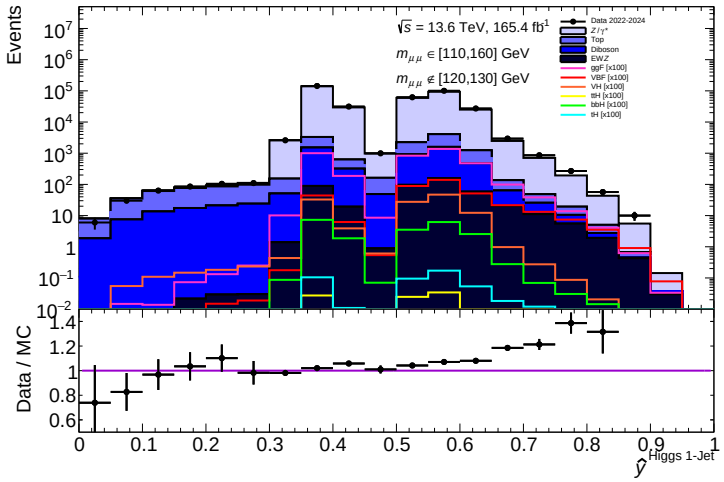


Figure 6.18: Distribution of the Higgs 1-Jet NN predictions, $\hat{y}^{\text{Higgs 1-Jet}}$, for signal and background MC compared with data in the sidebands for events passing Higgs 1-Jet selection. The NN is used as a classifier by dividing events into categories based on these predictions.

training the NN are shown in Table 6.14. The MC signal and background distributions for the training variables are shown in Appendix A.6. A DisCo hyperparameter of $\lambda = 100.0$ is used during training. The scores for the Higgs 1-Jet NN, $\hat{y}^{\text{Higgs 1-Jet}}$, can be seen in Figure 6.18.

6.4.8 Higgs 0-Jet Categories

This category targets VBF and ggF Higgs production events with no final state jets. These are referred to as "Higgs 0-Jet" categories. Events considered for this category must have exactly two muons, no additional charged leptons and no jets. There must be a missing transverse momentum of less than 120 GeV to avoid overlap with the VH2L category. The predicted and observed event yields after applying this selection are presented in Table 6.15.

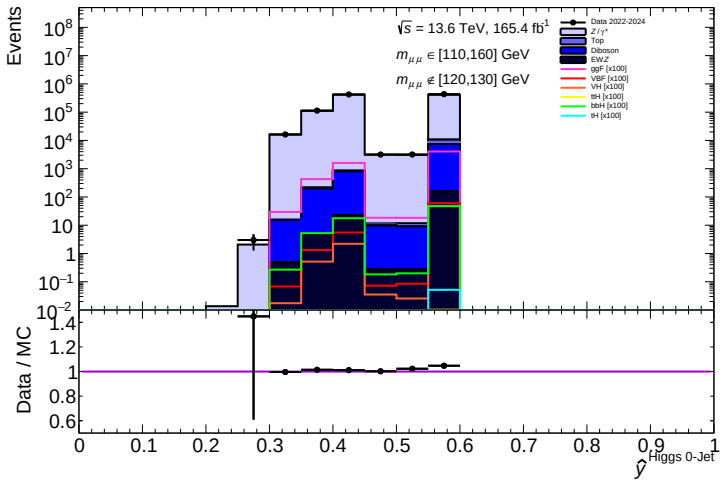


Figure 6.19: Distribution of the Higgs 0-Jet NN predictions, $\hat{y}^{\text{Higgs 0-Jet}}$, for signal and background MC compared with data in the sidebands for events passing Higgs 0-Jet selection. The NN is used as a classifier by dividing events into categories based on these predictions.

After selection, a NN is trained targeting both VBF and ggF Higgs signal events. The

Table 6.15: Observed and MC predicted event yields in the Z-boson control region (CR), sideband (SB), and signal window (SW) for events passing Higgs 0-Jet selection. The Z boson CR is defined by $m_{\mu\mu} \in [76, 106]\,\text{GeV}$. The SB is defined by $m_{\mu\mu} \in [110, 160]\,\text{GeV}$ with all processes blinded: $m_{\mu\mu} \notin [120, 130]\,\text{GeV}$. The SW is defined by $m_{\mu\mu} \in [120, 130]$.

$H \to \mu\mu$ selection	2 muons, $E_{\text{T}}^{\text{miss}} < 120\,\text{GeV}$, no jets		
$H \to \mu\mu$ dataset	2022-2024		
Region	CR	SB	SW
Data	71.65M	986.0k	301.2k
$Z \to \mu\mu$	73.27M	960.0k	292.1k
EW $Z \to \mu\mu$	1.20k	186	58.5
$t\bar{t}$	2.39k	2.15k	611
Single Top	1.42k	1.17k	330
$t\bar{t}V$	0.43	0.25	0.09
VV	34.1k	8.28k	2.47k
EW VV	18.4	4.30	1.52
ggF: $H \to \mu\mu$	7.31	63.7	454
VBF: $H \to \mu\mu$	0.08	0.69	4.70
ttH: $H \to \mu\mu$	0.00	0.00	0.00
VH: $H \to \mu\mu$	0.11	0.52	2.92
bbH: $H \to \mu\mu$	0.08	0.72	5.04
tH: $H \to \mu\mu$	0.00	0.00	0.00
MC bkg total	73.31M	971.7k	295.5k
Data/MC bkg	0.977	1.015	1.019
MC signal total	-	-	467
Sig : bkg	-	-	1 : 633

NN is trained against $Z \to \mu\mu$, $t\bar{t}$ and diboson backgrounds. The variables used for training the NN are shown in Table 6.16. The MC signal and background distributions for the training variables are shown in Appendix A.7. A DisCo hyperparameter of $\lambda = 100.0$ is used during training for both NNs. The scores for the Higgs 0-Jet NN, $\hat{y}^{\text{Higgs 0-Jet}}$, can be seen in Figure 6.19.

Table 6.16: The 11 variables used in training the NN for the Higgs 0-Jet category.

Variable	Description
$p_{\mathrm{T}}^{\mu_1}$	Transverse momentum of the leading muon
$p_{\mathrm{T}}^{\mu_2}$	Transverse momentum of the subleading muon
η_{μ_1}	Pseudorapidity of the leading muon
η_{μ_2}	Pseudorapidity of the subleading muon
ϕ_{μ_1}	Azimuthal angle of the leading muon
ϕ_{μ_2}	Azimuthal angle of the subleading muon
$\cos\theta^*$	The lepton decay angle $\cos\theta^*$ in the Collins–Soper frame [118]
$\Delta R(\ell_1^H, \ell_2^H)$	Angular separation of the Higgs candidate muons
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum
$\Delta\zeta$	Projection of the $\vec{p}_T^{\mathrm{miss}}$ on the bisector of the dimuon system in the transverse plane [119]
$\Delta\phi(H, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal separation of the Higgs boson and missing transverse momentum

6.4.9 Boundary Optimization

The NN scores assigned to each event for each Higgs production mode are used to further divide the events into categories to maximize the total significance. The total number of categories, N, varies for each production mode. The Asimov significance in each category is given by [107]

$$Z_i = \sqrt{2\left((s_i + b_i)\log\left(\frac{s_i + b_i}{b_i}\right) - s_i\right)}, \tag{6.14}$$

where s_i is the signal yield from MC simulation in the SW for the i^{th} category. The background yield, b_i, is the yield from MC simulation in the SW that has been scaled by a factor so that the number of events in the background sidebands match the total number of events in the data sidebands. Equation 6.14 can be approximated by $Z \approx s/\sqrt{b}$, when $s \ll b$, which is the case for the $H \to \mu\mu$ analysis. The total significance across all categories is then computed using Equation 6.1. To ensure

Table 6.17: The expected yields and counting significances based on MC yields in the $m_{\mu\mu} \in [120, 130]$ GeV region. To obtain a more realistic background estimate, the yields are from MC simulation have been scaled by a factor so the number of events in the background sidebands matches the number of events in the data sidebands. The signal purity is based on the yield of the targeted signal over the total signal yield in each category.

Category	Scores	Signal Yield	Background Yield	Signal Purity [%]	s/b [%]	Significance
VBF-High	0.9963-1.0000	3.0	16.6	94.5	18.0	0.715
VBF-Medium	0.9899-0.9963	3.5	32.3	88.9	11.0	0.613
VBF-Low	0.9772-0.9899	4.8	92.6	82.4	5.2	0.493
VBF-VeryLow	0.9493-0.9772	6.7	228	73.3	2.9	0.440
Higgs 2-Jet-High	0.7330-1.0000	11	743	98.4	1.5	0.397
Higgs 2-Jet-Medium	0.6710-0.7330	17	1600	97.2	1.0	0.415
Higgs 2-Jet-Low	0.5720-0.6710	88	16600	92.9	0.53	0.683
Higgs 2-Jet-VeryLow	0.0000-0.5720	92	41600	88.0	0.22	0.451
Higgs 1-Jet-High	0.6880-1.0000	6.5	490	99.4	1.3	0.292
Higgs 1-Jet-Medium	0.6170-0.6880	20	2840	98.6	0.70	0.370
Higgs 1-Jet-Low	0.5340-0.6170	190	55000	97.0	0.34	0.799
Higgs 1-Jet-VeryLow	0.0000-0.5340	90	54300	96.6	0.17	0.386
Higgs 0-Jet-High	0.5530-1.0000	310	131000	98.0	0.23	0.844
Higgs 0-Jet-Medium	0.4160-0.5530	130	129000	98.8	0.10	0.356
Higgs 0-Jet-Low	0.3760-0.4160	20	17200	98.7	0.12	0.153
Higgs 0-Jet-VeryLow	0.0000-0.3760	16	22700	98.9	0.07	0.104
VH4L	0.4340-1.0000	0.44	9.68	98.8	4.6	0.143
VH3L-High	0.9340-1.0000	0.45	5.63	96.5	8.1	0.189
VH3L-Medium	0.7080-1.0000	2.6	121	87.6	2.2	0.236
VH2L-High	0.8830-1.0000	0.65	28.7	69.9	2.3	0.120
VH2L-Medium	0.0130-0.8830	1.8	1420	24.1	0.13	0.047
ttH-High	0.9984-1.0000	1.2	19.0	97.9	6.5	0.280
ttH-Medium	0.9798-1.0000	2.4	119	96.3	2.1	0.223
Total	-	1010	475000	-	0.2	2.113

sufficient statistics for signal extraction when fitting, a minimum of 5 background MC events (after scaling to data sidebands) are required in the SW in each category.

The estimated yields, signal purity, and significance in each category after optimization are shown in Table 6.17. The signal purity is defined as the ratio of the signal processes targeted by the DisCo NNs over the total signal yield in each category. The signal composition in each category is shown in Table 6.18. Although most categories have a high signal purity, this analysis targets all Higgs production modes and it is more important to include as much signal as possible and maximize the significance rather than having categories with a high signal purity.

Table 6.18: The signal composition for each category based on the MC yields in the $m_{\mu\mu} \in [120, 130]\,\text{GeV}$ region. Most categories are dominated by the signal production modes targeted by the DisCo NNs trained for each category. The VH2L categories target ZH signal events but have significant contributions from WH and ggF production modes.

Category	Signal Composition [%]				
	ggF	VBF	VH	$t\bar{t}H$+tH	$b\bar{b}H$
VBF-High	5.5	94.5	0.0	0.0	0.0
VBF-Medium	10.9	88.9	0.1	0.0	0.0
VBF-Low	17.3	82.4	0.1	0.1	0.0
VBF-VeryLow	26.1	73.3	0.3	0.2	0.1
Higgs 2-Jet-High	62.6	35.8	1.2	0.3	0.1
Higgs 2-Jet-Medium	76.2	21.0	2.5	0.2	0.2
Higgs 2-Jet-Low	80.6	12.3	6.5	0.3	0.3
Higgs 2-Jet-VeryLow	80.3	7.7	10.9	0.6	0.5
Higgs 1-Jet-High	73.4	26.0	0.4	0.0	0.2
Higgs 1-Jet-Medium	85.7	12.9	1.1	0.0	0.3
Higgs 1-Jet-Low	89.0	8.0	2.5	0.0	0.4
Higgs 1-Jet-VeryLow	93.0	3.6	2.6	0.0	0.7
Higgs 0-Jet-High	96.7	1.4	0.9	0.0	1.1
Higgs 0-Jet-Medium	98.5	0.3	0.1	0.0	1.1
Higgs 0-Jet-Low	98.4	0.3	0.1	0.0	1.2
Higgs 0-Jet-VeryLow	98.7	0.2	0.1	0.0	1.0
VH3L-High	0.1	0.0	99.1	0.8	0.0
VH3L-Medium	0.6	0.2	94.5	4.7	0.0
VH4L	0.0	0.0	98.8	1.2	0.0
ttH-High	0.5	0.1	1.5	97.9	0.0
ttH-Medium	1.2	0.1	2.4	96.3	0.1
VH2L-High	0.9	0.5	97.7	0.8	0.0
VH2L-Medium	54.5	5.7	35.2	4.2	0.5

6.5 Modelling of the Dimuon Invariant Mass Distributions

After the events have been divided into categories, the $m_{\mu\mu}$ distributions need to be accurately described by a smooth function used in the statistical fitting procedure. The functional form needs to be determined and tested separately in each category to ensure that their behaviour can be appropriately modelled for signal extraction. The background modelling also tests that there is no shaping in the dimuon invariant mass distribution.

6.5.1 Modelling the $m_{\mu\mu}$ Signal Distributions

The $m_{\mu\mu}$ signal distributions are modelled using a double-sided crystal ball (DSCB) function. The DSCB is a probability density function (PDF) consisting of a Gaussian core with power law lower and higher tails. The DSCB is given by

$$f(x;\mu,\sigma,\alpha_L,n_L,\alpha_H,n_H) = N \cdot \begin{cases} A_L \cdot (B_L - \frac{x-\mu}{\sigma})^{-n_L} & \text{for } \frac{x-\mu}{\sigma} < -\alpha_L \\ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & \text{for } -\alpha_L \leq \frac{x-\mu}{\sigma} \leq \alpha_H \\ A_H \cdot (B_H + \frac{x-\mu}{\sigma})^{-n_H} & \text{for } \frac{x-\mu}{\sigma} > \alpha_H \end{cases}, \tag{6.15}$$

where

$$A_L = \left(\frac{n_L}{|\alpha_L|}\right)^{n_L} \cdot \exp\left(-\frac{\alpha_L^2}{2}\right), \tag{6.16}$$

$$B_L = \frac{n_L}{|\alpha_L|} - |\alpha_L|, \tag{6.17}$$

$$A_H = \left(\frac{n_H}{|\alpha_H|}\right)^{n_H} \cdot \exp\left(-\frac{\alpha_R^2}{2}\right), \tag{6.18}$$

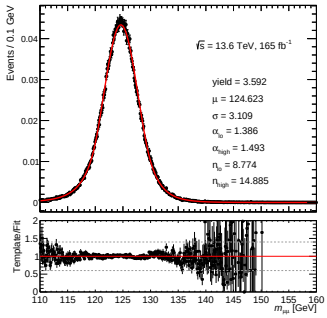
$$B_H = \frac{n_H}{|\alpha_H|} - |\alpha_H|. \tag{6.19}$$

The parameters represent:

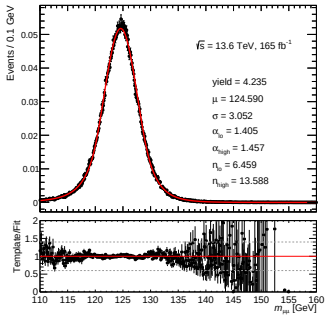
- μ: the mean value of the crystal ball function
- σ: the width of the crystal ball function
- α_L: the threshold for the left, low-end tail power-law
- α_H: the threshold for the right, high-end tail power-law
- n_L: the power in the left, low-end tail power-law
- n_H: the power in the left, high-end tail power-law
- N: the normalization factor

The values for the parameters in the DSCB function are determined by fitting the signal MC distributions in each category. The fit is performed to the total signal distribution including all Higgs production modes contributing to the signal in each category.

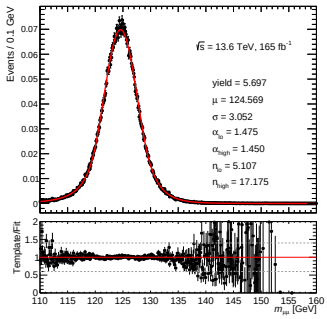
The signal $m_{\mu\mu}$ fits for each category are shown in Figure 6.20, 6.21 and 6.22.



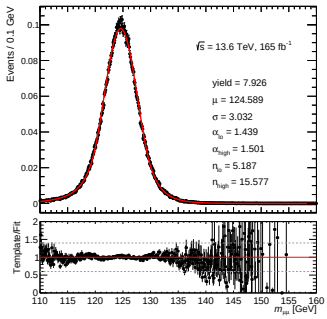
(a) VBF-High



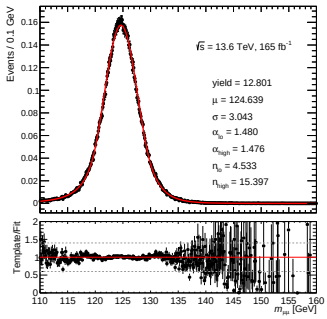
(b) VBF-Medium



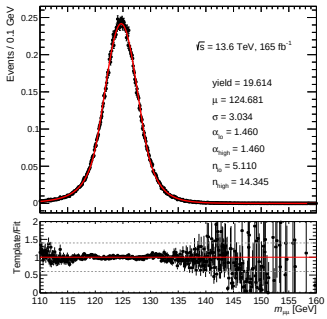
(c) VBF-Low



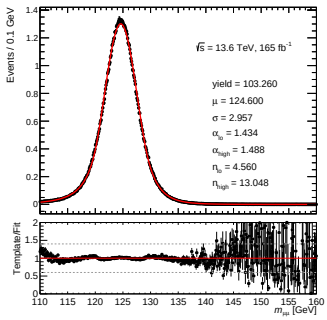
(d) VBF-Very Low



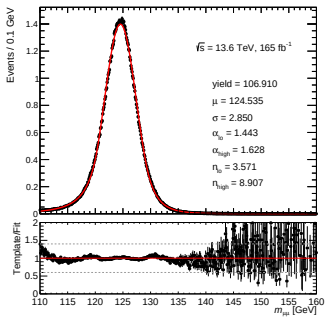
(e) Higgs 2-Jet-High



(f) Higgs 2-Jet-Medium

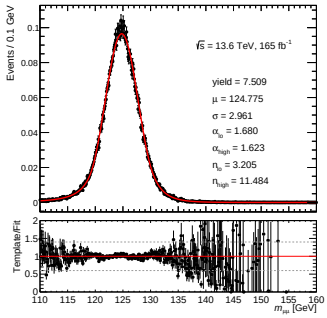


(g) Higgs 2-Jet-Low

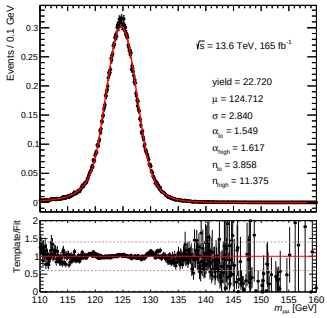


(h) Higgs 2-Jet-Very Low

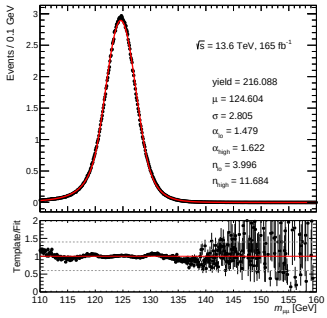
Figure 6.20: Signal $m_{\mu\mu}$ DSCB fits for the VBF and Higgs 2-Jet categories.



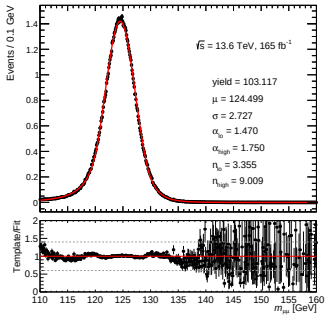
(a) Higgs 1-Jet-High



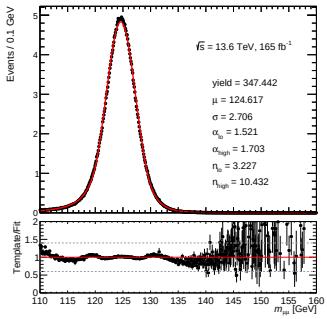
(b) Higgs 1-Jet-Medium



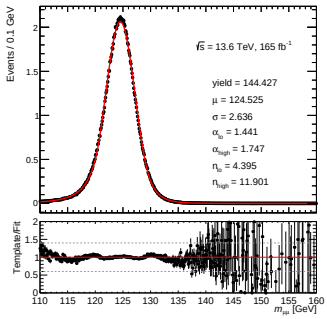
(c) Higgs 1-Jet-Low



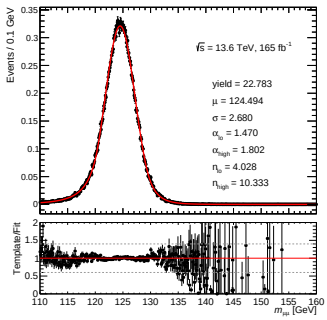
(d) Higgs 1-Jet-Very Low



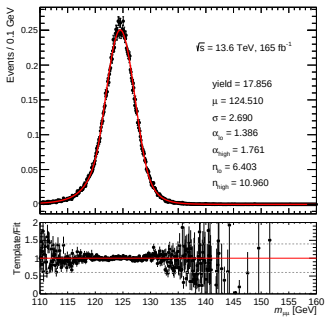
(e) Higgs 0-Jet-High



(f) Higgs 0-Jet-Medium

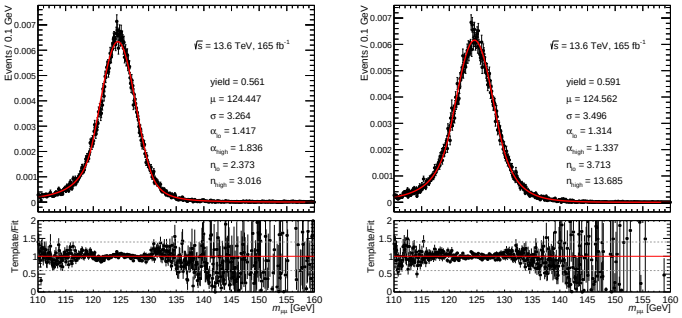


(g) Higgs 0-Jet-Low



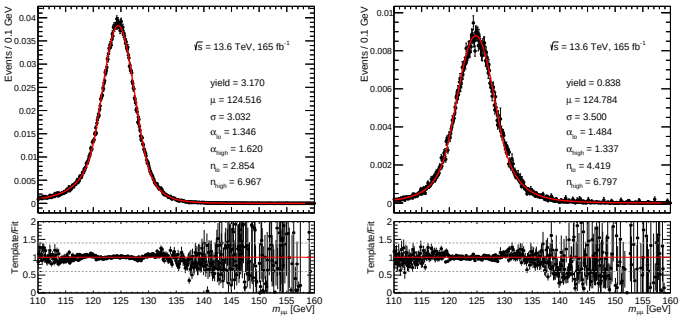
(h) Higgs 0-Jet-Very Low

Figure 6.21: Signal $m_{\mu\mu}$ DSCB fits for the Higgs 1-Jet and Higgs 0-Jet categories.



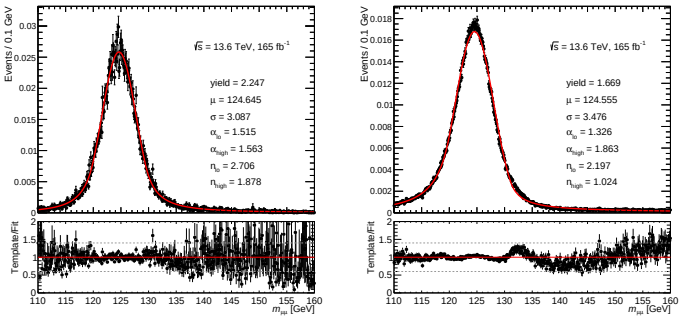
(a) VH4L

(b) VH3L-High



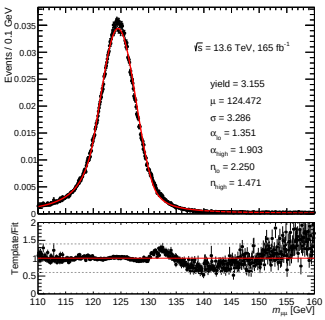
(c) VH3L-Medium

(d) VH2L-High



(e) VH2L-Low

(f) $t\bar{t}H$-High



(g) $t\bar{t}H$-Medium

Figure 6.22: Signal $m_{\mu\mu}$ DSCB fits for the VH and $t\bar{t}H$ categories.

6.5.2 Modelling the $m_{\mu\mu}$ Background Distributions

6.5.2.1 Background Models and Selection Procedure

The background for the $H \to \mu\mu$ process is dominated by DY, with significant contributions from $t\bar{t}$, diboson and other less common processes. The breakdown of the background varies significantly in each category defined by the Higgs production mode and NN scores, resulting in vastly different distributions that need to be modelled. To resolve this, a series of different background models are considered for each category, which consist of a core function multiplied by an empirical function.

The core function is meant to model the DY mass shape in each category. The core function is a rigid function with no free parameters and it is common to all categories. It is an analytical function based on the LO DY lineshape [120],

$$\mathrm{DY}(m_{\mu\mu}) = \sum_{q} \mathcal{L}_{q\bar{q}}(m_{\mu\mu}) \cdot \sigma_{q\bar{q}}(m_{\mu\mu}), \tag{6.20}$$

where $q = u, s, d, c, b$. The parton luminosity contribution, $\mathcal{L}_{q\bar{q}}$, is derived from the PDF4LHC21 PDF set as a function of $\hat{s} = m^2_{\mu\mu}$ using APFEL [121] with LHAPDF [122] and parametrized with a 6th order polynomial. The matrix element is given by,

$$\sigma_{q\bar{q}}(\hat{s}) = \frac{4\pi\alpha^2}{3\hat{s}N_c}\left[Q_q^2 - 2Q_q V_\ell V_q \chi_{Z\gamma}(\hat{s}) + (A_\ell^2 + V_\ell^2)(A_q^2 + V_q^2)\chi_Z(\hat{s})\right], \tag{6.21}$$

where

$$\chi_{Z\gamma}(\hat{s}) = \kappa \frac{\hat{s}(\hat{s} - m_Z^2)}{(\hat{s} - m_Z^2)^2 + \Gamma_Z^2 m_Z^2}, \tag{6.22}$$

$$\chi_{Z}(\hat{s}) = \kappa^2 \frac{\hat{s}^2}{(\hat{s} - m_Z^2)^2 + \Gamma_Z^2 m_Z^2}, \tag{6.23}$$

$$\kappa = \frac{\sqrt{2} G_F m_Z^2}{4\pi\alpha}. \tag{6.24}$$

Here, Q represents the electric charge, V represents the vector coupling and A represents the axial-vector coupling of the fermions, α and G_F represents the electroweak couplings, m_Z and Γ_Z represent the mass and width of the Z-boson and $N_c = 5$ is the number of QCD colour charges. This DY function is then convolved with a Gaussian function with a resolution that is dependent on the mass and derived from simulation.

The empirical functions are different functions that account for distortions in the core DY mass shape due to higher-order theory corrections and contributions from other background processes. The two groups of functions considered for the empirical functions are power-law functions ('Power') and exponentials of polynomials ('Epoly'), which are defined in Table 6.19.

Table 6.19: Forms of the empirical functions used for background modelling.

Function	Expression
PowerN	$m_{\mu\mu}^{(a_0 + a_1 m_{\mu\mu} + a_2 m_{\mu\mu}^2 + ... + a_N m_{\mu\mu}^N)}$
EpolyN	$e^{(a_1 m_{\mu\mu} + a_2 m_{\mu\mu}^2 + ... + a_N m_{\mu\mu}^N)}$

The combination of the core function and empirical functions results in multiple possible models for each category. The choice of background model is based on the following criteria:

- Goodness of fit: The function used for the background model must exhibit a χ^2 probability greater than 1%.

- Spurious signal (SS) test: the background model must pass the SS test (Section 6.5.2.2) with $\hat{s}_{SS}^{\max} < 0.2\,\sigma_{\hat{s}}$.

- If multiple models pass the above criteria, the model with the lowest number of free parameters is chosen. If multiple models with the same number of free parameters pass, the model with the lowest $\hat{s}_{SS}^{\max}$ is selected.

The χ^2 criteria ensures that the background model accurately describes the background shape. The SS test avoids using a background parametrization that has issues in the crucial signal region. Such a bias can be introduced into the backgrounds when defining categories based on machine learning classifiers. Selecting models that have the minimum number of free parameters and lowest $\hat{s}_{SS}^{\max}$ selects the function that will maximize the significance.

The background distributions used for selecting the models are obtained from the MC background samples for each category. This MC background distribution is then scaled by a factor so that the number of MC background events in the sidebands matches the number of data events in the sidebands. Fits are then performed to these scaled MC background distributions in each category to select the background models. It should be noted that the MC templates are only used to select the background models, which are flexible and contain free parameters that will be determined by a final fit to the data sidebands. This means that perfect agreement between the MC samples and the data sidebands is not essential and the background model selected for each category should be flexible enough to adapt to potential differences between simulation and data.

6.5.2.2 Spurious Signal Test

The spurious signal (SS) method is used to evaluate the systematic uncertainty of the background modelling in each category. For the SS test, a signal-free MC sample is prepared for each category composed of the different backgrounds. Each sample is then fit with a combined signal and background model. Since the samples are signal-free by construction, any signal events that appear in the fit are due to the inability of the background models to fully describe the background spectrum in that category. The measured number of signal events from the fit are referred to the *spurious signal* and are denoted $\hat{s}_{\text{SS}}$.

The background template for each category is fit with a combined signal and background model as a scan of m_H from 120–130 GeV in steps of 1 GeV. The fit is performed across the entire SR range of 110–160 GeV. The maximum number of signal events from the fits in the scan, $\hat{s}_{SS}^{\text{max}}$, is taken to be the total SS for that category. The $\hat{s}_{\text{SS}}^{\text{max}}$ for each category is taken to be the background modelling uncertainty.

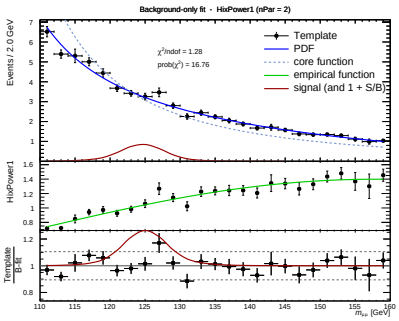
In order to pass the SS test for a given background model, it must satisfy $\hat{s}_{SS}^{\text{max}} < 0.2\,\sigma_{\hat{s}}$, where $\sigma_{\hat{s}}$ is the expected statistical error on the signal measurement in data in the range $m_{\mu\mu} \in [120, 130]$ GeV. When applying this requirement, the MC statistical error on the signal is subtracted from the value of $\hat{s}_{SS}^{\text{max}}$.

6.5.2.3 Background Modelling Results

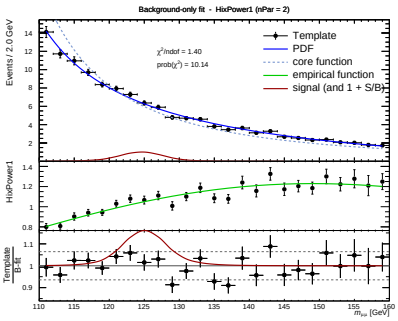
The background model selected for each category and the $\hat{s}_{SS}^{\text{max}}$ for that model can be seen in Table 6.20. The background $m_{\mu\mu}$ fits to MC in each category can be seen in Figure 6.23, 6.24 and 6.25.

Table 6.20: The selected $m_{\mu\mu}$ PDF for each category based on a fit to the MC background simulation and the number of parameters for that model. The $\hat{s}_{SS}^{\max}$ is the maximum spurious signal found in the scan range of 120–130 GeV. $\hat{s}_{SS}^{\max}/\sigma_{\hat{s}}$ is the ratio of the spurious signal to the expected statistical error on the signal measurement in data, after the MC statistical error is subtracted from the absolute value of the $\hat{s}_{SS}^{\max}$. This must be less than 20% to pass the SS test.

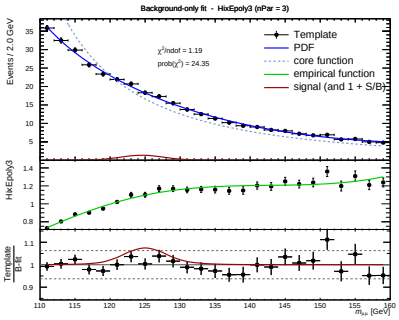
Category	Selected $m_{\mu\mu}$ PDF	$\hat{s}_{SS}^{\max}$	$\hat{s}_{SS}^{\max}/\sigma_{\hat{s}}$ [%]	Number of Parameters	P(χ^2) [%]
VBF-High	HixPower1	1.02	9.4	2	16.8
VBF-Medium	HixPower1	2.14	15	2	10.1
VBF-Low	HixEpoly3	4.66	18	3	24.4
VBF-Very Low	HixPower1	5.83	13	2	37.1
Higgs 2-Jet-High	HixEpoly1	-12.5	-15	1	5.76
Higgs 2-Jet-Medium	HixEpoly2	-20.6	-14	2	88.6
Higgs 2-Jet-Low	HixPower2	-40.8	-16	3	12.2
Higgs 2-Jet-Very Low	HixPower2	-198	-12	3	9.63
Higgs 1-Jet-High	HixPower1	8.83	-6.9	2	8.32
Higgs 1-Jet-Medium	HixEpoly4	-13.1	-5.0	4	1.62
Higgs 1-Jet-Low	HixEpoly2	-161	-17	2	5.84
Higgs 1-Jet-Very Low	HixEpoly2	62.6	0	2	48.8
Higgs 0-Jet-High	HixEpoly4	140	0	4	11.1
Higgs 0-Jet-Medium	HixEpoly3	-159	0	3	3.08
Higgs 0-Jet-Low	HixPower1	22.9	0	2	16.4
Higgs 0-Jet-Very Low	HixPower2	61.0	0	3	1.5
VH4L	HixEpoly2	0.496	4.2	2	63.8
VH3L-High	HixPower0	1.29	13	1	92.3
VH3L-Medium	HixEpoly3	5.82	14	3	24.0
VH2L-High	HixPower1	2.42	11	2	58.8
VH2L-Low	HixEpoly3	20.4	15	3	13.5
ttH-High	HixPower0	2.05	13	1	3.87
ttH-Medium	HixEpoly2	2.52	0	2	87.7



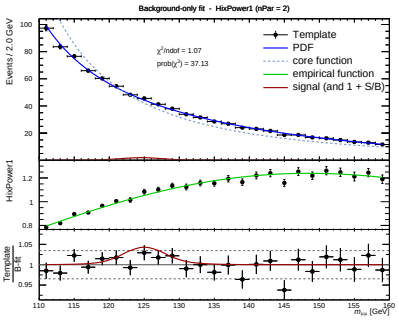
(a) VBF-High Category



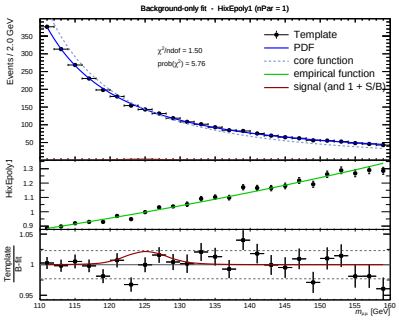
(b) VBF-Medium Category



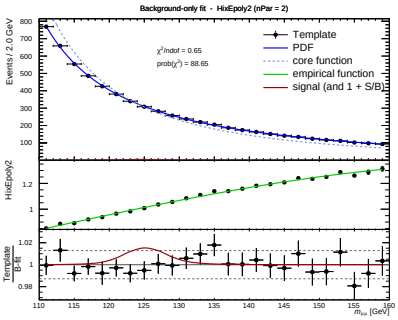
(c) VBF-Low Category



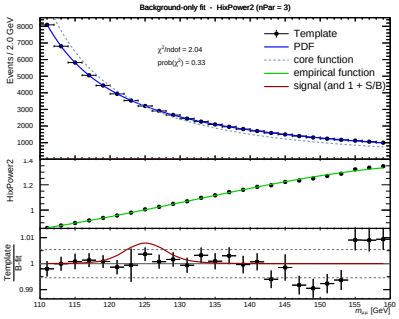
(d) VBF-Very Low Category



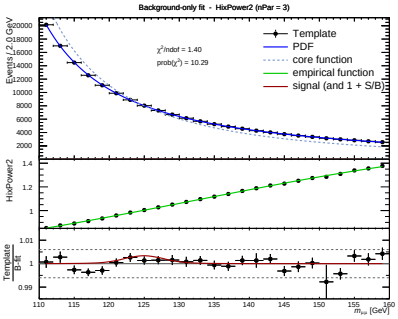
(e) Higgs 2-Jet-High Category



(f) Higgs 2-Jet-Medium Category

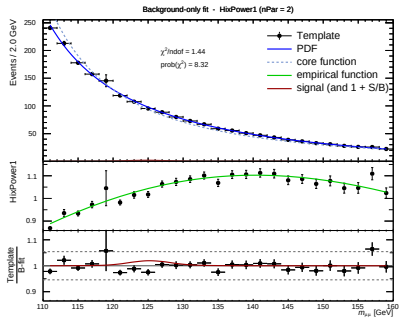


(g) Higgs 2-Jet-Low Category

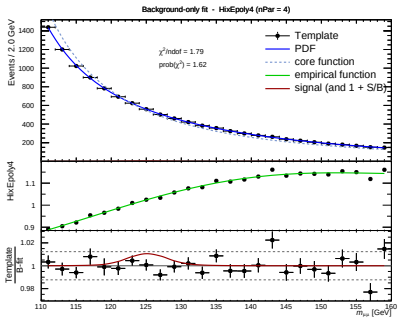


(h) Higgs 2-Jet-Very Low Category

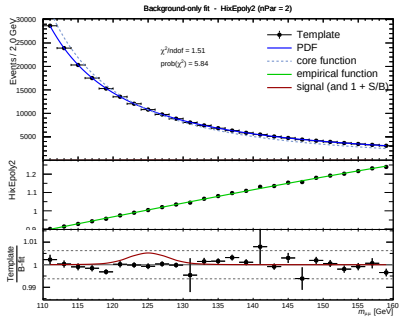
Figure 6.23: Selected background $m_{\mu\mu}$ models for the VBF and Higgs 2-Jet categories. The functions are derived as fits to background-only simulation (top panels), and the expected signal is shown as a red curve.



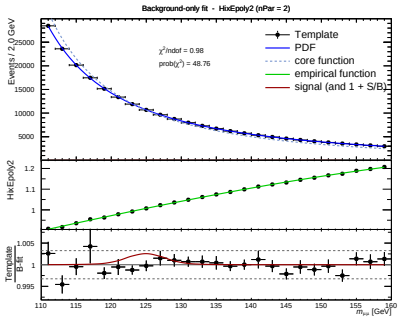
(a) Higgs 1-Jet-High Category



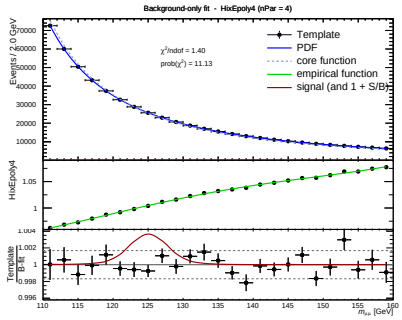
(b) Higgs 1-Jet-Medium Category



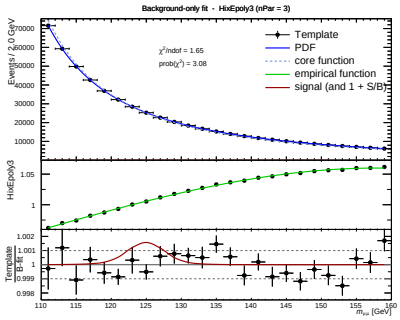
(c) Higgs 1-Jet-Low Category



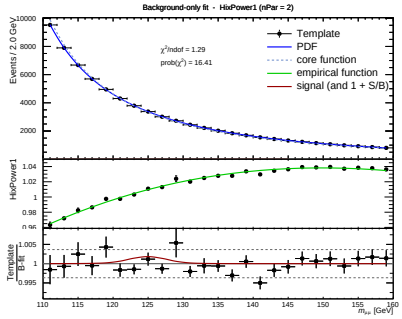
(d) Higgs 1-Jet-Very Low Category



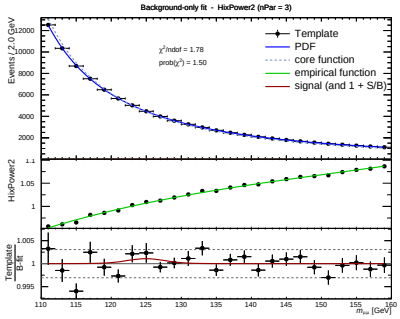
(e) Higgs 0-Jet-High Category



(f) Higgs 0-Jet-Medium Category

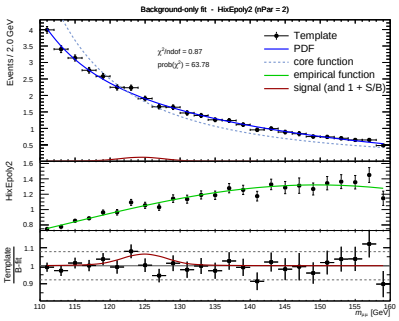


(g) Higgs 0-Jet-Low Category

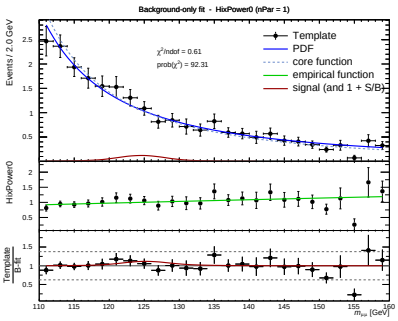


(h) Higgs 0-Jet-Very Low Category

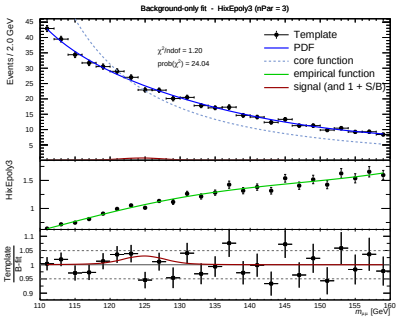
Figure 6.24: Selected background $m_{\mu\mu}$ models for the Higgs 1-Jet and Higgs 0-Jet categories. The functions are derived as fits to background-only simulation (top panels), and the expected signal is shown as a red curve.



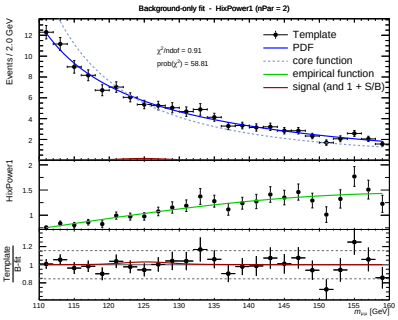
(a) VH4L Category



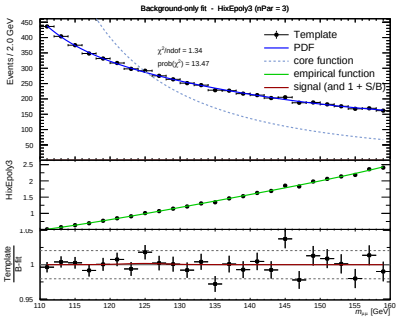
(b) VH3L-High Category



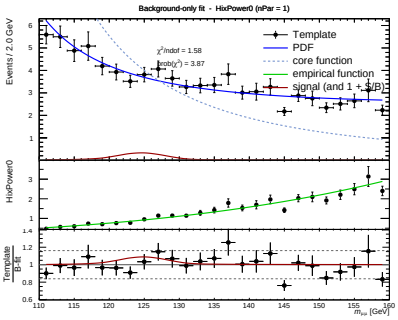
(c) VH3L-Medium Category



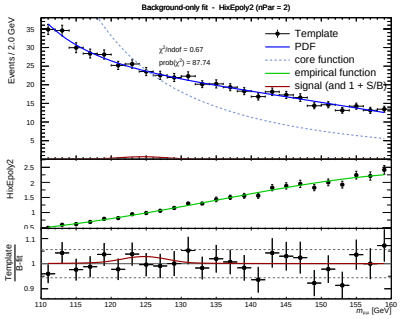
(d) VH2L-High Category



(e) VH2L-Low Category



(f) $t\bar{t}H$-High Category



(g) $t\bar{t}H$-Medium Category

Figure 6.25: Selected background $m_{\mu\mu}$ models for the VH and $t\bar{t}H$ categories. The functions are derived as fits to background-only simulation (top panels), and the expected signal is shown as a red curve.

Chapter 7

Results

This chapter reports the expected significance for the $H \to \mu\mu$ signal using the approach outlined in the previous chapter, and compares these results with categories defined using BDTs. A combination of DisCo NNs and BDTs were used for the most recent ATLAS $H \to \mu\mu$ search, and the unblinded results from this analysis are also presented.

The expected significance for the for the $H \to \mu\mu$ signal for 165 fb^{-1} of Run-3 data using categories defined with DisCo NNs is shown in Section 7.1. The expected significance using categories defined using BDTs from the most recent ATLAS $H \to \mu\mu$ search is then presented for comparison in Section 7.2. The unblinded results combining Run-2 and Run-3 ATLAS data using a combination of BDTs and DisCo NNs is shown in Section 7.3, which established ATLAS standalone evidence for the $H \to \mu\mu$ decay.

7.1 Expected Significance from Run-3 Data

To estimate the expected significance of the $H \to \mu\mu$ signal for $165\,\text{fb}^{-1}$ of Run-3 data collected at $\sqrt{s} = 13.6\,\text{TeV}$, a binned maximum likelihood fit to the Asimov $m_{\mu\mu}$ distributions in the range $m_{\mu\mu} \in [110, 160]\,\text{GeV}$ using the categories defined in Chapter 6 is performed. The Asimov datasets in each category are composed of data in the sidebands combined with signal and background MC simulation samples in the SW, using a bin size of 0.1 GeV. The spurious signal uncertainties listed in Section 6.5.2.3 are implemented in the fit as nuisance parameters that affect the signal yields and the Higgs boson mass is assumed to be $m_H = 125.09\,\text{GeV}$ [13].

A simultaneous fit to all 23 categories is performed with a floating signal strength, μ, for the signal-plus-background fit, and with $\mu = 0$ for the background-only hypothesis. The significance is then extracted from the difference in log-likelihood between the two fits. The signal strength is defined as the ratio of the observed signal yield to the one expected in the SM. For a fit to Asimov data consisting of MC samples based on SM expectations, a signal strength of $\mu = 1.0$ is obtained by construction. The results from the fits to Asimov data show an expected significance of 1.56σ for the $H \to \mu\mu$ signal with respect to the background-only hypothesis. This corresponds to a best-fit value of the signal strength parameter of $\mu = 1.0 \pm 0.6$. The breakdown of the significance for the groups of categories is shown in Table 7.1.

The dimuon invariant mass of the signal-plus-background fit to the combined categories is shown in Figure 7.1.

Table 7.1: The expected significance for each category and for the combined case using categories defined using DisCo NNs. The expected significances with statistical uncertainties only, and with statistical and spurious signal uncertainties are shown.

Category	Expected Significance [σ] (Stat. Only)	Expected Significance [σ] (Stat.+SS)
VBF	0.94	0.91
Higgs 2-Jet	0.74	0.70
Higgs 1-Jet	0.80	0.74
Higgs 0-Jet	0.67	0.65
VH4L	0.11	0.11
VH3L	0.22	0.21
VH2L	0.11	0.10
$t\bar{t}H$	0.31	0.29
Combined	1.63	1.56

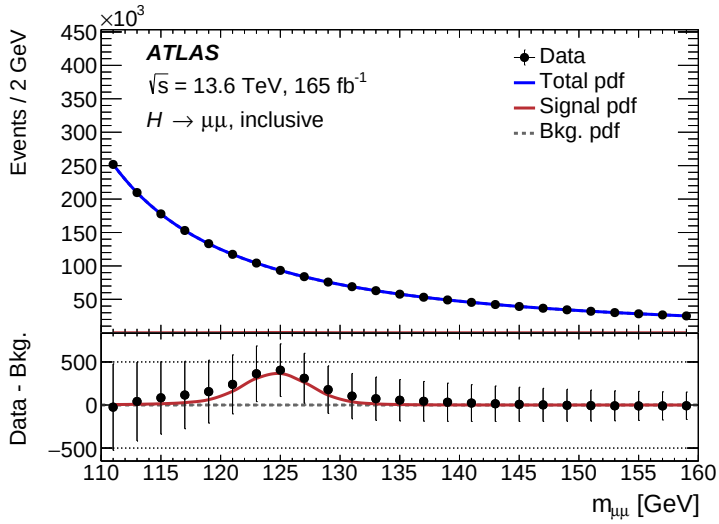


Figure 7.1: The Asimov dimuon invariant mass spectrum combining all 23 Run-3 categories that are defined using DisCo NNs. A fit is performed in the mass range $m_{\mu\mu} \in [110, 160]$ GeV, providing a signal strength of $\mu = 1.0 \pm 0.6$.

7.2 Comparison with BDTs

The most recent ATLAS search for the Higgs to dimuon decay [2] used 165 fb^{-1} of Run-3 data collected at $\sqrt{s} = 13.6\,\text{TeV}$ and combined this with the analysis of the 140 fb^{-1} of Run-2 data collected at $\sqrt{s} = 13\,\text{TeV}$ [3]. The analysis strategy for the Run-3 data follows the same approach presented in this thesis. The only difference is the choice of ML technique used to divide events into categories. With the exception of the $t\bar{t}H$ categories, boosted decision trees (BDTs) were used to categorize events instead of DisCo NNs. This allows for a direct comparison between the two approaches used to search for the $H \to \mu\mu$ signal.

The ATLAS Run-3 search used DisCo NNs as a classifier for the $t\bar{t}H$ categories following the approach outlined in this thesis. DisCo NNs were used because they were shown to increase the significance by 37% compared to BDTs based on MC simulation for 56 fb^{-1} of data (2022–2023).

BDTs are a ML algorithm that combines the outputs of many individual *decision trees* to create a powerful classifier. A decision tree makes predictions by recursively splitting data based on the values of input variables, which creates a series of decision boundaries that divide signal from background. Boosting improves the performance by training an ensemble of trees sequentially. Before building each new tree, events that were misclassified by the previous tree are given higher weights, which makes the algorithm focus more on these difficult events. This improves the overall performance of the BDT. The final classifier is typically a weighted sum of the predictions of all of the trees. BDTs have widely been used for classification problems in particle physics, including in the Run-2 and Run-3 ATLAS $H \to \mu\mu$ searches [2, 3] as they are robust

and easy to train. They can train well on small datasets and with continuous or discrete variables, and non-standardized data.

Although the training procedure is quite different for BDTs and NNs, the other aspects of the analysis were generally kept the same. The number of categories defined using BDTs and NNs is the same, although the training variables used and the boundaries defined based on the BDT and NN outputs does differ. BDTs can create shaping of the background distributions similar to that caused by a NN. In order to prevent shaping, training variables for the BDTs were carefully chosen to avoid any with correlations with $m_{\mu\mu}$. In most cases, the NNs trained in this thesis contain all of the BDT training variables used, with the addition of extra variables as the DisCo loss term would prevent correlations with $m_{\mu\mu}$.

Previous studies showed that DisCo NNs outperform BDTs for the $t\bar{t}H$ categories. A comparison of the estimated counting significance using NNs outlined in Section 6.4.9 with the counting significance using BDTs for the remaining categories is shown in Table 7.2. This estimate shows that the DisCo NNs also perform better for the VH4L and VH3L categories and perform worse for the remaining categories. This results in an overall decrease in significance from 2.22σ using BDTs, to 2.11σ using NNs. This overall decrease in significance is due to decreases in the VBF and Higgs N-Jet categories, which have the largest contribution to the significance.

An Asimov fit can be used to extract the expected significance for the $H \to \mu\mu$ signal using categories defined using BDTs or NNs. The signal and background distributions for the categories defined using BDTs are parametrized using the same procedure described in Section 6.5. Table 7.3 shows the expected significance from the Asimov fits using categories defined by BDTs, NNs, and the best combination of the two that

Table 7.2: The expected counting significance based on signal and background MC yields with $m_{\mu\mu} \in [120, 130]\,\text{GeV}$ corresponding to $165\,\text{fb}^{-1}$ of data. The categories determined using the outputs from a BDT used in the ATLAS Run-3 analysis and categories determined using DisCo NNs outlined in this thesis. The total counting significance for the BDT column uses the NN $t\bar{t}H$ significance.

Category	BDT	NN	Improvement with NN [%]
VBF	1.218	1.115	−9.2
Higgs 2-Jet	1.036	1.000	−3.6
Higgs 1-Jet	1.076	1.005	−7.1
Higgs 0-Jet	0.984	0.935	−4.8
VH4L	0.137	0.143	+4.2
VH3L	0.293	0.302	+3.0
VH2L	0.144	0.129	−11.6
$t\bar{t}H$	–	0.350	–
Total	2.22	2.11	-5.0

maximizes expected significance. After accounting for differences in the background modelling and the statistical and SS uncertainties, NNs outperform BDTs in the $t\bar{t}H$ and VH3L categories. There is a 16.7% increase in significance for VH3L when using a NN compared to using a BDT. The VH4L category now has a slight decrease in significance using NNs, and all other categories still have a decrease in significance using NNs.

NNs showed a better counting significance for the VH4L category, but have a lower expected Asimov significance than BDTs. This is likely due to the fact that the background model selected for the NN VH4L category has two free parameters (HixEpoly2), while the background model for the BDT VH4L category has one free parameter (HixPower0). This extra degree of freedom in the background model increases the uncertainty on the background shape in the SW and the fit then has more freedom to absorb a small signal-like excess into the background model. This reduces

Table 7.3: The expected significance based on a fit to Asimov datasets corresponding to 165 fb^{-1} of data for categories determined using the outputs from a BDT, NN, or the best option for that category. The total significance for the BDT column uses the NN $t\bar{t}H$ significance.

Category	Expected Significance [σ]		
	BDT	NN	Best
Stat. Only			
VBF	1.13	0.94	1.13
Higgs 2-Jet	0.85	0.74	0.85
Higgs 1-Jet	0.85	0.80	0.85
Higgs 0-Jet	0.82	0.67	0.82
VH4L	0.12	0.11	0.12
VH3L	0.19	0.22	0.22
VH2L	0.11	0.11	0.11
$t\bar{t}H$	–	0.31	0.31
Combined	1.88	1.63	1.89
Stat.+SS			
VBF	1.11	0.91	1.11
Higgs 2-Jet	0.83	0.70	0.83
Higgs 1-Jet	0.82	0.74	0.82
Higgs 0-Jet	0.77	0.65	0.77
VH4L	0.12	0.11	0.12
VH3L	0.18	0.21	0.21
VH2L	0.11	0.10	0.11
$t\bar{t}H$	–	0.29	0.29
Combined	1.82	1.56	1.83

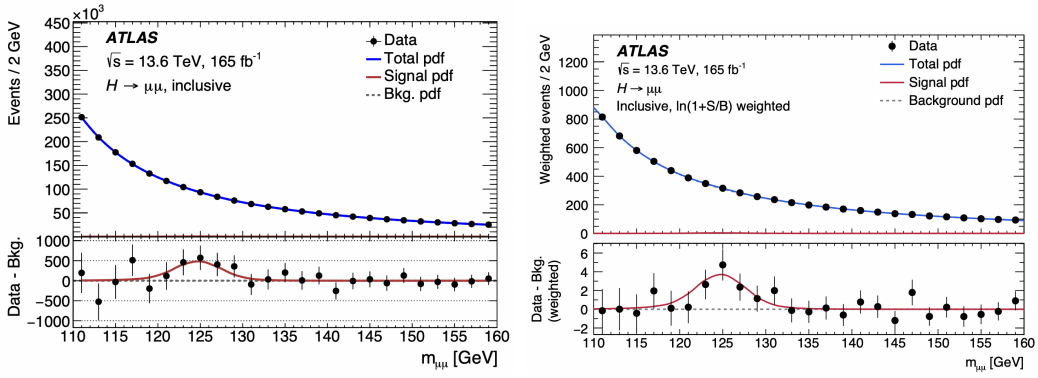
the likelihood difference between the signal-plus-background and background-only hypotheses, resulting in a lower Asimov significance. The NN VH4L category also has a slightly larger SS uncertainty, which would reduce the stat+SS expected significance. Since the NN provides a similar significance to the BDT for the VH4L category, it is possible that optimization of the training variables and NN hyperparameters could allow the NN to outperform the BDT.

The maximum expected significance for the $H \to \mu\mu$ analysis occurs when a NN is used for the $t\bar{t}H$ and VH3L categories, and a BDT is used for classification in the VBF, Higgs 2-Jet, Higgs 1-Jet, Higgs 0-Jet, VH4L, and VH2L categories. Using this scheme, the expected significance is 1.83σ.

7.3 Unblinded Results

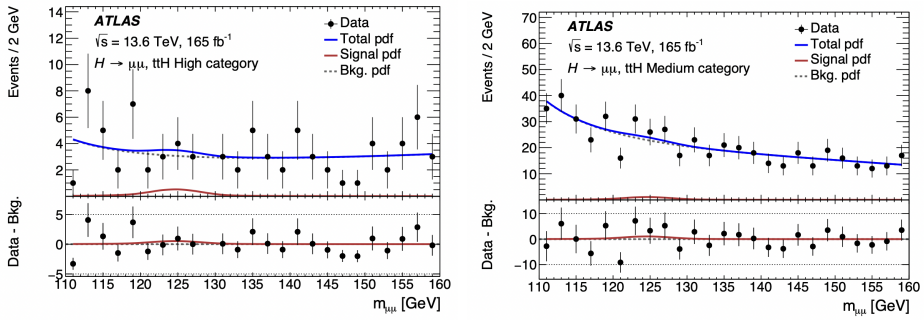
Unblinded results from the ATLAS Run-3 search [2] for $H \to \mu\mu$ using $165\,\text{fb}^{-1}$ of data collected at $\sqrt{s} = 13.6\,\text{TeV}$ are presented in this section. The analysis follows the same approach outlined in this thesis. BDTs are used to classify events in most categories, and the DisCo NNs described in this thesis are used for the $t\bar{t}H$ categories.

A binned maximum likelihood fit to the Run-3 $m_{\mu\mu}$ distributions show an observed (expected) significance of 2.8σ (1.8σ) with respect to the background-only hypothesis. The signal strength, μ, is defined as the ratio of the observed signal yield to the signal yield expected in the SM. The best-fit value of the signal strength parameter is $\mu = 1.6^{+0.6}_{-0.5}(\text{stat.})^{+0.2}_{-0.2}(\text{syst.})$. The results from the fit are shown in Figure 7.2. The fits for the two $t\bar{t}H$ categories defined using a NN with a DisCo loss function can be seen in Figure 7.3.



(a) (b)

Figure 7.2: The observed dimuon invariant mass spectrum combining all 23 Run-3 categories for (a) unweighted events and (b) events weighted by $\log(1 + S/B)$, where S and B are the signal and background yields evaluated in $m_{\mu\mu} \in [120, 130]$ GeV. A fit is performed in the mass range $m_{\mu\mu} \in [110, 160]$ GeV, providing a signal strength of $\mu = 1.6 \pm 0.6$. Figures taken from Ref. [2].



(a) (b)

Figure 7.3: The observed dimuon invariant mass spectra for the Run-3 (a) $t\bar{t}H$-High and (b) $t\bar{t}H$-Medium categories. Figures taken from Ref. [2].

The Run-3 analysis results are combined with the previous Run-2 ATLAS analysis [3]. The Run-2 dataset consists of $140\,\text{fb}^{-1}$ of data collected at $\sqrt{s} = 13\,\text{TeV}$ and the analysis is outlined in Ref. [3]. The analysis follows the same approach defined in this thesis and 20 mutually exclusive categories are defined using BDTs. The experimental uncertainties are taken as uncorrelated between the Run-2 and Run-3 datasets and theory uncertainties are treated as correlated with the exception of those related to PDFs. The combined Run-2 and Run-3 datasets give an observed (expected) significance of 3.4σ (2.5σ), establishing evidence for the $H \to \mu\mu$ decay with the ATLAS detector. The measured signal strength for the combined dataset is $\mu = 1.4^{+0.4}_{-0.4}(\text{stat.})^{+0.1}_{-0.1}(\text{syst.})$. Since $\mu = \sigma_{H\to\mu\mu}/\sigma^{\text{SM}}_{H\to\mu\mu} = \mathcal{B}_{H\to\mu\mu}\,\sigma_H/\sigma^{\text{SM}}_H$ and the Higgs boson cross section has been precisely measured using several other decay channels, the signal strength measurement can be translated into a measurement of the branching fraction of $H \to \mu\mu$. This gives $\mathcal{B}(H \to \mu\mu) = (3.0 \pm 0.9) \times 10^{-4}$.

Chapter 8

Conclusion

This thesis presents the first use of NNs trained using a DisCo loss function for an ATLAS resonance search. These NNs are used to divide events into categories to increase the significance of the $H \to \mu\mu$ signal. The DisCo loss term was found to successfully prevent the NN from introducing shaping into the background distributions. This allowed more variables to be used when training the NN, even if they are correlated with $m_{\mu\mu}$, providing more information to the NN to discriminate between signal and background events. The background distributions in the categories defined by the NN outputs passed rigorous background modelling criteria, indicating that no spurious signal was introduced into the background by the NN.

The NNs outperformed BDTs in terms of expected significance for the $t\bar{t}H$ and VH3L categories, and were competitive in the VH4L category. BDTs provided a higher significance in the VBF, Higgs 2-Jet, Higgs 1-Jet, Higgs 0-Jet and VH2L categories. One possible explanation for this is related to the complexity of the final states targeted in the different categories. The $t\bar{t}H$, $ZH \to \mu\mu\ell\ell$ (targeted in VH4L), and

$WH \to \mu\mu\ell\nu$ (targeted in VH3L) final states involve extra physics objects, including charged leptons, missing transverse momentum, jets, and b-jets. Particularly for $t\bar{t}H$ events, the final state is very busy with two energetic b-jets and two W leptons decaying leptonically or hadronically, which can all be provided to the NN for training. NNs are great for solving complex problems with many inputs that involve non-linear correlations between variables [123]. In contrast, BDTs often perform well with tabular data[1], where a list of input variables are provided for training, and with features that have heavy tails or skewed distributions [124, 125]. This would explain why the BDTs used in this analysis were found to be an excellent choice for the VBF category, where there are minimal final state particles (two muons and two jets) and the events are dominated by a few powerful training variables, such as the pseudorapidity separation of the two forward jets. It is the author's recommendation that future $H \to \mu\mu$ analyses continue to consider DisCo NNs for the $t\bar{t}H$ category, as well as for the VH4L and VH3L categories, which also exhibit more complex correlations among input variables.

It should also be noted that even though the DisCo loss term allows a NN to train on variables correlated with $m_{\mu\mu}$, this term penalizes the NN and reduces its ability to discriminate between signal and background events. If these variables do not greatly improve the discriminating power of the NN, it may be better to train a ML classifier on a reduced list of variables without the punitive DisCo loss term. The DisCo loss term also introduces an additional hyperparameter, λ, which adjusts the strength of the DisCo loss term. This hyperparameter needs to be adjusted based on other parameters, such as the category targeted, the number of training variables, and the batch size used. It is possible that further optimization of the hyperparameter is

[1] as opposed to non-tabular data, such as images or graphs

required for some of the categories in the $H \to \mu\mu$ analysis in order to increase the significance while still preventing shaping of the background.

One potential avenue to improve the significance for the $H \to \mu\mu$ analysis is to provide information on the muon momentum resolution of events to the NN during training. The resolution is different for each muon depending on the momentum and which detector components the muon interacts with. Events with a better muon momentum resolution will produce a narrower Higgs peak in the dimuon invariant mass distribution, resulting in a more distinct peak to extract from the background. The resolution can be provided to the NN as an input variable or as a weight for each event during training. By weighting events by the inverse of the resolution of the dimuon invariant mass, events with a better resolution will be given a stronger consideration by the NN. This approach was used in the Run-2 CMS search for the $H \to \mu\mu$ decay [8].

DisCo NNs were used in the latest ATLAS search for $H \to \mu\mu$ [2]. This search combined Run-2 and Run-3 data collected by the ATLAS detector and saw an observed (expected) significance over the background-only hypothesis of 3.4σ (2.5σ). Using SM values for the Higgs boson production cross section, this corresponds to a branching fraction of $\mathcal{B}(H \to \mu\mu) = (3.0 \pm 0.9) \times 10^{-4}$. This analysis provides ATLAS standalone evidence for the decay of the Higgs boson into two muons, and hence for a Higgs boson Yukawa coupling to a second-generation fermion.

The next step is to establish 5σ significance (discovery) of the $H \to \mu\mu$ decay. To accomplish this with the ATLAS detector, a much larger dataset is required. The LHC will finish Run-3 in 2026, after which it will undergo upgrades for the HL-LHC. The HL-LHC plans to deliver about $4000\,\text{fb}^{-1}$ of data, which will provide sufficient

statistics for discovery of $H \to \mu\mu$. The ATLAS ITk upgrade will be operational during the high luminosity phase of the LHC, and this will provide improved tracking information for the muons used in the $H \to \mu\mu$ analysis.

A key component of the ITk is the ABCStar readout chip for the ITk Strip Detector. On-wafer production testing of the ABCStar was completed in 2025, providing the chips necessary for installation of the ITk. This required testing of over 350,000 chips, half of which were tested by Carleton University through a partnership with DA-Integrated. This partnership with a semiconductor testing company was highly successful and represented a first for a particle physics project. As particle physics continues to move to higher energies and requires larger accelerators and detectors, industrial partnerships such as this one will be necessary to produce and test the components required.

The overall yield of ABCStar chips usable in the ITk through production probing was 85.88%, which was lower than expected due to issues encountered during testing, including a problem with the SRAM of the chip. This issue arose because the timing models used in the design were outdated and based on earlier best practices. Since the SRAM had functioned correctly in earlier prototypes, it was considered silicon proven, giving designers confidence to not recompute the timing models. For future experiments, it is important to recognize that even silicon proven components can exhibit unexpected behaviour, and effort should be made to keep all aspects of the simulation up to date. To improve the ability to identify possible risks as early as possible, prototype wafers should undergo a process called striping, to ensure they cover a representative range of process variations. Properly characterizing striped wafers can reveal potential design issues early and lead to more accurate yield predictions before full-scale production.

The Higgs boson was first discovered over a decade ago at the LHC, and the HL-LHC will provide additional data necessary to better understand its properties. Discovery of the Higgs to dimuon decay to establish a Higgs boson Yukawa coupling to a second generation fermion is a highly anticipated result that is attainable at the HL-LHC. This will provide further insights into the origin of the fermion mass hierarchy and help us understand the mechanism behind mass generation.

References

[1] M. Basso et al.,
A Starry Byte — proton beam measurements of single event upsets and other radiation effects in ABCStar ASIC Versions 0 and 1 for the ITk strip tracker,
Journal of Instrumentation **17** (2022) P03017,
arXiv: 2203.12641 [physics.ins-det].

[2] ATLAS Collaboration, *Evidence for the dimuon decay of the Higgs boson in pp collisions with the ATLAS detector*, Phys. Rev. Lett. **135** (23 2025) 231802,
arXiv: 2507.03595 [hep-ex].

[3] ATLAS Collaboration, *A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector*, Phys. Lett. B **812** (2021) 135980,
arXiv: 2007.07830 [hep-ex].

[4] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,
Phys. Rev. Lett. **13** (16 1964) 508.

[5] F. Englert and R. Brout,
Broken Symmetry and the Mass of Gauge Vector Mesons,
Phys. Rev. Lett. **13** (9 1964) 321.

[6] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*,
Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex].

[7] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30,
arXiv: 1207.7235 [hep-ex].

[8] CMS Collaboration, *Evidence for Higgs boson decay to a pair of muons*,
JHEP **01** (2021) 148, arXiv: 2009.04363 [hep-ex].

[9] M. Thomson, *Modern particle physics*,
New York: Cambridge University Press, 2013,
ISBN: 978-1-107-03426-6, 978-1-139-52536-7.

[10] D. J. Griffiths, *Introduction to Elementary Particles*, John Wiley & Sons, 1987,
ISBN: 978-0-471-60386-3, 978-3-527-61846-0.

[11] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Reading, USA: Addison-Wesley, 1995, ISBN: 978-0-201-50397-5, 978-0-429-50355-9, 978-0-429-49417-8.

[12] S. Navas et al., *Review of particle physics*, Phys. Rev. D **110** (2024) 030001.

[13] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv: 1503.07589 [hep-ex].

[14] Y. Fukuda et al., *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81** (8 1998) 1562.

[15] Q. R. Ahmad et al., *Measurement of the Rate of $\nu_e + d \to p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (7 2001) 071301.

[16] Q. R. Ahmad et al., *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (1 2002) 011301.

[17] E. Noether, *Invariant variation problems*, Transport Theory and Statistical Physics **1** (1971) 186.

[18] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579.

[19] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264.

[20] T. Saito and K. Shigemoto, *Weak and Electromagnetic Interactions in Superconductivity Model*, Prog. Theor. Phys. **57** (1977) 242.

[21] A. Karlberg et al., *Ad interim recommendations for the Higgs boson production cross sections at $\sqrt{s} = 13.6$ TeV*, (2024), arXiv: 2402.09955 [hep-ph].

[22] ATLAS Collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, Nature **607** (2022) 52, arXiv: 2207.00092 [hep-ex].

[23] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, en, 2012, URL: http://cds.cern.ch/record/1416519.

[24] G. Antchev et al., *First measurement of elastic, inelastic and total cross-section at s=13 TeV by TOTEM and overview of cross-section data at LHC energies: TOTEM Collaboration*, European Physical Journal C **79** (2019), arXiv: 1712.06153 [hep-ex].

[25] F. Gross et al.,
50 Years of quantum chromodynamics: Introduction and Review,
The European Physical Journal C **83** (2023), arXiv: 2212.11107 [hep-ph].

[26] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*,
Nuclear Physics B **126** (1977) 298.

[27] J. Gao, L. Harland-Lang, and J. Rojo,
The structure of the proton in the LHC precision era,
Physics Reports **742** (2018) 1, arXiv: 1709.04922 [hep-ph].

[28] L. Lönnblad, *Correcting the Colour-Dipole Cascade Model with Fixed Order Matrix Elements*, Journal of High Energy Physics **2002** (2002) 046.

[29] L. Lönnblad and S. Prestel,
Matching tree-level matrix elements with interleaved showers,
Journal of High Energy Physics **2012** (2012), arXiv: 1109.4829 [hep-ph].

[30] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*,
Journal of High Energy Physics **2012** (2012), arXiv: 1209.6215 [hep-ph].

[31] C. Bierlich et al.,
A comprehensive guide to the physics and usage of PYTHIA 8.3, (2022),
arXiv: 2203.11601 [hep-ph].

[32] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand,
Parton fragmentation and string dynamics, Physics Reports **97** (1983) 31.

[33] B. Andersson, *The Lund Model*, vol. 7, Cambridge University Press, 1998,
ISBN: 978-1-009-40129-6, 978-1-009-40125-8, 978-1-009-40128-9,
978-0-521-01734-3, 978-0-521-42094-5, 978-0-511-88149-7.

[34] B. R. Webber,
A QCD Model for Jet Fragmentation Including Soft Gluon Interference,
Nucl. Phys. B **238** (1984) 492.

[35] O. S. Brüning et al., *LHC Design Report*, CERN Yellow Reports: Monographs,
Geneva: CERN, 2004, URL: https://cds.cern.ch/record/815187.

[36] E. Mobs, *The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019*, 2019, URL: https://cds.cern.ch/record/2684277.

[37] A. Lopes and M. L. Perrey, "FAQ-LHC The guide," 2022,
URL: https://cds.cern.ch/record/2809109.

[38] CERN, *Project Schedule*, Accessed: 2025-01-12, URL: https://project-hl-lhc-industry.web.cern.ch/content/project-schedule.

[39] ATLAS Collaboration,
The ATLAS Experiment at the CERN Large Hadron Collider,
JINST **3** (2008) S08003.

[40] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[41] ALICE Collaboration, *The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment*, JINST **3** (2008) S08002.

[42] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[43] CERN, *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*, CERN Yellow Reports: Monographs, Geneva: CERN, 2020, URL: https://cds.cern.ch/record/2749422.

[44] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3*, (2023), arXiv: 2305.16623 [physics.ins-det].

[45] ATLAS Collaboration, *Beam Spot Public Results*, Accessed: 2025-07-29, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BeamSpotPublicResults.

[46] P. Vankov, *ATLAS Upgrade for the HL-LHC: meeting the challenges of a five-fold increase in collision rate*, EPJ Web of Conferences **28** (2012) 12069.

[47] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, Journal of Instrumentation **9** (2014) P08009.

[48] ATLAS Collaboration, *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*, Journal of Instrumentation **3** (2008) P02013.

[49] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter: Technical Design Report*, ATLAS-TDR-2; CERN-LHCC-96-041, 1996, URL: https://cds.cern.ch/record/331061.

[50] R. C. Fernow, *Introduction to experimental particle physics*, Cambridge university press, 1986.

[51] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, ATLAS-TDR-3; CERN-LHCC-96-042, 1996, URL: https://cds.cern.ch/record/331062.

[52] T. Davidek, *Performance and Calibration of the ATLAS Tile Calorimeter*, Instruments **6** (2022).

[53] ATLAS Collaboration, *ATLAS Muon Spectrometer: Technical Design Report*, ATLAS-TDR-10; CERN-LHCC-97-022, CERN, 1997, URL: https://cds.cern.ch/record/331068.

[54] ATLAS Collaboration, *ATLAS New Small Wheel: Technical Design Report*, ATLAS-TDR-020; CERN-LHCC-2013-006, 2013, URL: https://cds.cern.ch/record/1552862.

[55] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017.

[56] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, The European Physical Journal C **83** (2023).

[57] ATLAS Collaboration, *Luminosity public results – Run 2*, 2025, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[58] ATLAS Collaboration, *Luminosity public results – Run 3*, 2025, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3.

[59] ATLAS Collaboration, *A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade: Technical Design Report*, ATLAS-TDR-031; CERN-LHCC-2020-007, 2020, URL: https://cds.cern.ch/record/2719855.

[60] ATLAS Collaboration, *ATLAS Muon Spectrometer Phase-II Upgrade: Technical Design Report*, ATLAS-TDR-026; CERN-LHCC-2017-017, 2017, URL: https://cds.cern.ch/record/2285580.

[61] ATLAS Collaboration, *ATLAS LAr Calorimeter Phase-II Upgrade: Technical Design Report*, ATLAS-TDR-027; CERN-LHCC-2017-018, 2017, URL: https://cds.cern.ch/record/2285582.

[62] ATLAS Collaboration, *ATLAS Tile Calorimeter Phase-II Upgrade: Technical Design Report*, ATLAS-TDR-028; CERN-LHCC-2017-019, 2017, URL: https://cds.cern.ch/record/2285583.

[63] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design Report*, ATLAS-TDR-025; CERN-LHCC-2017-005, 2017, URL: https://cds.cern.ch/record/2257755.

[64] ATLAS Collaboration,
ATLAS Inner Tracker Pixel Detector: Technical Design Report,
ATLAS-TDR-030; CERN-LHCC-2017-021, 2017,
URL: https://cds.cern.ch/record/2285585.

[65] W. Lu et al., *Development of the ABCStar front-end chip for the ATLAS silicon strip upgrade*, Journal of Instrumentation **12** (2017) C04017.

[66] K. Cormier et al., *Development of the front end amplifier circuit for the ATLAS ITk silicon strip detector*,
Journal of Instrumentation **16** (2021) P07061.

[67] R. G. Larson and T. J. Rehg, "Spin Coating,"
Liquid Film Coating: Scientific principles and their technological implications,
ed. by S. F. Kistler and P. M. Schweizer,
Dordrecht: Springer Netherlands, 1997 709, ISBN: 978-94-011-5342-3,
URL: https://doi.org/10.1007/978-94-011-5342-3_20.

[68] F. Faccio and G. Cervelli,
Radiation-induced edge effects in deep submicron CMOS transistors,
IEEE Transactions on Nuclear Science **52** (2005) 2413.

[69] F. Faccio et al., *Total ionizing dose effects in shallow trench isolation oxides*,
Microelectronics Reliability **48** (2008) 1000, 2007 Reliability of Compound Semiconductors (ROCS) Workshop.

[70] J. John, *ABCStar, the ATLAS ITk Strip Analog Front-End ASIC Production*,
https://cds.cern.ch/record/2885951/files/ATL-ITK-SLIDE-2023-668.pdf, Presented at the 13th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors, 2023.

[71] R. Goering, *Variability upends designers' plans*,
Electronic Engineering Times (2005) 39.

[72] M. Choi, S. Kim, T. Noh, D. Kang, and S. Jung, *Si Characterization on Thinning and Singulation Processes for 2.5/3D HBM Package Integration*,
Materials **17** (2024) 5529.

[73] ATLAS Collaboration,
Prospects for the measurement of the rare Higgs boson decay $H \to \mu\mu$ with 3000 fb^{-1} of pp collisions collected at $\sqrt{s} = 14$ TeV by the ATLAS experiment,
ATL-PHYS-PUB-2018-006, CERN, 2018,
URL: https://cds.cern.ch/record/2319741.

[74] T. Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*,
Journal of Physics: Conference Series **119** (2008) 032014.

[75] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*,
Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex].

[76] ATLAS Collaboration,
Software Performance of the ATLAS Track Reconstruction for LHC Run 3,
Computing and Software for Big Science **8** (2024).

[77] ATLAS Collaboration, *ATLAS Tracking Software Tutorial 2024*,
Accessed: 2025-07-09,
URL: https://atlassoftwaredocs.web.cern.ch/internal-links/tracking-tutorial/idoverview/.

[78] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*,
Nucl. Instrum. Meth. A **262** (1987) 444.

[79] J. Illingworth and J. Kittler, *A survey of the hough transform*,
Computer Vision, Graphics, and Image Processing **44** (1988) 87.

[80] I. Sanderswood,
Development of ATLAS Primary Vertex Reconstruction for LHC Run 3,
(2019), arXiv: 1910.08405 [hep-ex].

[81] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490,
arXiv: 1603.02934 [hep-ex].

[82] "Performance of the ATLAS Particle Flow Algorithm for Jets in Run 2," 2017,
URL: https://cds.cern.ch/record/2255735.

[83] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV*,
Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex].

[84] ATLAS Collaboration, *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*,
ATL-PHYS-PUB-2017-022, CERN, 2017,
URL: https://cds.cern.ch/record/2298955.

[85] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*,
Journal of High Energy Physics **2008** (2008) 063,
arXiv: 0802.1189 [hep-ph].

[86] R. Atkin, *Review of jet reconstruction algorithms*,
Journal of Physics: Conference Series **645** (2015) 012008.

[87] M. Aaboud et al.,
Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,
Phys. Rev. D **96** (2017) 072002, arXiv: 1703.09665.

[88] ATLAS Collaboration, *Transforming jet flavour tagging at ATLAS*, (2025),
arXiv: 2505.19689 [hep-ex].

[89] ATLAS Collaboration,
Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment,
Journal of High Energy Physics **2024** (2024), arXiv: 2308.13362 [hep-ex].

[90] ATLAS Collaboration, *The performance of missing transverse momentum reconstruction and its significance with the ATLAS detector using 140 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions*, The European Physical Journal C **85** (2025), arXiv: 2402.05858 [hep-ex].

[91] ATLAS Collaboration,
Public ATLAS Online Luminosity Plots for Run-3 of the LHC,
Accessed: 2025-07-08, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3.

[92] S. Alioli, P. Nason, C. Oleari, and E. Re,
A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,
Journal of High Energy Physics **2010** (2010), arXiv: 1002.2581 [hep-ph].

[93] J. Alwall et al.,
The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,
Journal of High Energy Physics **2014** (2014), arXiv: 1405.0301 [hep-ph].

[94] R. Frederix et al.,
The automation of next-to-leading order electroweak calculations,
Journal of High Energy Physics **2018** (2018), arXiv: 1804.10017 [hep-ph].

[95] D. J. Lange, *The EvtGen particle decay simulation package*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) 152,
BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[96] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*,
ATL-PHYS-PUB-2014-021, CERN, 2014,
URL: https://cds.cern.ch/record/1966419.

[97] R. D. Ball et al., *Parton distributions with LHC data*,
Nuclear Physics B **867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[98] E. Bothmann et al., *Event generation with Sherpa 2.2*,
SciPost Physics **7** (2019), arXiv: 1905.09127 [hep-ph].

[99] R. D. Ball et al., *Parton distributions for the LHC run II*,
Journal of High Energy Physics **2015** (2015), arXiv: 1410.8849 [hep-ph].

[100] S. Agostinelli et al., *Geant4—a simulation toolkit*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250.

[101] K. Werner, F.-M. Liu, and T. Pierog,
Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at the BNL Relativistic Heavy Ion Collider,
Physical Review C **74** (2006).

[102] ATLAS Collaboration,
Software and computing for Run 3 of the ATLAS experiment at the LHC,
The European Physical Journal C **85** (2025) 234,
arXiv: 2404.06335 [hep-ex].

[103] ATLAS Collaboration,
The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model,
ATL-PHYS-PUB-2016-017, CERN, 2016,
URL: https://cds.cern.ch/record/2206965.

[104] A. Buckley, *Computational challenges for MC event generation*,
Journal of Physics: Conference Series **1525** (2020) 012023,
arXiv: 1908.00167 [hep-ph].

[105] K. Danziger, S. Höche, and F. Siegert,
Reducing negative weights in Monte Carlo event generation with Sherpa,
(2021), arXiv: 2110.15211 [hep-ph].

[106] W. Buttinger, *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*,
ATL-COM-SOFT-2015-119, CERN, 2015,
URL: https://cds.cern.ch/record/2014726.

[107] G. Cowan, K. Cranmer, E. Gross, and O. Vitells,
Asymptotic formulae for likelihood-based tests of new physics,
The European Physical Journal C **71** (2011) 1554,
arXiv: 1007.1727 [physics.data-an].

[108] C. C. Aggarwal et al., *Neural networks and deep learning*, vol. 10, 978,
Springer, 2018.

[109] G. J. Székely, M. L. Rizzo, and N. K. Bakirov,
Measuring and testing dependence by correlation of distances,
The Annals of Statistics **35** (2007) 2769.

[110] G. J. Székely and M. L. Rizzo, *Brownian distance covariance*,
The Annals of Applied Statistics **3** (2009) 1236.

[111] G. J. Székely and M. L. Rizzo,
The distance correlation t-test of independence in high dimension,
Journal of Multivariate Analysis **117** (2013) 193.

[112] G. J. Székely and M. L. Rizzo,
Partial distance correlation with methods for dissimilarities,
The Annals of Applied Statistics **42** (2014) 2382.

[113] G. Kasieczka and D. Shih, *Robust Jet Classifiers through Distance Correlation*,
Physical Review Letters **125** (2020).

[114] S. Han and H. D. Kim, *Elaborating Higgs to dimuon decay from gluon fusion by decorrelation and jet substructure*,
The European Physical Journal C **85** (2025) 324,
arXiv: 2406.11961 [hep-ph].

[115] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov,
Improving neural networks by preventing co-adaptation of feature detectors,
(2012), arXiv: 1207.0580 [cs.NE].

[116] F. Chollet et al., *Keras*, https://keras.io, 2015.

[117] Martín Abadi et al.,
TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,
Software available from tensorflow.org, 2015,
URL: https://www.tensorflow.org/.

[118] J. C. Collins and D. E. Soper,
Angular distribution of dileptons in high-energy hadron collisions,
Phys. Rev. D **16** (7 1977) 2219.

[119] ATLAS Collaboration,
Search for the neutral Higgs bosons of the Minimal Supersymmetric Standard Model in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector,
JHEP **02** (2013) 095, arXiv: 1211.6956 [hep-ex].

[120] M. Aaboud et al., *Measurement of the Drell-Yan triple-differential cross section in pp collisions at $\sqrt{s} = 8$ TeV*,
Journal of High Energy Physics **2017** (2017).

[121] V. Bertone, S. Carrazza, and J. Rojo,
APFEL: A PDF evolution library with QED corrections,
Computer Physics Communications **185** (2014) 1647,
arXiv: 1310.1394 [hep-ph].

[122] A. Buckley et al., *LHAPDF6: parton density access in the LHC precision era*,
The European Physical Journal C **75** (2015), arXiv: 1412.7420 [hep-ph].

[123] P. Baldi, P. Sadowski, and D. Whiteson,
Searching for exotic particles in high-energy physics with deep learning,
Nature Communications **5** (2014) 4308, arXiv: 1402.4735 [hep-ph].

[124] R. Shwartz-Ziv and A. Armon, *Tabular data: Deep learning is not all you need*,
Information Fusion **81** (2022) 84, arXiv: 2106.03253 [cs.LG].

[125] D. McElfresh et al.,
When do neural nets outperform boosted trees on tabular data?
Advances in Neural Information Processing Systems **36** (2023) 76336,
arXiv: 2305.02997 [cs.LG].

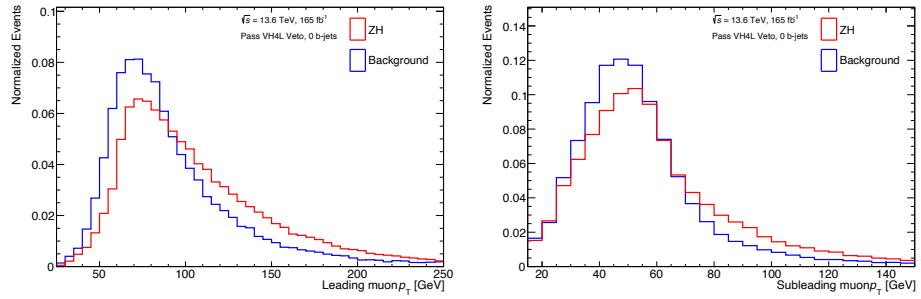
Appendix A

Distributions of NN Training Variables

The signal and background distributions for variables used in training the NNs outlined in Chapter 6 are presented here.

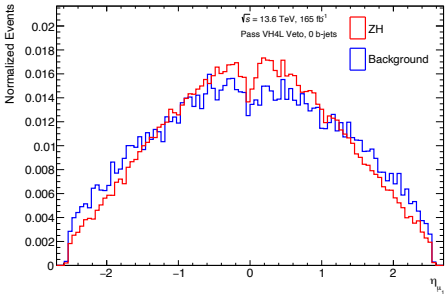
A.1 VH4L Training Variables

MC signal and background distributions for the 14 features used to train the NN VH4L classifier are shown in Figures A.1 and A.2.

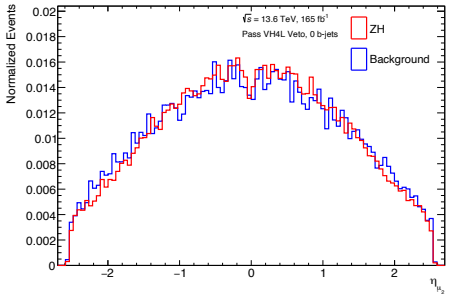


(a) Leading muon p_{T}.

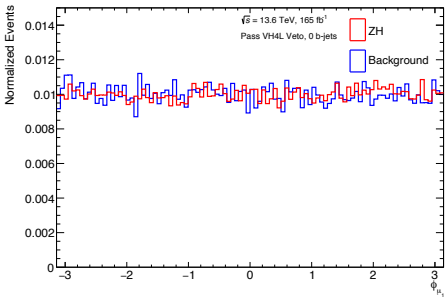
(b) Subleading muon p_{T}.



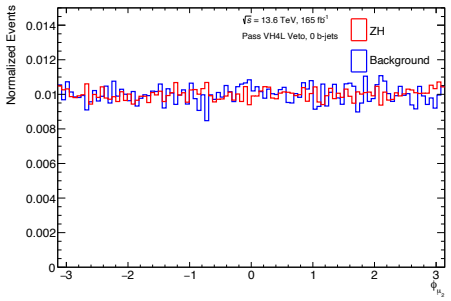
(c) Leading muon η.



(d) Subleading muon η.

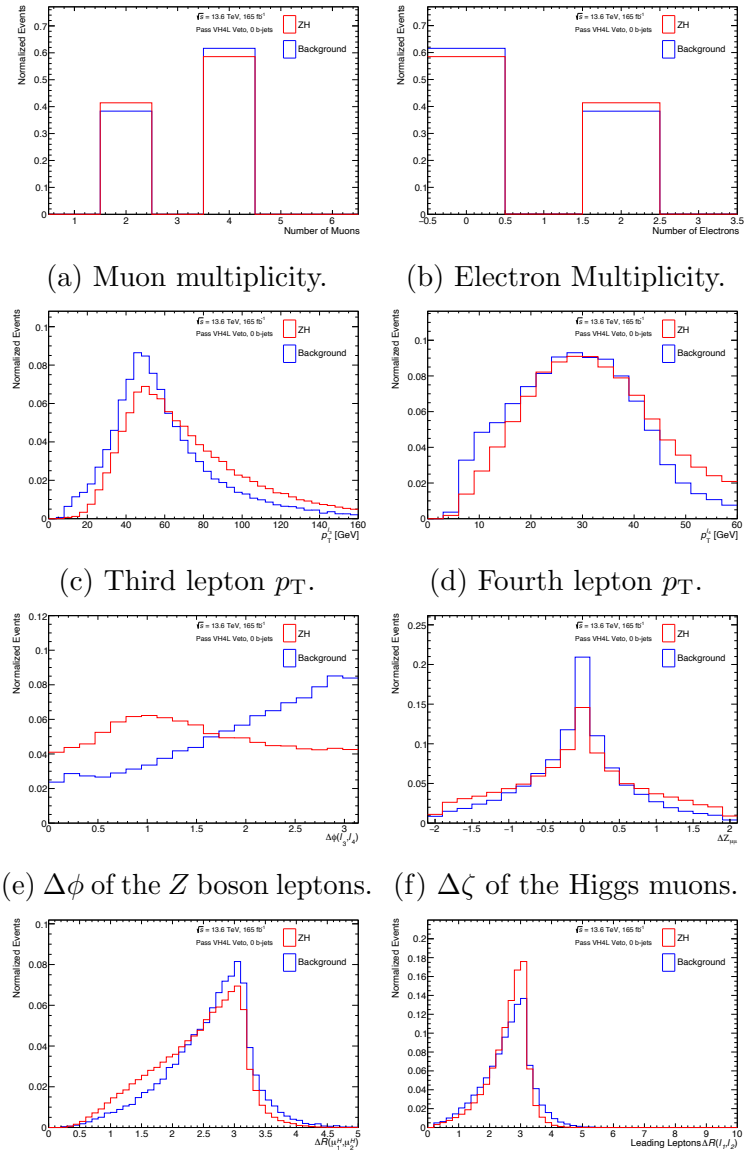


(e) Leading muon ϕ.



(f) Subleading muon ϕ

Figure A.1: Distributions of features obtained from muons that are used to train the NN VH4L classifier. The muons from ZH signal events tend to have a higher transverse momentum than background processes.



(a) Muon multiplicity.

(b) Electron Multiplicity.

(c) Third lepton p_{T}.

(d) Fourth lepton p_{T}.

(e) $\Delta\phi$ of the Z boson leptons.

(f) $\Delta\zeta$ of the Higgs muons.

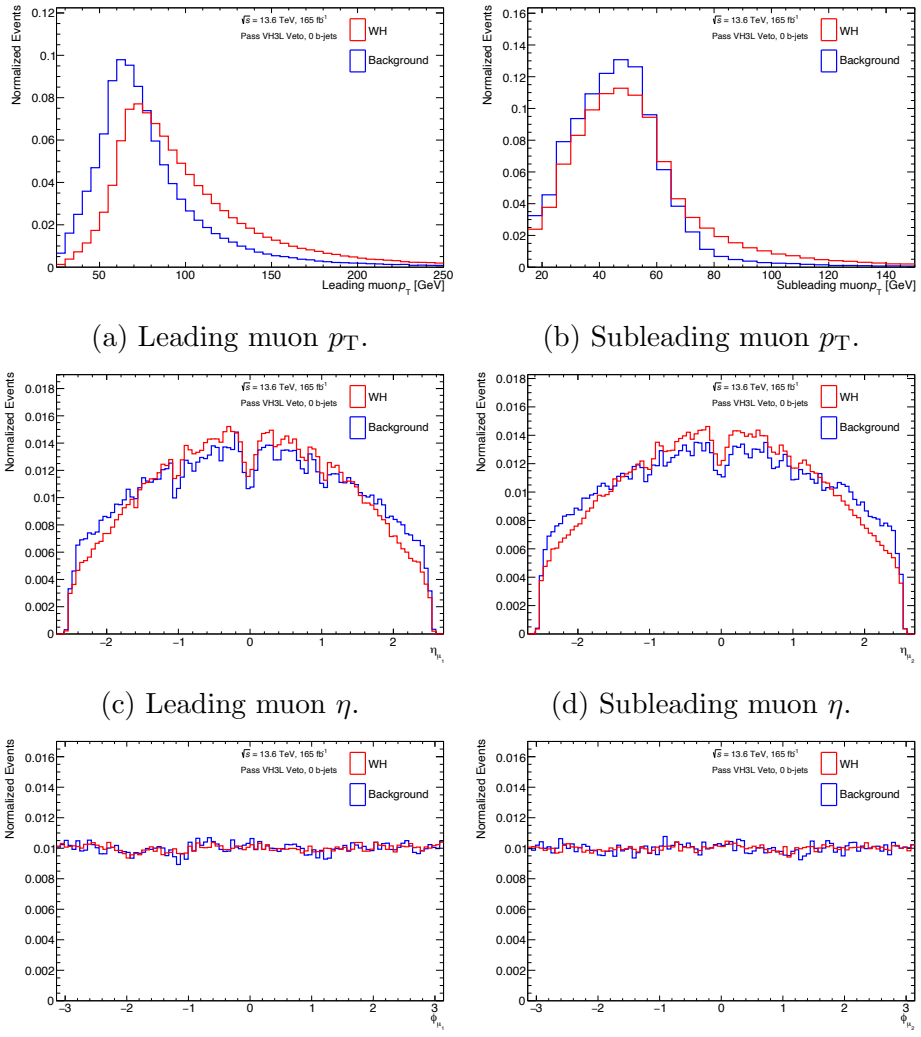
(g) Angular separation of the Higgs boson muons.

(h) Angular separation of the two leading leptons.

Figure A.2: Distributions of additional leptonic features that are used to train the NN VH4L classifier. Additional leptons in ZH signal events tend to have a higher p_{T} than background events. The azimuthal separation of the Z candidate leptons, the angular separation of the Higgs candidate leptons and the angular separation of the two leading leptons tends to be lower in ZH signal events compared to background events. $\Delta\zeta$ is broader for ZH signal events.

A.2 VH3L Training Variables

MC signal and background distributions for the 20 features used to train the NN VH3L classifier are shown in Figures A.3, A.4, A.5, and A.6.



(a) Leading muon p_{T}.

(b) Subleading muon p_{T}.

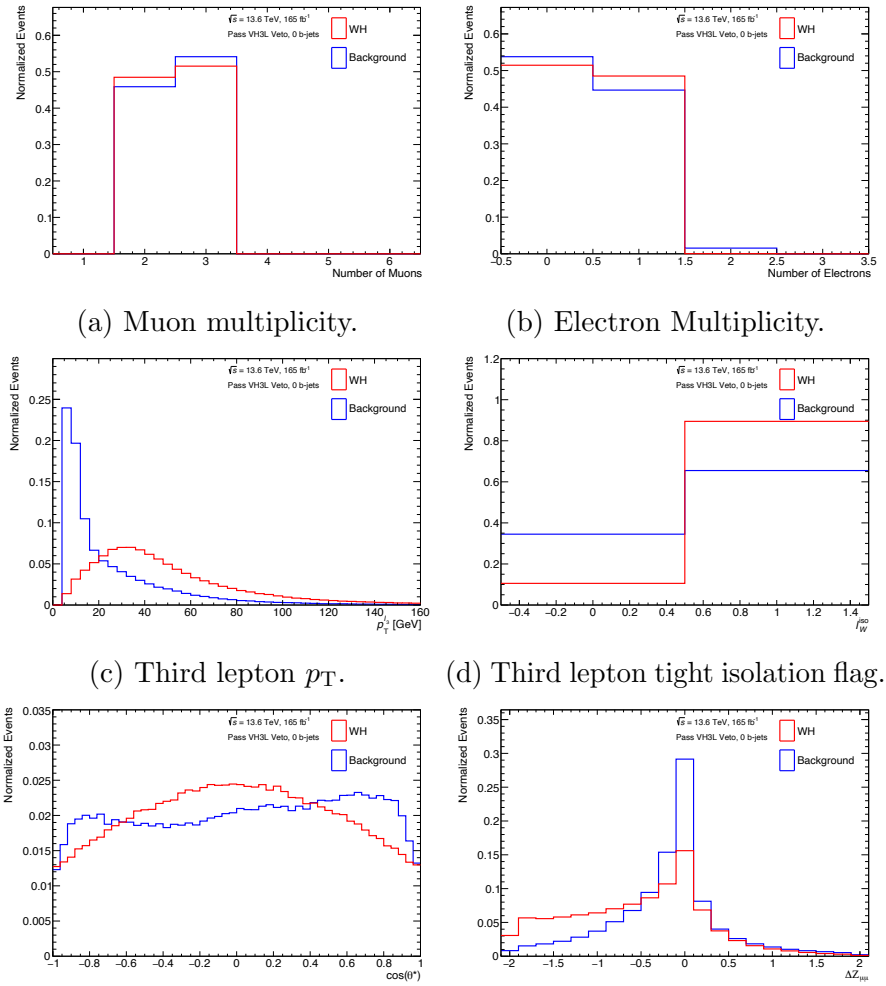
(c) Leading muon η.

(d) Subleading muon η.

(e) Leading muon ϕ.

(f) Subleading muon ϕ.

Figure A.3: Distributions of features obtained from muons that are used to train the NN VH3L classifier. The muons from WH signal events tend to have a higher momentum than background processes.



(a) Muon multiplicity.

(b) Electron Multiplicity.

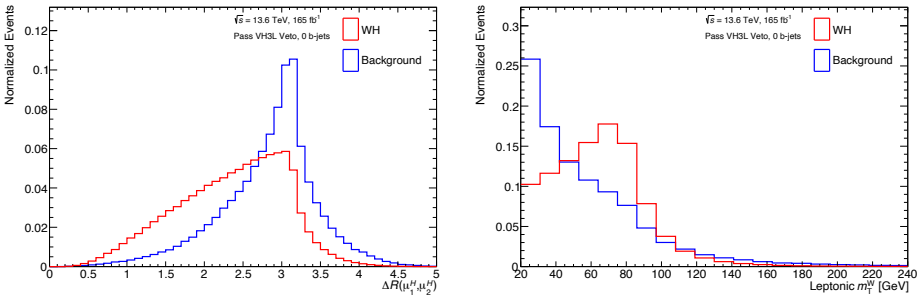
(c) Third lepton p_{T}.

(d) Third lepton tight isolation flag.

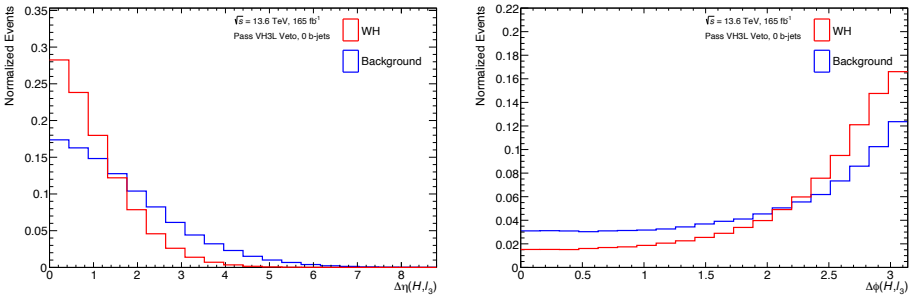
(e) Dimuon system $\cos\theta^*$.

(f) $\Delta\zeta$ of the Higgs muons.

Figure A.4: Distributions of leptonic features that are used to train the NN VH3L classifier. Additional leptons in WH signal events tend to have a higher p_{T} than background events and are more likely to pass the tight isolation working point. $\Delta\zeta$ is broader for WH signal events and the $\cos\theta^*$ distributions are different for signal and background events.

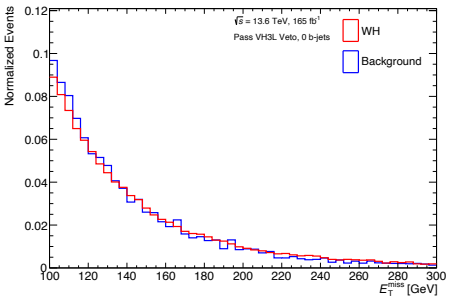


(a) ΔR of the Higgs boson muons. (b) Transverse mass of the W boson.

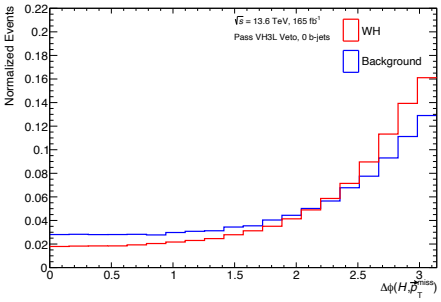


(c) Pseudorapidity separation of the Higgs boson and W lepton. (d) Azimuthal separation of the Higgs boson and W lepton.

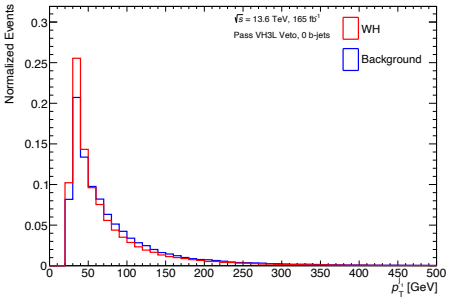
Figure A.5: Distributions of features that are used to train the NN VH3L classifier. The angular separation of the Higgs candidate leptons tends to be lower in WH signal events compared to background events. There is a peak in the leptonic transverse mass around the W mass in signal events. The pseudorapidity separation of the Higgs candidate and the W candidate lepton tends to be lower in WH signal events compared to background events.



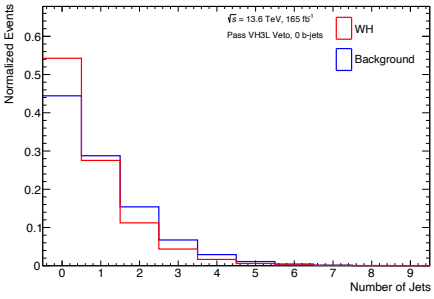
(a) Missing transverse momentum.



(b) $\Delta\phi$ of the Higgs boson and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.



(c) Leading jet p_{T}.

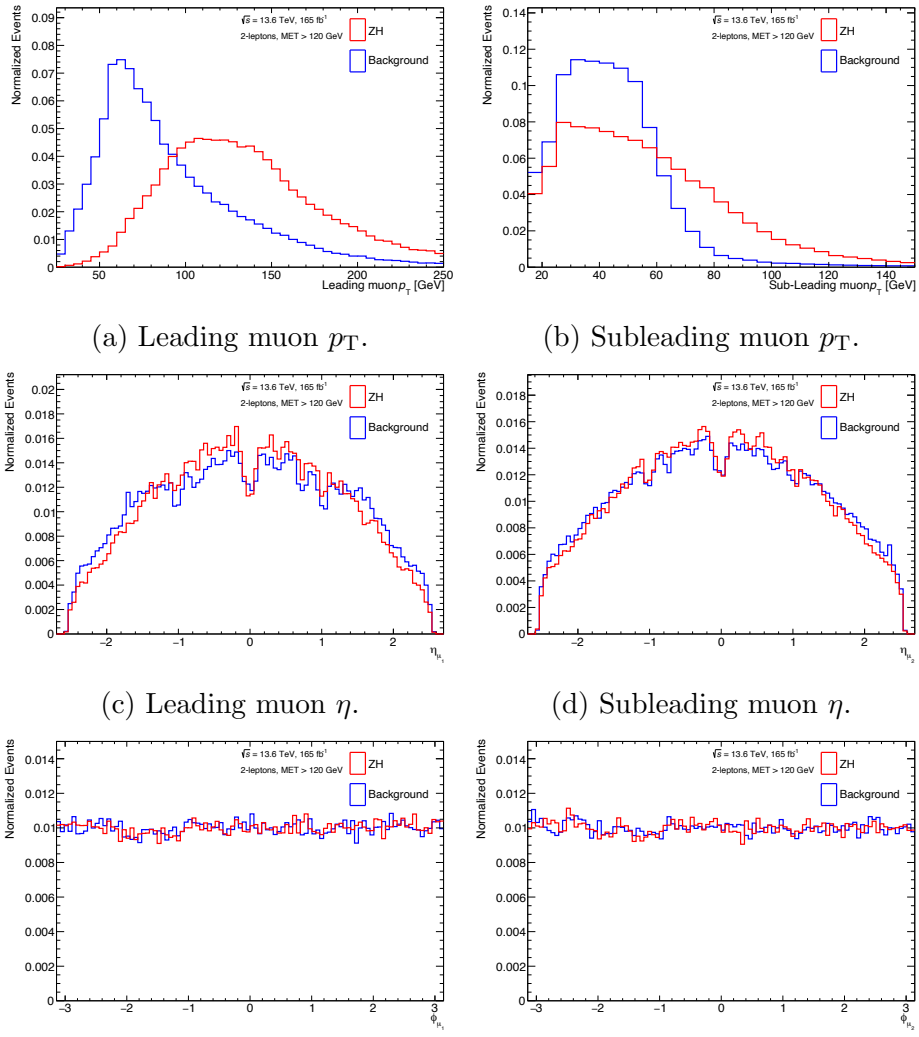


(d) Jet multiplicity.

Figure A.6: Distributions of features that are used to train the NN VH3L classifier. The azimuthal separation between the Higgs boson and W lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ tends to be higher for WH signal events compared to background events. WH signal events tend to have a lower jet multiplicity than background events and the leading jet tends to be slightly softer in signal events.

A.3 VH2L Training Variables

MC signal and background distributions for the 15 features used to train the NN VH2L classifier are shown in Figures A.7, A.8, and A.9.



(a) Leading muon p_{T}.

(b) Subleading muon p_{T}.

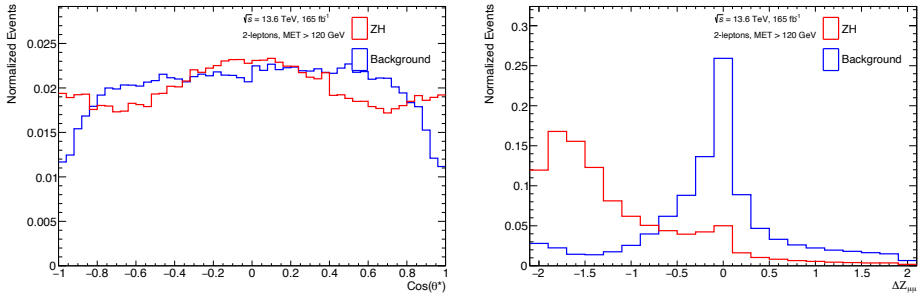
(c) Leading muon η.

(d) Subleading muon η.

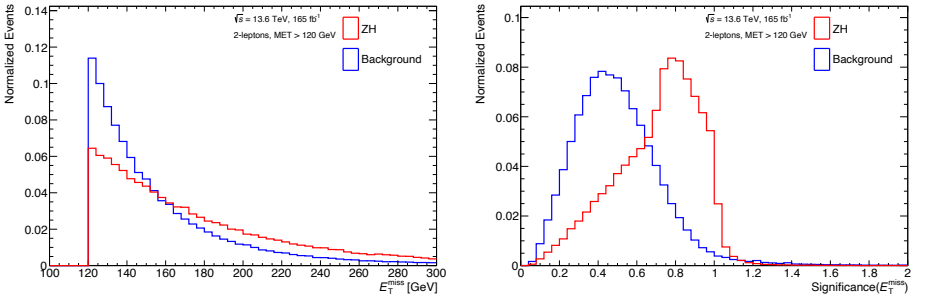
(e) Leading muon ϕ.

(f) Subleading muon ϕ.

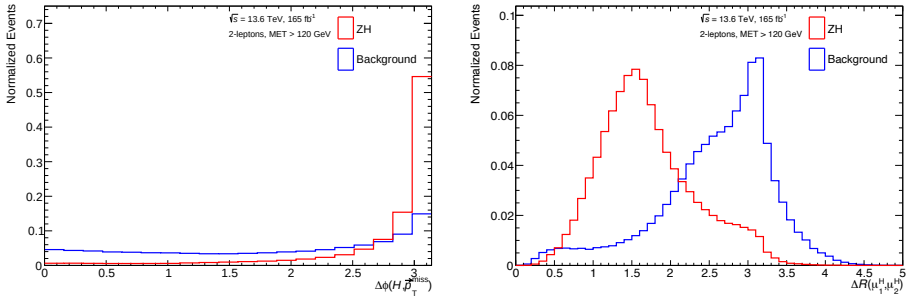
Figure A.7: Distributions of features obtained from muons that are used to train the NN VH2L classifier. The muons from ZH signal events tend to have a higher momentum than background processes.



(a) Dimuon system $\cos\theta^*$. (b) $\Delta\zeta$ of the Higgs candidate muons.

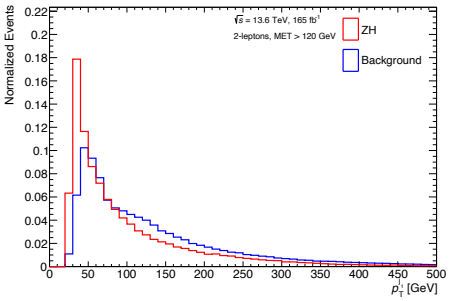


(c) Missing transverse momentum. (d) Significance of the $E_{\mathrm{T}}^{\mathrm{miss}}$.

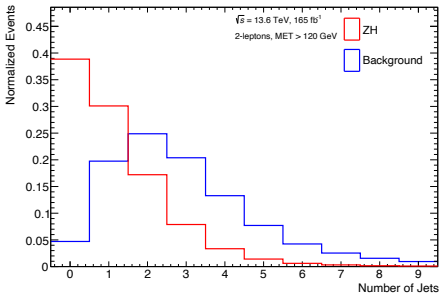


(e) Azimuthal separation of the Higgs boson and $E_{\mathrm{T}}^{\mathrm{miss}}$. (f) Angular separation of the Higgs boson muons.

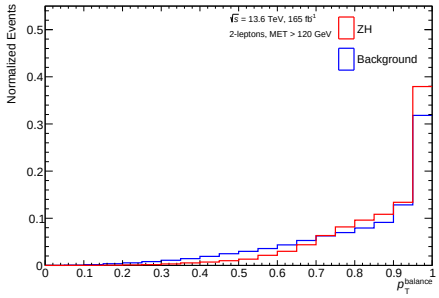
Figure A.8: Distributions of features that are used to train the NN VH2L classifier. ZH signal events have a distinctly different $\Delta\zeta$ distribution. ZH signal events tend to have a higher $E_{\mathrm{T}}^{\mathrm{miss}}$, a higher $E_{\mathrm{T}}^{\mathrm{miss}}$ significance and the azimuthal separation between the Higgs boson and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ tends to be higher than for background processes. The angular separation of the Higgs candidate muons is lower in signal events compared to background events.



(a) Leading jet p_{T}.



(b) Jet multiplicity.

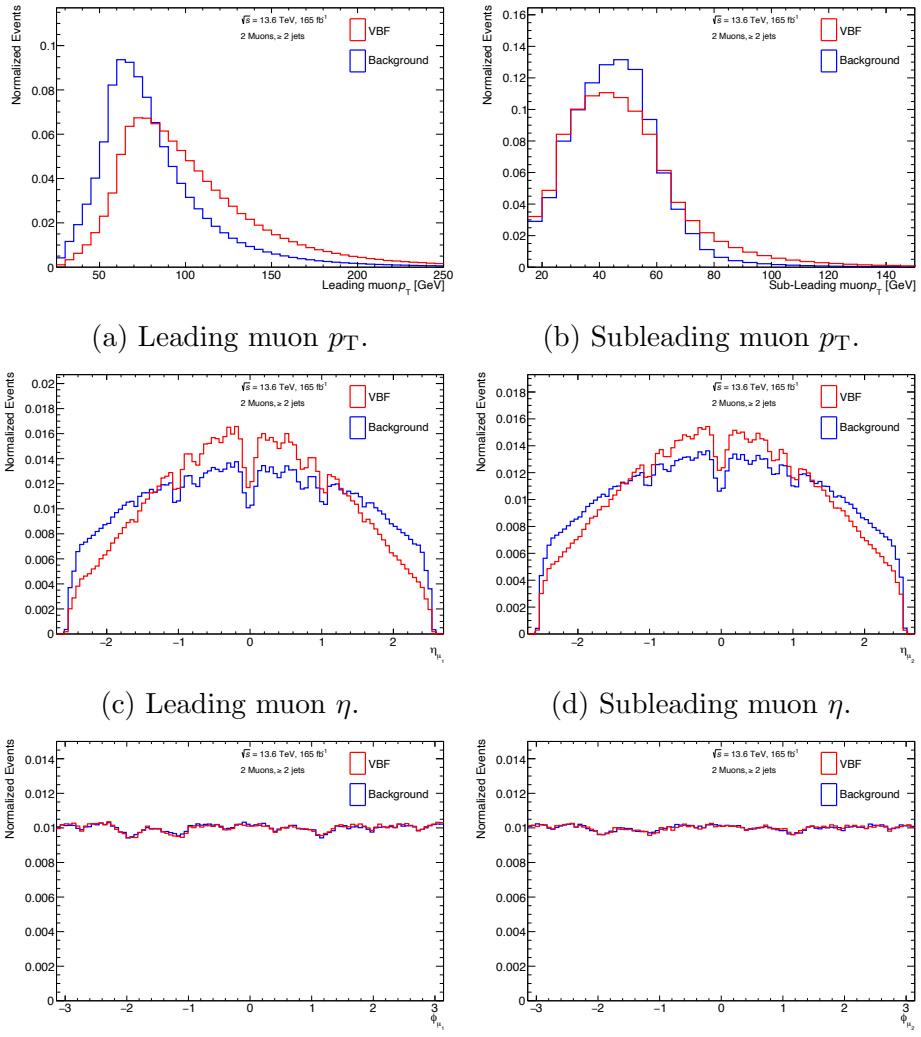


(c) p_{T} balance of the dimuon and dijet system.

Figure A.9: Distributions of features that are used to train the NN VH2L classifier. ZH signal events tend to have a lower jet multiplicity, a lower p_{T} for the leading jet, and a different p_{T} balance than background events.

A.4 VBF Training Variables

MC signal and background distributions for the 23 features used to train the NN VBF classifier are shown in Figures A.10, A.11, A.12, and A.13.



(a) Leading muon p_{T}.

(b) Subleading muon p_{T}.

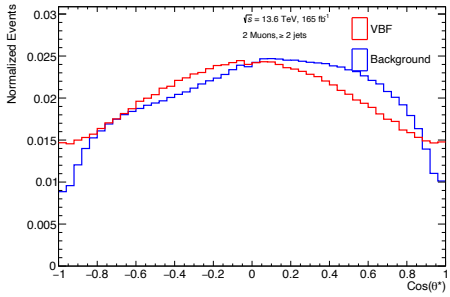
(c) Leading muon η.

(d) Subleading muon η.

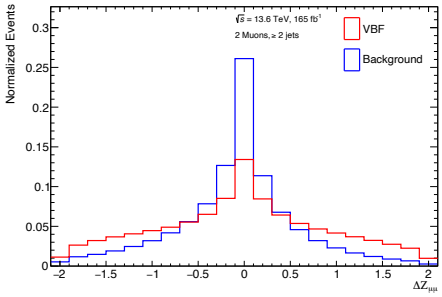
(e) Leading muon ϕ.

(f) Subleading muon ϕ.

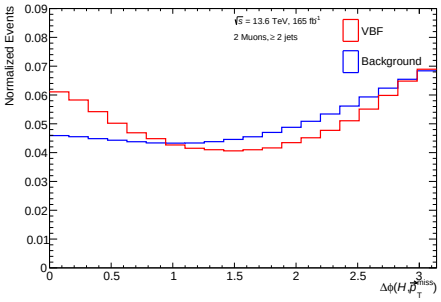
Figure A.10: Distributions of features obtained from muons that are used to train the NN VBF classifier. The muons from VBF signal events tend to have a higher momentum than background processes.



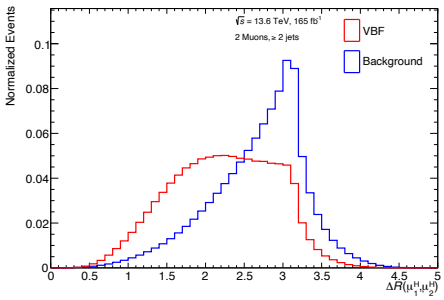
(a) Dimuon system $\cos\theta^*$.



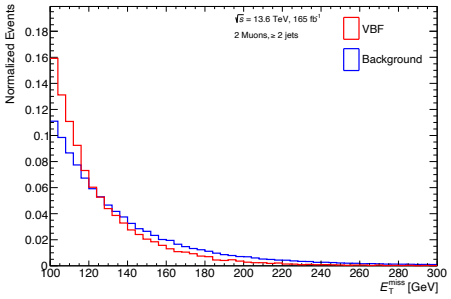
(b) $\Delta\zeta$ of the Higgs candidate muons.



(c) Azimuthal separation of the Higgs boson and $E_{\mathrm{T}}^{\mathrm{miss}}$.

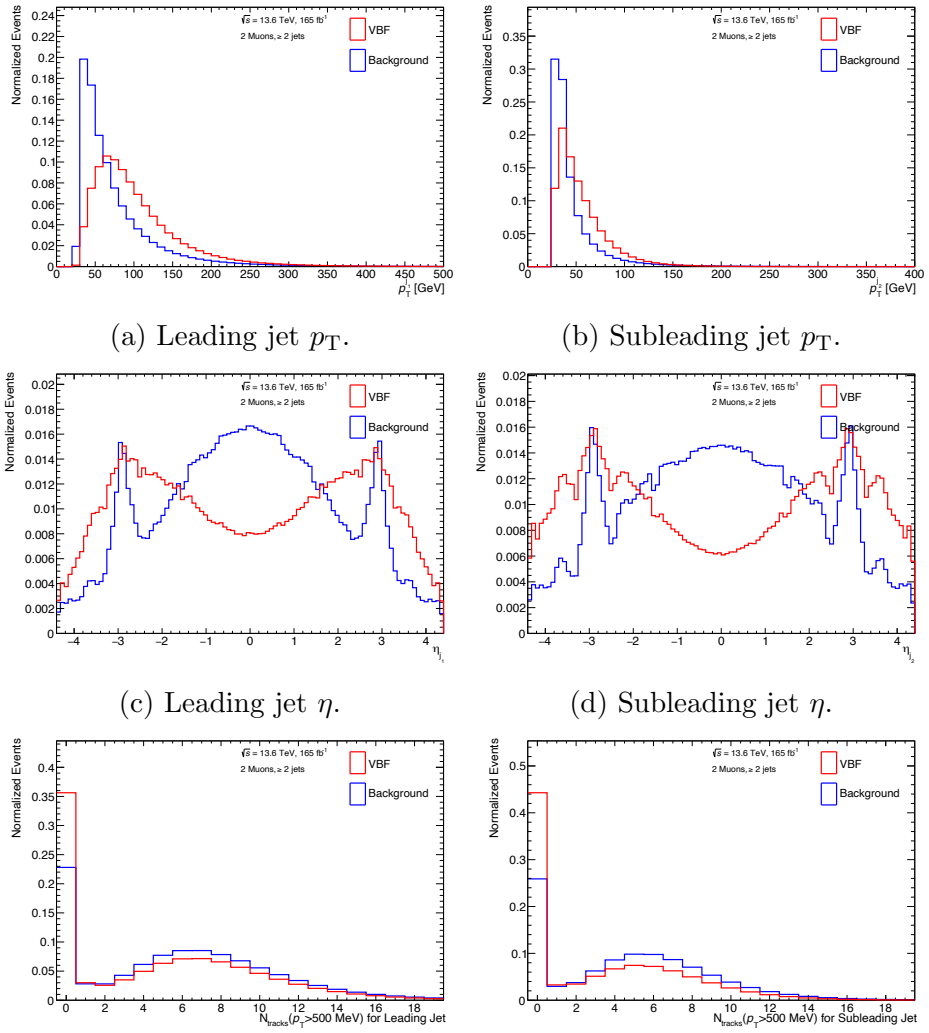


(d) Angular separation of the Higgs boson muons.



(e) Missing transverse momentum.

Figure A.11: Distributions of features that are used to train the NN VBF classifier. VBF signal events have a broader $\Delta\zeta$ distribution. VBF signal events tend to have a lower $E_{\mathrm{T}}^{\mathrm{miss}}$ and a different azimuthal separation between the Higgs boson and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. The angular separation of the Higgs candidate muons is lower for signal events compared to background events.

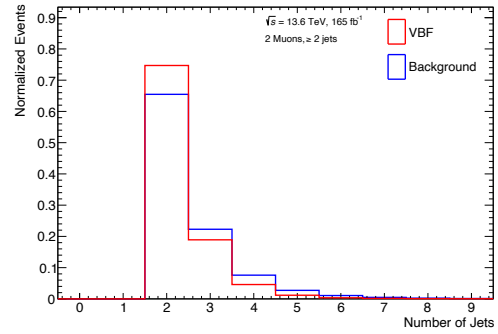


(a) Leading jet p_{T}. (b) Subleading jet p_{T}.

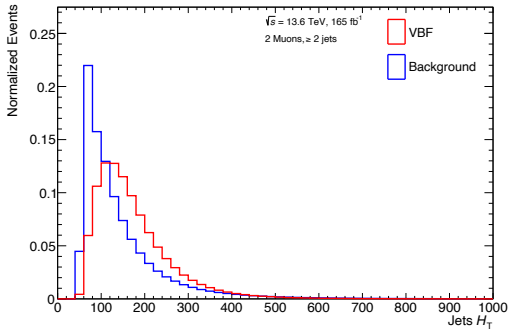
(c) Leading jet η. (d) Subleading jet η.

(e) Leading jet track multiplicity. (f) Subleading jet track multiplicity.

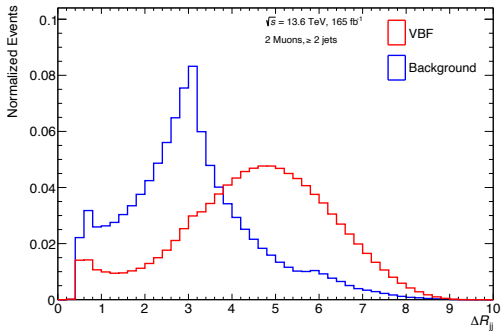
Figure A.12: Distributions of features obtained from jets that are used to train the NN VBF classifier. VBF signal events tend to have jets with a higher p_{T} that are more forward and have a lower track multiplicity compared to background events.



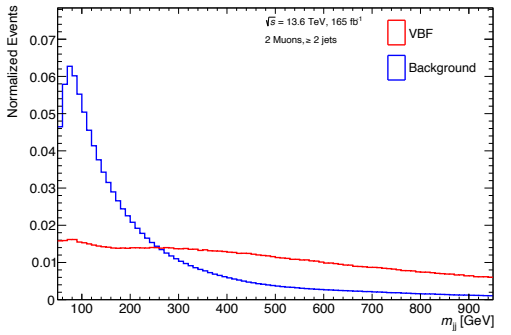
(a) Jet multiplicity.



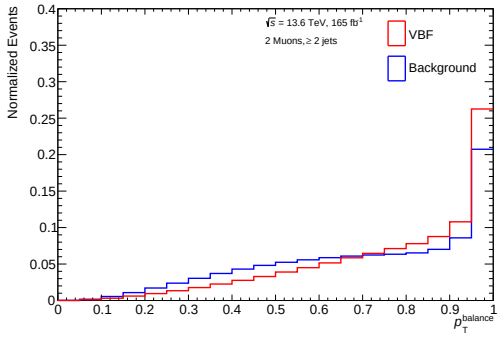
(b) Scalar sum of jet p_{T}.



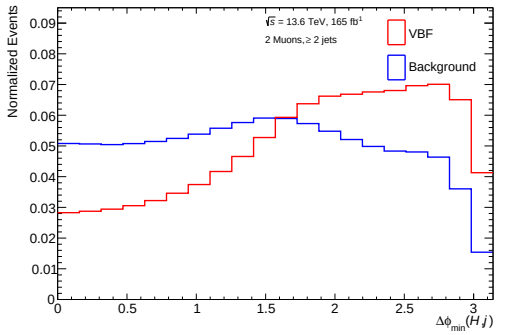
(c) The angular separation of leading jets.



(d) Dijet mass.



(e) p_{T} balance of the dimuon and dijet system.

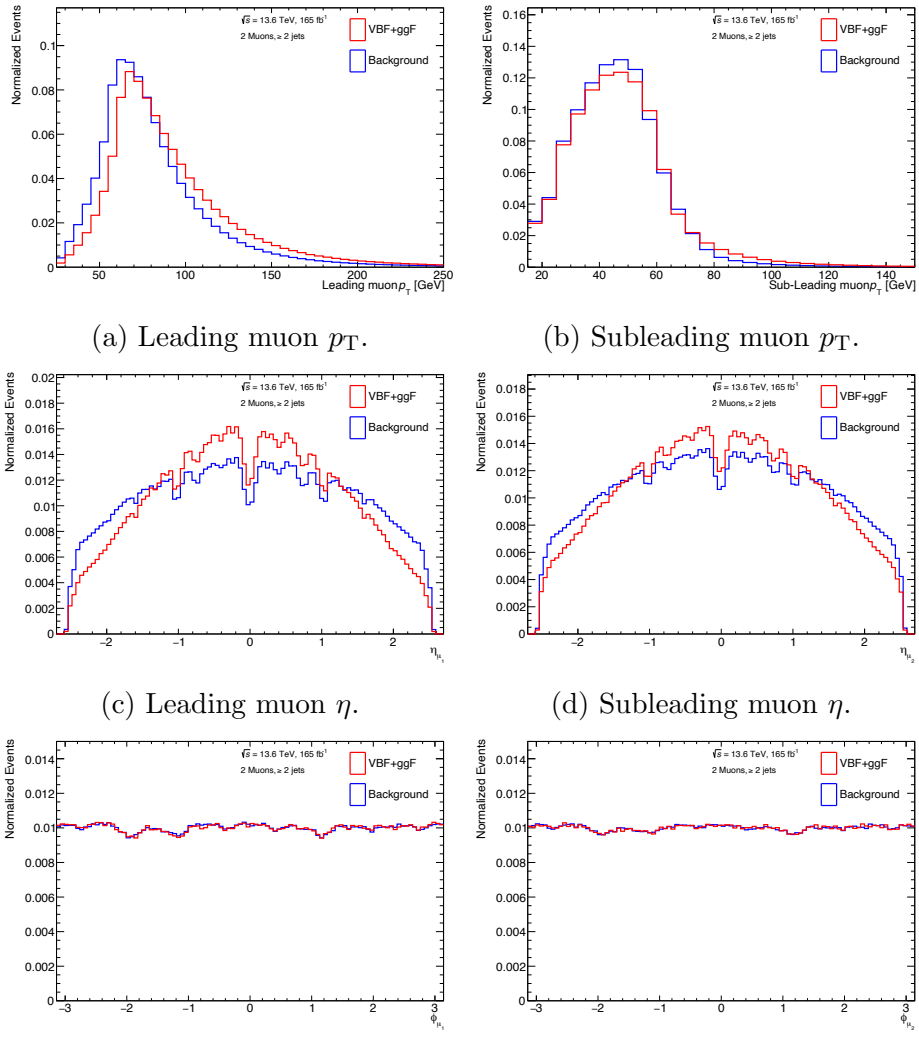


(f) Minimum $\Delta\phi$ between dimuon system and leading or subleading jet.

Figure A.13: Distributions of features obtained from jets that are used to train the NN VBF classifier. VBF signal events tend to have a lower jet multiplicity, but higher dijet mass and dijet angular separation than background events. Signal events also have a larger minimum azimuthal separation between the Higgs candidate and leading or subleading jet and a p_{T} balance closer to 1.

A.5 Higgs 2-Jet Training Variables

MC signal and background distributions for the 23 features used to train the NN Higgs 2-Jet classifier are shown in Figures A.14, A.15, A.16, and A.17.



(a) Leading muon p_{T}.

(b) Subleading muon p_{T}.

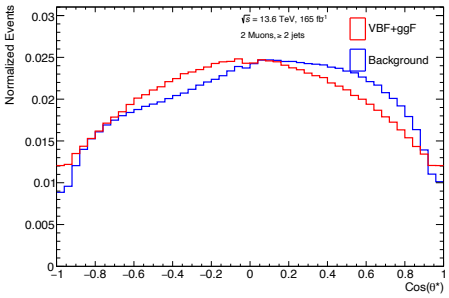
(c) Leading muon η.

(d) Subleading muon η.

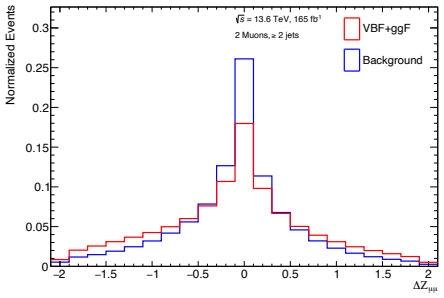
(e) Leading muon ϕ.

(f) Subleading muon ϕ.

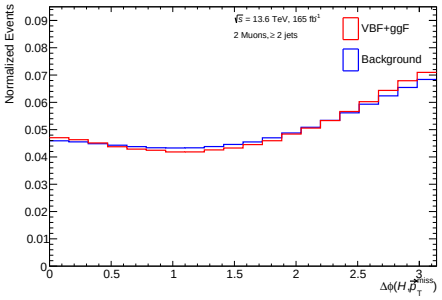
Figure A.14: Distributions of features obtained from muons that are used to train the NN Higgs 2-Jet classifier. The muons from VBF and ggF signal events tend to have a higher momentum than background processes.



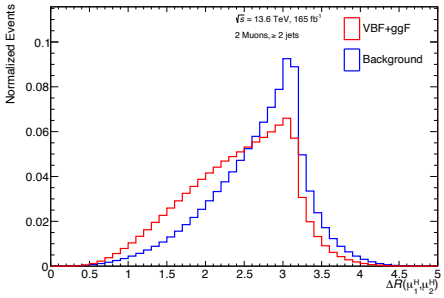
(a) Dimuon system $\cos\theta^*$.



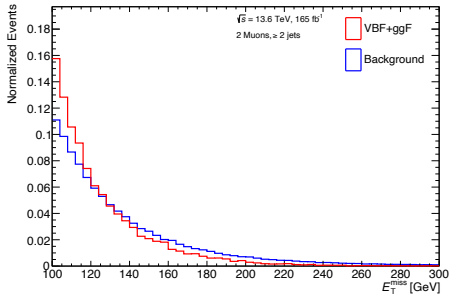
(b) $\Delta\zeta$ of the Higgs candidate muons.



(c) Azimuthal separation of the Higgs boson and $E_{\mathrm{T}}^{\mathrm{miss}}$.

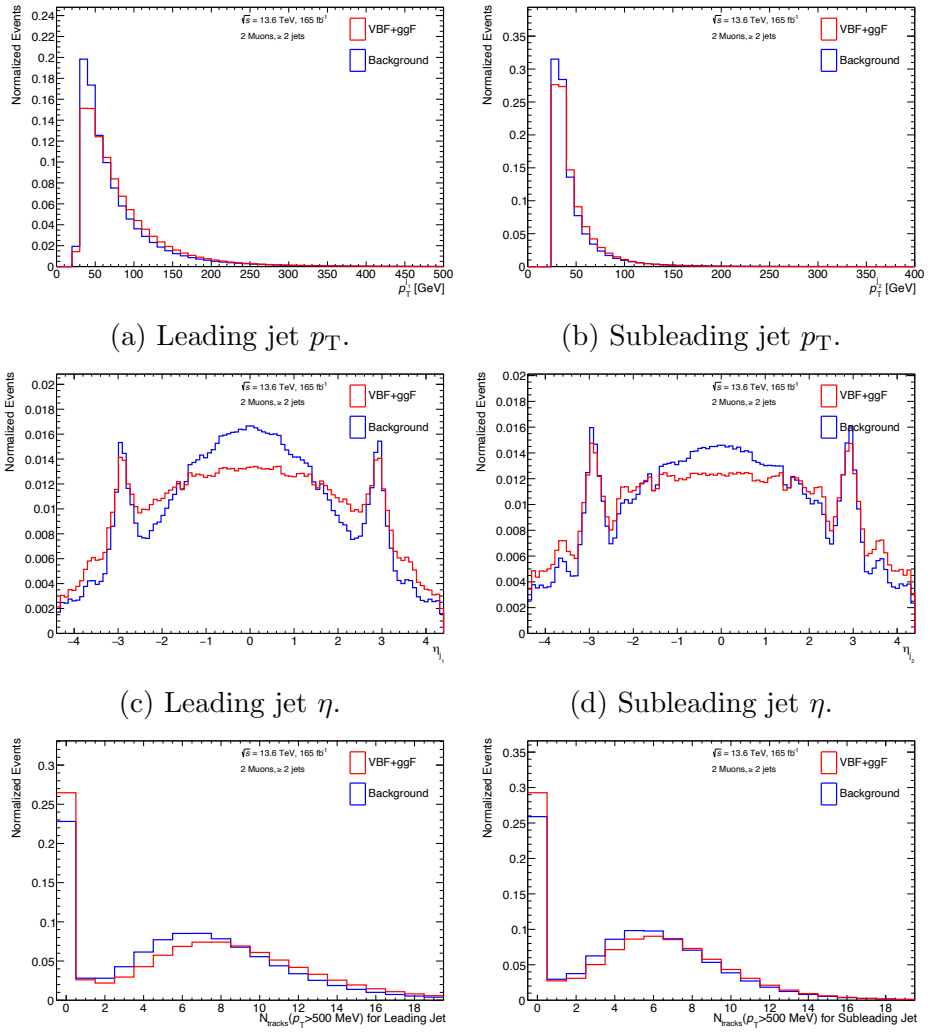


(d) Angular separation of the Higgs boson muons.



(e) Missing transverse momentum.

Figure A.15: Distributions of features that are used to train the NN Higgs 2-Jet classifier. VBF and ggF signal events have a broader $\Delta\zeta$ distribution. Signal events tend to have a lower $E_{\mathrm{T}}^{\mathrm{miss}}$ compared to background events. The angular separation of the Higgs candidate muons is lower in signal events compared to background events.

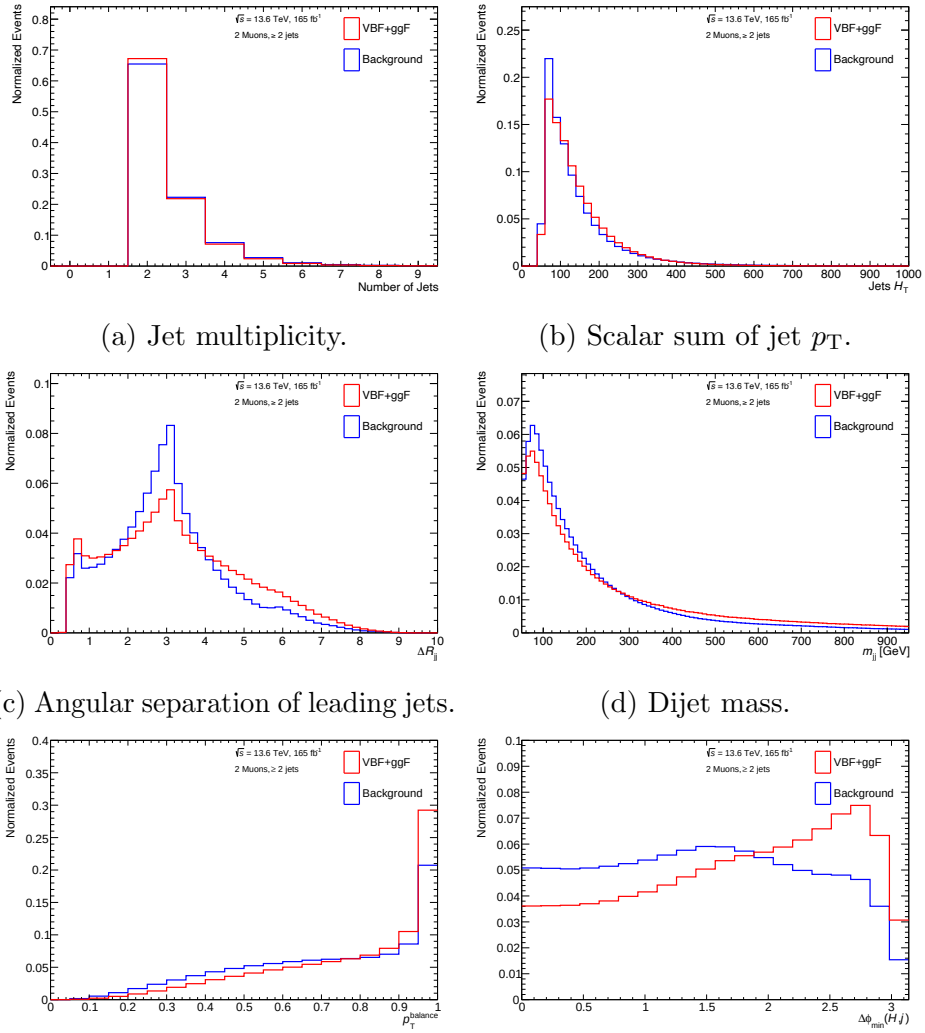


(a) Leading jet p_{T}. (b) Subleading jet p_{T}.

(c) Leading jet η. (d) Subleading jet η.

(e) Leading jet track multiplicity. (f) Subleading jet track multiplicity.

Figure A.16: Distributions of features obtained from jets that are used to train the NN Higgs 2-Jet classifier. Signal events tend to have jets with a higher p_{T} that are more forward and have a lower track multiplicity compared to background events.



(a) Jet multiplicity.

(b) Scalar sum of jet p_{T}.

(c) Angular separation of leading jets.

(d) Dijet mass.

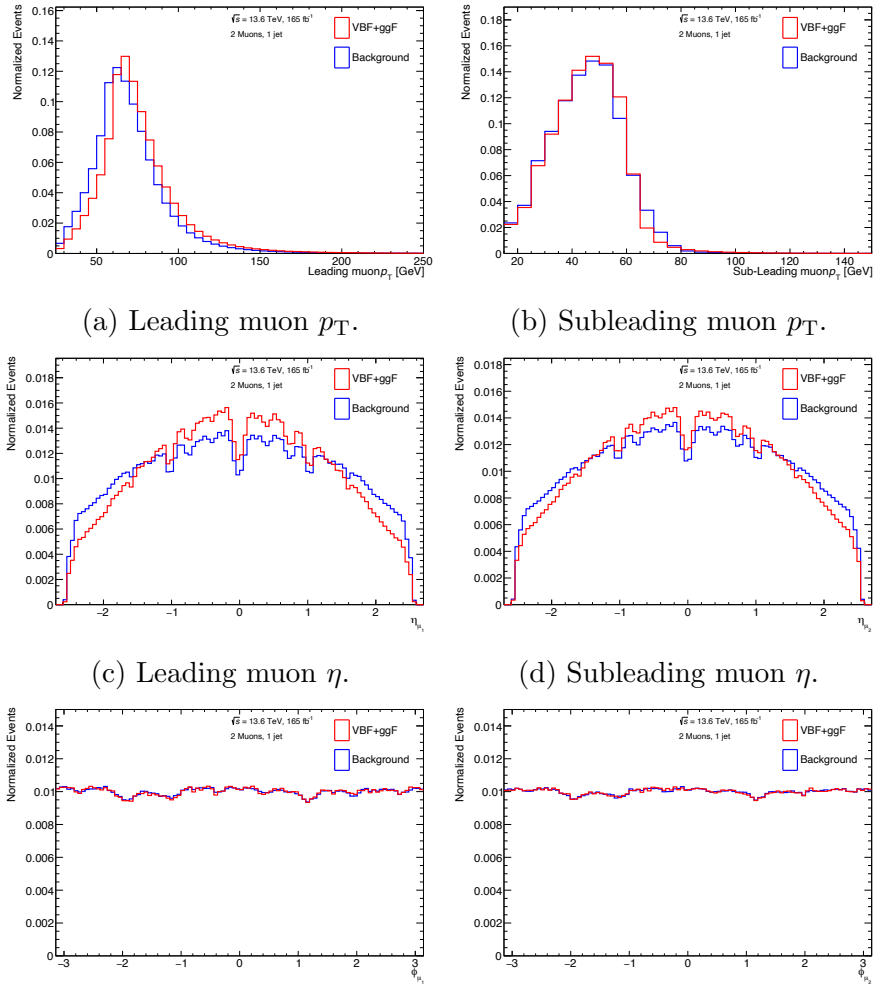
(e) p_{T} balance of the dimuon and dijet system.

(f) Minimum $\Delta\phi$ between dimuon system and leading or subleading jet.

Figure A.17: Distributions of features obtained from jets that are used to train the NN Higgs 2-Jet classifier. VBF and ggF signal events tend to have a slightly higher scalar sum of jet p_{T} and dijet mass. Signal events also have a larger minimum azimuthal separation between the Higgs candidate and leading or subleading jet and a p_{T} balance closer to 1 compared to background events.

A.6 Higgs 1-Jet Training Variables

MC signal and background distributions for features for the 16 used to train the NN Higgs 1-Jet classifier are shown in Figures A.18, A.19, and A.20.



(a) Leading muon p_{T}.

(b) Subleading muon p_{T}.

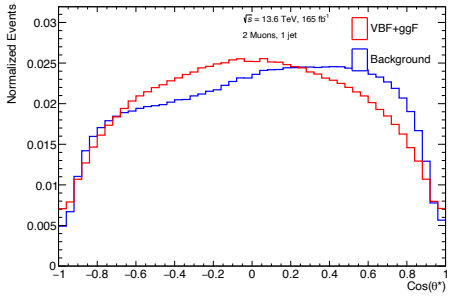
(c) Leading muon η.

(d) Subleading muon η.

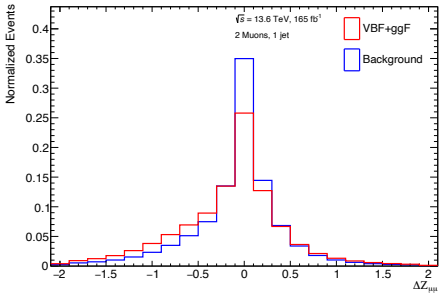
(e) Leading muon ϕ.

(f) Subleading muon ϕ.

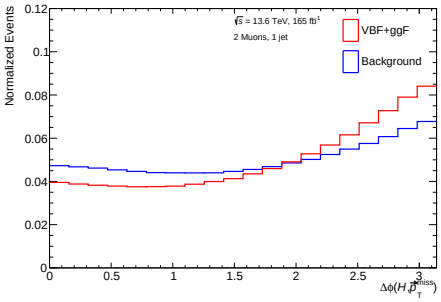
Figure A.18: Distributions of features obtained from muons that are used to train the NN Higgs 1-Jet classifier. The leading muons from VBF and ggF signal events tend to have a higher momentum than background processes and the muons tend to be more central.



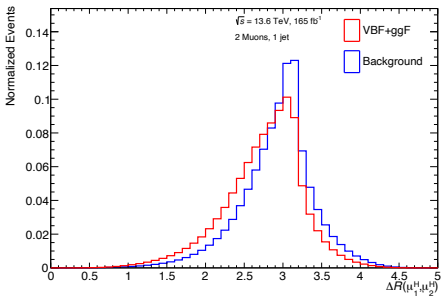
(a) Dimuon system $\cos\theta^*$.



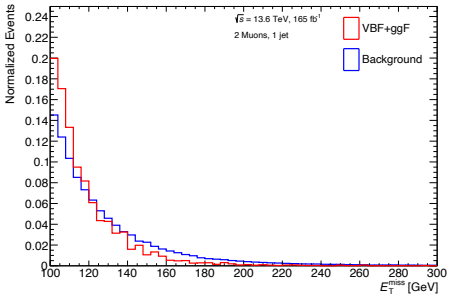
(b) $\Delta\zeta$ of the Higgs candidate muons.



(c) Azimuthal separation of the Higgs boson and $E_{\mathrm{T}}^{\mathrm{miss}}$.

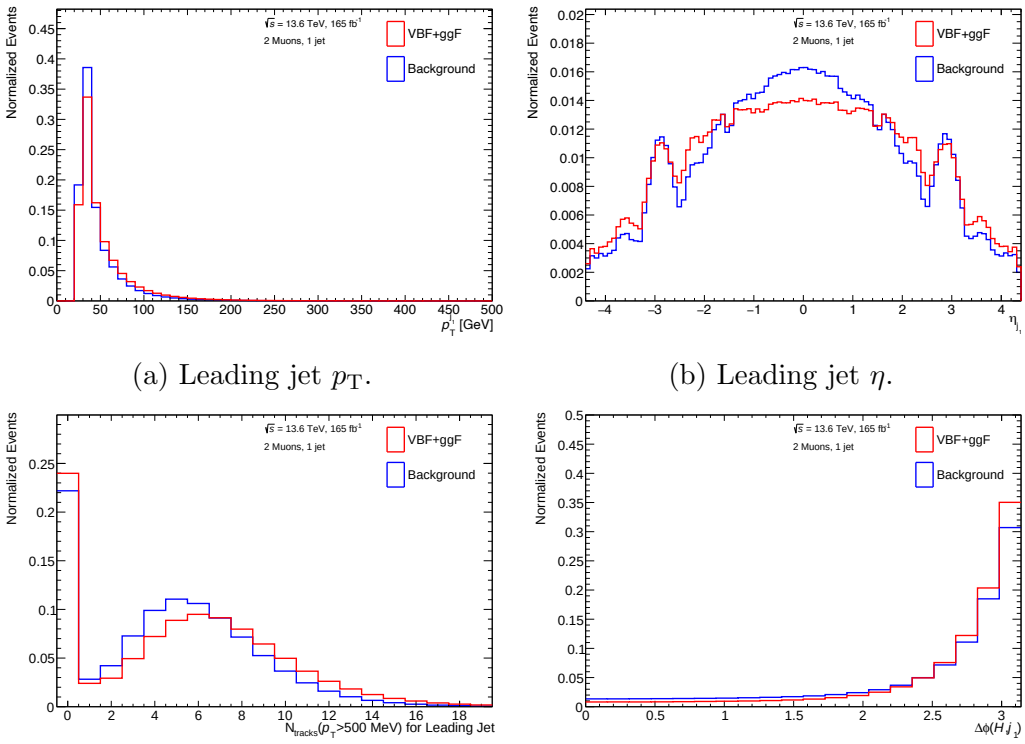


(d) Angular separation of the Higgs boson muons.



(e) Missing transverse momentum.

Figure A.19: Distributions of features that are used to train the NN Higgs 1-Jet classifier. VBF and ggF signal events have a broader $\Delta\zeta$ distribution. Signal events tend to have a lower $E_{\mathrm{T}}^{\mathrm{miss}}$ and larger azimuthal separation between the Higgs boson and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. The angular separation of the Higgs candidate muons is lower in signal events compared to background events.



(a) Leading jet p_{T}.

(b) Leading jet η.

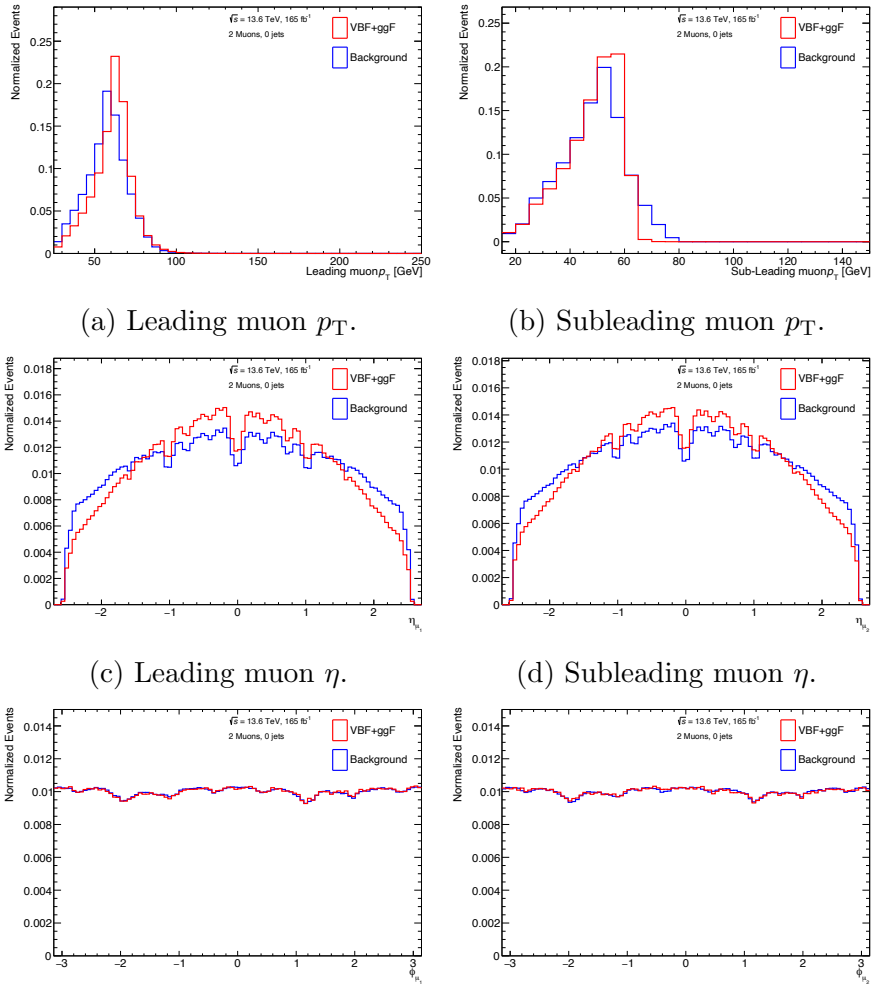
(c) Leading jet track multiplicity.

(d) $\Delta\phi$ of the Higgs and leading jet.

Figure A.20: Distributions of features obtained from jets that are used to train the NN Higgs 1-Jet classifier. VBF and ggF signal events tend to have jets with a higher p_{T} that are more forward and have a lower track multiplicity compared to background events. Signal events also have a larger azimuthal separation between the Higgs candidate and leading jet

A.7 Higgs 0-Jet Training Variables

MC signal and background distributions for the 11 features used to train the NN Higgs 0-Jet classifier are shown in Figures A.21 and A.22



(a) Leading muon p_{T}.

(b) Subleading muon p_{T}.

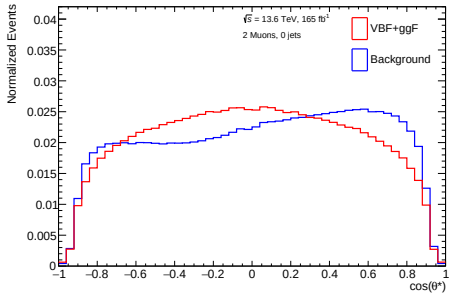
(c) Leading muon η.

(d) Subleading muon η.

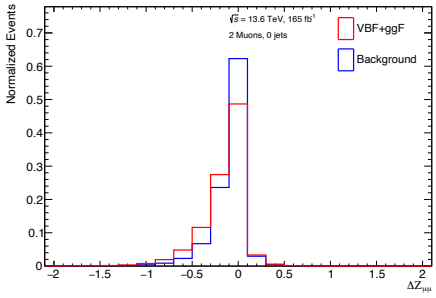
(e) Leading muon ϕ.

(f) Subleading muon ϕ.

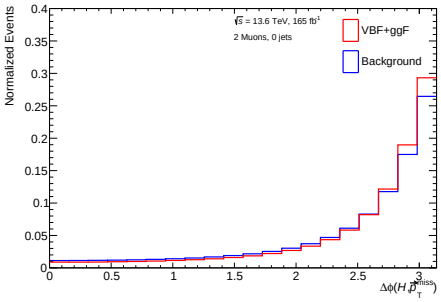
Figure A.21: Distributions of features obtained from muons that are used to train the NN Higgs 0-Jet classifier. The leading muons from VBF and ggF signal events tend to have a higher momentum than background processes and the muons tend to be more central.



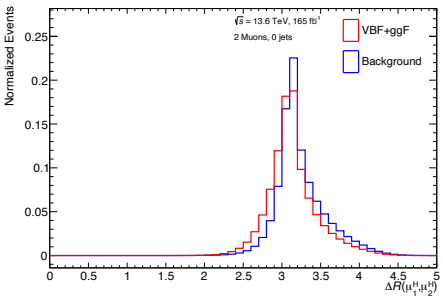
(a) Dimuon system $\cos\theta^*$.



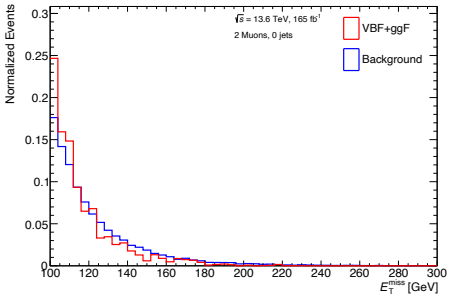
(b) $\Delta\zeta$ of the Higgs candidate muons.



(c) Azimuthal separation of the Higgs boson and $E_{\mathrm{T}}^{\mathrm{miss}}$.



(d) Angular separation of the Higgs boson muons.



(e) Missing transverse momentum.

Figure A.22: Distributions of features that are used to train the NN Higgs 0-Jet classifier. VBF and ggF signal events have a broader $\Delta\zeta$ distribution. Signal events tend to have a lower $E_{\mathrm{T}}^{\mathrm{miss}}$ and larger azimuthal separation between the Higgs boson and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. The angular separation of the Higgs candidate muons is lower in signal events compared to background events.