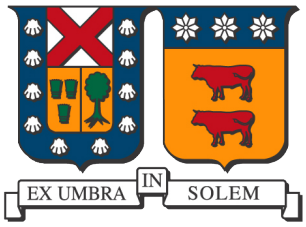
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE FÍSICA
VALPARAÍSO – CHILE



SEARCH FOR LONG-RANGE STRUCTURES IN TWO-PARTICLE CORRELATIONS IN THE EG2 EXPERIMENT, AND FOR A HEAVY SCALAR RESONANCE IN THE BOOSTED $HH \to b\bar{b}\tau^+\tau^-$ DECAY CHANNEL IN THE ATLAS EXPERIMENT

LUCIANO GABRIEL ARELLANO BERT

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS, MENCIÓN FÍSICA

Profesor Guía :	Dr. William Brooks
Profesor Co-Guía :	Dr. Edson Carquín
Co-referente interno :	Dr. Hayk Hakobyan
Co-referente externo :	Dr. Miguel Arratia

Enero – 2021

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☐ Memoria o trabajo de título ☒ Tesis de Postgrado

Título del trabajo: Search for long-range structures in two-particle correlations in the EG2 experiment, and for a heavy scalar resonance in the boosted hh->bbtautau decay channel in the atlas experiment

Nombre del candidato(a): Luciano Gabriel Arellano Bert

Carrera / Grado: Magíster en física, mención física

Campus: Casa Central **Departamento:** Departamento de física

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, William King Brooks, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por **(marcar una opción)**:

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 11/11/2026 **Firma:** William King Brooks

Estudiante o Candidato(a):

Fecha: 10/11/2026 **Firma:**

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Agradecimientos

El presente trabajo fue posible en parte gracias al financiamiento otorgado por la Dirección de Postgrado y Programas a través de la beca de arancel durante la duración de mi Magíster y la beca financiera en el periodo entre Septiembre 2018 y Febrero 2019. El trabajo en ATLAS fue financiado además a través del proyecto FONDECYT 1190886 "Search for resonant and non-resonant higgs pair production at the LHC with the ATLAS experiment".

Abstract

This thesis is divided in two parts. In the first part, a search for long-range structures in the rapidity-azimuth space is performed using two-particle correlations for the EG2 run data of the CLAS experiment. A far-side long-range structure typically associated to back-to-back pairs is observed for low Z_h bins of the associated particle, however no near-side "ridge-like" structure such as those observed in p–p or p–A collisions is observed. The one-dimensional projection of the correlations in azimuth is fitted using a third order Fourier series, and the second Fourier component v_2 is found to have no dependence on the interaction hardness variable Q^2.

In the second part, the ATLAS search for the decay of a heavy scalar resonance into a Higgs boson pair, and their subsequent decay in the boosted $b\bar{b}\tau^+\tau^-$ channel is presented. The ATLAS analysis and di-jet reconstruction method are discussed in detail. Then, an improvement of the signal region sensitivity using a two-dimensional mass window is presented and shown to lower the expected limits of the cross section across the entire search range 1–3 TeV. Finally, an alternative data-driven background estimation method using two-dimensional sidebands is presented, as an adaptation of the analysis done for the *bbbb* decay channel. Despite a good modeling of the entries yield in the signal region, the kinematic distribution of the estimated background is poor in some regions, and possible improvements to the method are discussed.

Contents

1 Introduction **5**

1.1 The Standard Model . . . 5

1.1.1 A brief history of particle physics . . . 5

1.1.2 Gauge bosons as mediators . . . 7

1.1.3 The Higgs field and the Higgs boson . . . 8

1.2 QCD and the internal structure of the proton . . . 11

1.2.1 The parton model and the internal structure of the proton . . . 11

1.2.2 Electron scattering . . . 11

1.2.3 Hadronization . . . 12

1.2.4 Medium effects and collective behavior . . . 13

1.3 Double Higgs production at LHC . . . 14

1.3.1 Combined searches in ATLAS . . . 15

1.3.2 Jet physics . . . 16

2 CLAS **20**

2.1 CLAS detector at Jefferson Lab . . . 20

2.1.1 The CLAS detector . . . 20

2.1.2 The EG2 experiment and Double Target System . . . 21

2.2 EG2 data analysis . . . 23

2.2.1 Compact File Format . . . 23

2.2.2 Event and particle pre-selection . . . 23

2.2.3 Virtual photon frame of reference . . . 24

2.3 Two particle correlations . . . 26

2.3.1 Event selection and analysis . . . 27

2.3.2 Nuclear medium dependence . . . 28

2.3.3 Q^2 dependence of v_2 . . . 28

3 ATLAS **36**

3.1 The ATLAS experiment at CERN . . . 36

3.1.1 LHC . . . 36

3.1.2 The ATLAS detector . . . 38

3.2 Boosted $X \to HH \to bb\tau\tau$ analysis . . . 41

3.2.1 Simulated events . . . 42

3.2.2 Object reconstruction and identification . . . 42

3.2.3 Event selection and data driven approach to background estimation . . . 44

3.3 Data driven estimation of QCD background using sidebands . . . 46

3.3.1 2D mass window cut . . . 46

3.3.2 Sideband and validation regions . . . 52

1 Introduction

This thesis will present the results of the research done for the CLAS and ATLAS experiments. It is divided in three parts. In this first part, we will present an introduction to the field of particle physics, introducing the relevant theoretical aspects needed to understand the experiments and results we will describe, and put emphasis on the physics constructs and observables that will be used in the analyses. In the second part, we will present the work related to the CLAS experiment, the novel search for a ridge structure and other long scale effects using electron nuclei collisions at relatively low energy, using two-particle correlations. In the third part, we will present the work related to the ATLAS experiment. We will describe the search for a heavy resonance into a double-Higgs boson pair in the decay channel $X \to HH \to b\bar{b}\tau^+\tau^-$. We will then present an optimization to the signal region using a two-dimensional mass window, and the estimation of multi-jet background using a data driven two-dimensional sideband method.

This first part is itself divided in three sections. First, in section 1.1 we will discuss some aspects of the Standard Model of particle physics. We will start with a brief historical introduction, then introduce the key concepts of gauge theories and spontaneous symmetry breaking from which the Standard Model is built. In the second section 1.2 we will focus in the strong interaction and contextualize in terms of deep inelastic electron scattering experiments. We will describe the current theoretical ideas used to describe hadronization, and introduce the ideas of the medium effect on parton propagation, and the collective phenomenon known as "the ridge". In the third section, 1.3, we will discuss double Higgs production in-and-beyond the Standard Model, and present the current state of searches at the Large Hadron Collider. We will discuss possible new-physics candidates that might enhance the probability of pair production, and finally we will introduce the concept of jet physics.

1.1 The Standard Model

1.1.1 A brief history of particle physics

The history of particle physics can be said to begin with the discovery of the electron in 1897 by J. J. Thomson, who in his experimental work with cathode rays showed that they were composed of electrified particles[1]. This happened in the midst of the development of atomic theory, which is inspired by the idea that all matter in the universe can be divided only up to a certain point, and the atoms (from Greek, meaning indivisible) of the different elements of the periodic table combine to create the wealth of observed diversity in the universe. This culminated in the Rutherford-Bohr model of the atom in 1913, which in addition to electrons includes protons as fundamental particles[2].

The fact that atoms had an internal structure meant they were not truly indivisible. Furthermore the electron is 1800 times smaller than the smallest atom, Hydrogen, so the term *elementary particle* was eventually coined to represent truly indivisible particles. Starting around 1930 new supposedly fundamental particles were being constantly discovered in cosmic rays, reactors, and especially with the advent of particle accelerators[3]. By the decade of 1950 the variety of known particles was so large that it was becoming

obvious that these could not all be fundamental, but rather the combination of a smaller set of truly fundamental particles, which could explain this ever increasing *particle zoo*.

The quark model proposed independently by Murray Gell-Mann[4] and George Zweig (who called them "aces")[5] posits that the hundreds of observed mesons and baryons stemmed from only 3 fundamental quarks (and their corresponding antiparticles), *up*, *down* and *strange*. Today we also know of the existence of three more, *charm*, *bottom* and *top* and we call all of these varieties *flavors* of quarks. Quarks also possess a quantum number referred to as color; single quarks cannot be observed due to *color confinement*, only bound states of a quark-antiquark pair (known as *mesons*) or three quarks (or three antiquarks, known as *baryons*) of different color may be observed as real particles.

In addition to quarks, there exists another set of particles called *leptons*, which includes the old *electron* and the more recently discovered *muon* and *tau* particles; these three all have electric charge. Three additional neutral leptons called neutrinos (one for each *flavor*, i.e. electron neutrino, muon neutrino and tau neutrino) complete the known fundamental *fermions*, all particles of spin 1/2, which also include quarks. A diagram with all fundamental particles in the Standard Model along with their classifications can be found in 1.1.

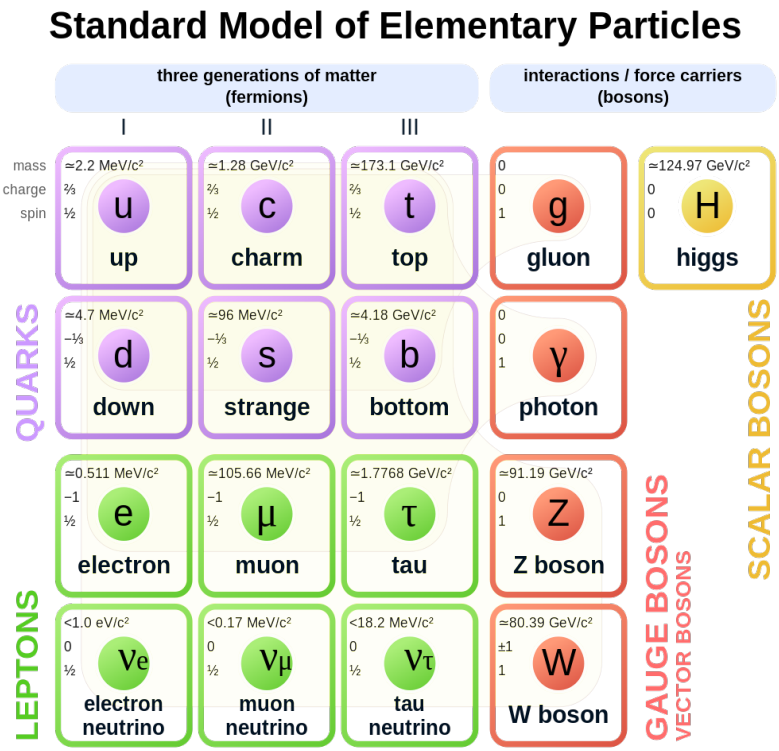


Figure 1.1: The fundamental particles composing the Standard Model. A corresponding antiparticle exists for all fermions

In parallel to the discovery of these particles, different quantum field theories were constructed to explain their interactions in the context of *fundamental interactions* or fundamental forces, of which four are known[1]. With time, the most successful quantum field theories proved to be *gauge* theories, in which so called gauge *bosons* (particles of

[1] One of these four fundamental forces, gravitation, remains unexplained in terms of renormalizable quantum field theories and thus is not included in the Standard Model.

integer spin) act as mediators of the interaction, their existence stemming due to an internal symmetry of the system, and is formally described by group theory. Quantum electrodynamics (QED) explains the electromagnetic interactions concerning electrically charged particles, which would be mediated by the *photon*. Quantum Chromodynamics (QCD) explains the strong nuclear interaction among quarks, which would be mediated by *gluons*. Fermi Theory first explained the beta decay process, and eventually served as the basis for the gauge-based Electroweak Theory, which explains all flavored processes due to the weak interaction and also unifies it with electromagnetism, by proposing that the $W^{\pm}$ and Z bosons along with the photon are the "physical" mixes of unobservable gauge fields. The W [6] and Z [7] bosons were discovered in 1983 at CERN.

The final major theoretical breakthrough was the Brout-Englert-Higgs mechanism [8, 9], by which a new scalar field known as the Higgs field allows the gauge bosons to acquire mass via spontaneous symmetry breaking. The Higgs field would also be responsible for the masses of the quarks and charged leptons via Yukawa couplings, and the existence of a new scalar particle known as the Higgs boson was predicted. A particle with such characteristics was found by the ATLAS [10] and CMS [11] experiments at LHC in 2012, thus validating the theory almost 50 years after it was conceived.

1.1.2 Gauge bosons as mediators

We can better understand the role of gauge invariance in the Standard Model with the study of Quantum Electrodynamics. The Dirac equation, the equation of motion which describes electrons,

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi = 0, \tag{1.1}$$

can be obtained from the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma_{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi. \tag{1.2}$$

This Lagrangian has an explicit global symmetry $\psi \to \psi' = \psi e^{i\alpha}$, with α a constant phase factor. If we want the more general *local* symmetry $\psi \to \psi' = \psi e^{i\alpha(x)}$, where α can now depend on spacetime coordinates, invariance can be achieved by introducing a new *gauge* field A_{μ} such that we replace the regular derivative for a covariant derivative $\partial_{\mu} \to D_{\mu} = \partial_{\mu} - ieA_{\mu}$ in which the new field transforms as $A_{\mu} \to A_{\mu} + \frac{1}{e}\partial_{\mu}\alpha$. The Lagrangian becomes

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}D_{\mu}\psi - m\bar{\psi}\psi \tag{1.3}$$

Let us now show the Lagrangian in terms of this new vector field A_{μ} we introduced, and give this field the kinetic term $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$. The resulting Lagrangian is known as the QED Lagrangian,

$$\mathcal{L}_{\text{QED}} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi + e\bar{\psi}\gamma^{\mu}\psi A_{\mu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.4}$$

A local symmetry has appeared in the Lagrangian but we are forced to introduce a new field, the gauge field A_{μ} which represents a boson, namely the photon. The photon interacts with the charged fermions ψ via the $\bar{\psi}A_{\mu}\psi$ term, which is why we say that this particle *mediates* the interaction. This procedure of introducing new fields in order to achieve local gauge invariance in the Lagrangian is known as *gauge theory* and has been shown to be incredibly fruitful, and even more importantly, was shown to lead to renormalizable theories by t'Hooft [12]. A more detailed overview of gauge theories in particle physics and of the following discussion on the Higgs field can be found in any modern particle physics textbook such as refs. [13, 14]

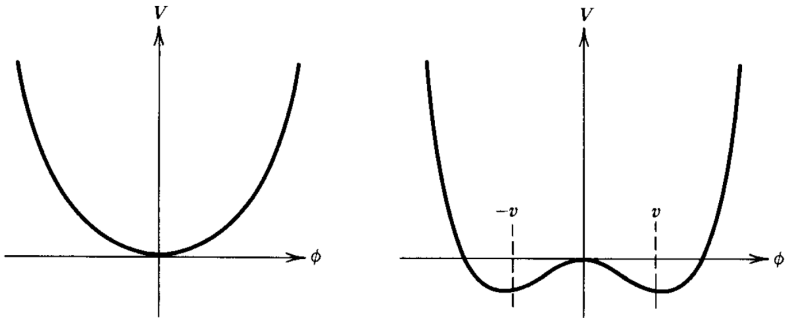


Figure 1.2: Potential $V(\phi)$ in the cases (a) $\mu^2 > 0$ and $\lambda > 0$, (b) $\mu^2 < 0$ and $\lambda > 0$. In case (b), the vacuum expectation value is located at a distance v from the origin in either direction. Source: reference [13]

1.1.3 The Higgs field and the Higgs boson

One stumbling block while trying to construct a gauge theory for the weak force, was that a procedure such as the one used in section 1.1.2 can only produce massless bosons, and it was known at the time that the mediating particles of the weak force were massive, as evidenced by the limited range of the interaction. Furthermore, disregarding local gauge invariance and manually introducing a mass term in the Lagrangian introduces unrenormalizable divergences.

Spontaneous symmetry breaking

The breakthrough that allowed to solve this theoretical problem is based around a phenomenon called spontaneous symmetry breaking (SSB), which occurs when the Lagrangian and ground state do not share a symmetry. This can be best understood with an example: consider the Lagrangian for a one-dimensional field $\phi(x)$

$$\mathcal{L} = \frac{1}{2}(\partial^\mu \phi)(\partial_\mu \phi) - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4. \tag{1.5}$$

This Lagrangian has a $\phi \to -\phi$ symmetry, however the ground state of the field is dictated by the minima of its potential term

$$V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4.$$

This potential will have different behaviors depending on the signs of λ and μ^2, as shown in figure 1.2. The most interesting case is for $\mu^2 < 0$ and $\lambda > 0$, as in this case the potential has only an unstable equilibrium at $\phi(0) = 0$ and stable minima at the points $\pm v = \pm\sqrt{\mu^2/\lambda}$. This suggests that we introduce a field $\eta(x)$ properly centered around one of these points $\pm v$ in order to expand around a minimum. Choosing $+v$ we replace $\phi(x) = v + \eta(x)$, and the Lagrangian reads

$$\mathcal{L} = \frac{1}{2}(\partial^\mu \eta)(\partial_\mu \eta) - \lambda\eta^2 v^2 - \lambda\eta^3 v - \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda v^4.$$

The η^3 term doesn't respect $\eta \to -\eta$, the symmetry of the Lagrangian has been spontaneously broken. We also take note that the field η has aquired a mass and self-interaction terms.

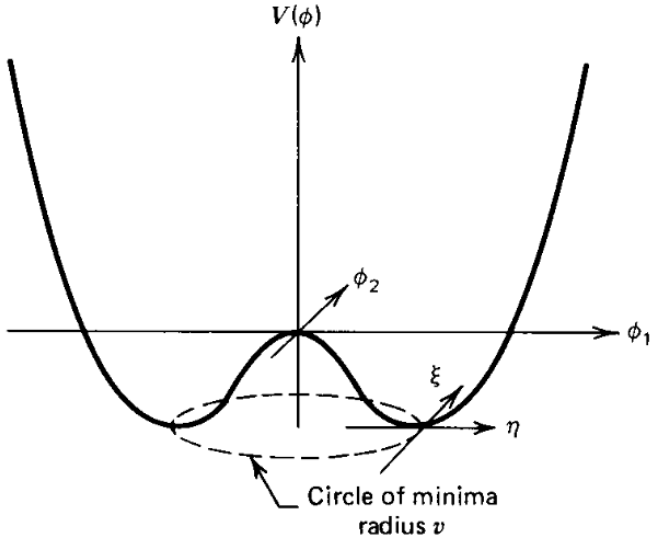


Figure 1.3: Potential $V(\phi)$ for the two dimensional scalar field. There is a continuous region of minima along which ξ runs unopposed introducing massless Nambu-Goldstone bosons in the theory. Source: reference [13]

The Brout-Englert-Higgs mechanism

The SSB of a local gauge theory is known as the Brout-Englert-Higgs mechanism. We consider a complex scalar field $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$ and the Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial^\mu \phi)^*(\partial_\mu \phi) - \frac{1}{2}\mu^2(\phi\phi^*) - \frac{1}{4}\lambda(\phi\phi^*)^2. \tag{1.6}$$

Let us introduce local gauge invariance as we did in section 1.1.2 by the $U(1)$ transformation $\phi \to \phi' = \phi e^{i\alpha(x)}$, and the introduction of a covariant derivative D_μ and gauge field A_μ

$$\partial'_\mu \to D_\mu = \partial_\mu - ieA_\mu \qquad\qquad A'_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha, \tag{1.7}$$

and adding a kinetic term for A_μ, the invariant Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}(D^\mu \phi)^*(D_\mu \phi) - \frac{1}{2}\mu^2(\phi\phi^*) - \frac{1}{4}\lambda(\phi\phi^*)^2 - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{1.8}$$

Now the potential has two dimensions, and again the case that allows for SSB is for $\mu^2 < 0$ and $\lambda > 0$, which appears in figure 1.3. This potential now has a continuum of minima, from which we choose $\phi_1 = v, \phi_2 = 0$. The expansion is now in the complex plane

$$\phi(x) = \frac{1}{\sqrt{2}}\left(v + \eta(x) + i\xi(x)\right),$$

with the field ξ running along the minima. Replacing in the Lagrangian we will obtain

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \xi)^2 + \frac{1}{2}(\partial \eta)^2 - \lambda v^2 \eta^2 + \frac{1}{2}e^2 v^2 A_\mu A^\mu - evA_\mu \partial\xi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{1.9}$$

The gauge boson A_μ is now massive, this is what we were originally looking for. However, we have also obtained a massless scalar field ξ representing a particle known as a Nambu-Goldstone boson. Such a particle will always appear when spontaneously breaking a

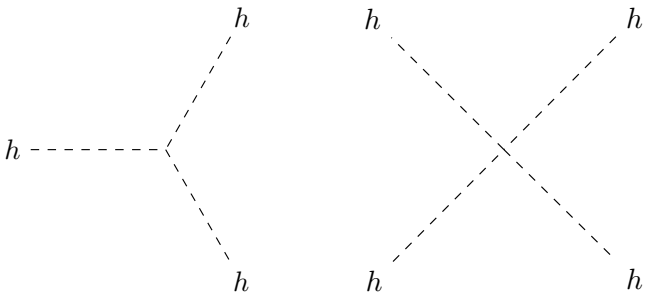


Figure 1.4: Trilinear (left) and quadrilinear (right) coupling diagrams of the Higgs boson as predicted by the Standard Model

continuous symmetry, a result known as Goldstone's theorem. This is a problem, as no such massless boson has ever been observed in an experiment. The solution that the Brout-Englert-Higgs mechanism delivers is introducing in the Lagrangian (1.6) the gauge choices

$$\phi' \to \frac{1}{\sqrt{2}}(v + h(x))e^{\theta(x)/v} \qquad A'_\mu \to A_\mu + \frac{1}{ev}\partial_\mu\theta. \tag{1.10}$$

With this the Lagrangian now becomes

$$\begin{aligned}\mathcal{L} =& \frac{1}{2}(\partial_\mu h)^2 - \lambda v^2 h^2 + \frac{1}{2}e^2 v^2 A_\mu^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4 \\ &+ \frac{1}{2}e^2 A_\mu^2 h^2 + v e^2 A_\mu^2 h - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\end{aligned}$$

This Lagrangian has everything we wished for and more. It gives the gauge boson A_μ its mass, and is devoid of massless Nambu-Goldstone bosons. It also predicts that the scalar field h, known as the Higgs field, will produce a boson of mass $m_h = \sqrt{2\lambda v^2}$, and a potential

$$V(h) = \frac{m_h^2}{2}h^2 + \lambda_3 v h^3 + \frac{\lambda_4}{4}h^4 \tag{1.11}$$

with $\lambda_3 = \lambda_4 = \lambda$. These trilinear and quadrilinear couplings give rise to the diagrams shown in figure 1.4.

The measurement of these couplings still remain out of experimental reach due to their low predicted cross section. In fact, the measurement of the Higgs trilinear coupling is one of the main motivations for the searches of the double Higgs decay we will discuss in the future section 1.3.

The Higgs field in the Standard Model

The Higgs field in the SM is obtained from the SSB of the $SU(2) \times U(1)$ symmetry of the electroweak interactions. In this case three massless Nambu-Goldstone bosons are generated and with the Higgs mechanism the degrees of freedom transform into the longitudinal polarizations of the three massive vector bosons while the photon remains massless.

In the SM the same Higgs doublet which generates the masses of the $W^\pm, Z$ bosons can also be used to give masses to the quarks and charged leptons via Yukawa couplings. However, the strength of these couplings are not predicted by the theory, only that they will be proportional to the fermion masses. It has indeed been found that the Higgs couplings

to third generation charged fermions are consistent with the SM predictions[15], but the couplings to first and second generation fermions remain experimentally challenging to observe[16, 17].

1.2 QCD and the internal structure of the proton

1.2.1 The parton model and the internal structure of the proton

While the quark model successfully applied group theory to describe the quantum numbers of the known hadrons and explain the particle zoo, it does not explain the dynamics of their constituents by itself. This was better explained with the parton model proposed by Feynman[18] which posits protons have an internal structure. Bjorken and Paschos[19], applied this to deep inelastic electron-proton scattering experiments (DIS)[2]. These developments paved the way for the formulation of the modern theory of QCD.

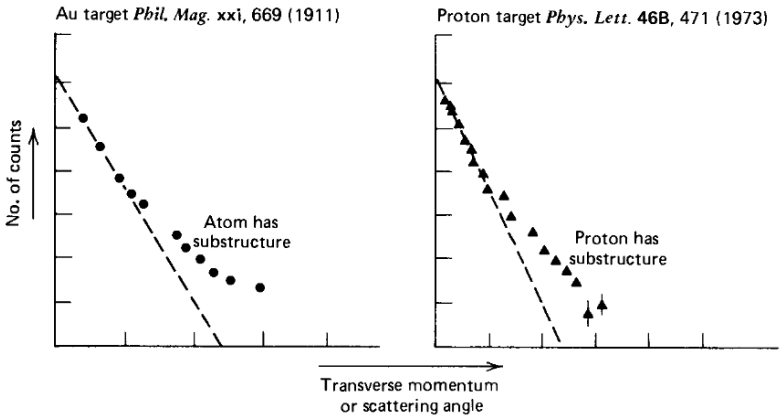


Figure 1.5: Comparison of experimental data showing in the left the scattering of alpha particles off atoms, proving atoms possess an internal structure (namely a nucleus) as shown by the famous Rutherford experiment. In the right picture, the scattering of electrons off protons is shown, suggesting they have an internal structure as well. Source: ref. [13] (original articles in the image itself)

1.2.2 Electron scattering

In an electron scattering experiment, a given electron beam of four-momentum k scatters off a nucleon target. By measuring the energy and momentum k' of the scattered electron, and assuming a one photon interaction, we can completely determine the characteristics of the virtual photon (denoted as γ^*) of four-momentum q, as illustrated in 1.6. After this we can ignore the electron and focus on the picture in (c) which contains all the information we need.

In the laboratory frame, with M the rest mass of the nucleon, we have

$$q = (E - E', -p_x, -p_y, p_z - p'_z) \qquad p = (M, 0, 0, 0) \tag{1.12}$$

[2]Deep inelastic scattering refers to the processes in which a highly energetic particle scatters of a hadron, revealing its internal structure. This can be achieved with any number of probe particles, such as electrons, muons or neutrinos. In this thesis we will focus on electron scattering mostly, and will use the terms DIS and electron scattering interchangeably

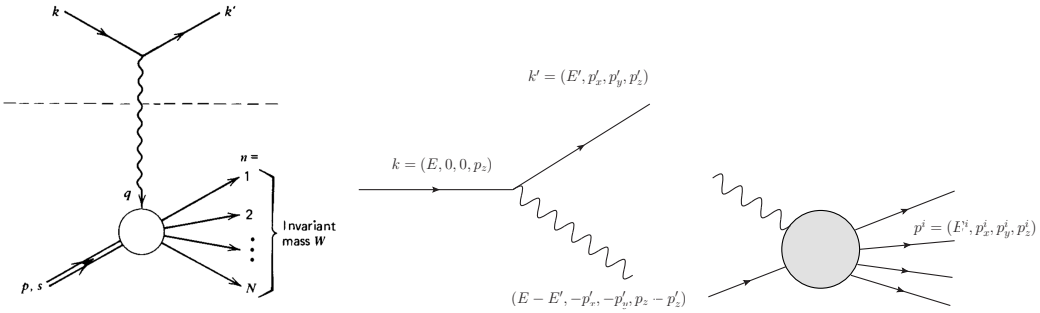


Figure 1.6: Left: full process. Center: electron side. Right: proton side

and so $p \cdot q = M(E - E')$. We denote the energy transfer from the electron to the nucleon by ν,

$$\nu \equiv E - E' = \frac{p \cdot q}{M}. \tag{1.13}$$

The total invariant mass of the γ^*-nucleon system is on the other hand denoted by $W^2 = (p + q)^2$, and so

$$W = \sqrt{M^2 + 2M\nu + q^2} = \sqrt{M^2 + 2M\nu - Q^2} \tag{1.14}$$

where we have used the more commonly used variable $Q^2 = -q^2$.

In the high momentum transfer regime $Q^2 \gg 1$ GeV^2 the underlying process can be described to first approximation as the scattering of a virtual photon with a free quark. The invariant mass W^2 is a measure of the hadronization of the system; in the elastic special case with $W^2 = M^2$ and $Q^2 = 2M\nu$ there are no new hadrons generated. The next threshold in $e^- p \to e^- X$ (assuming a struck proton) would be $X = p + \pi^0$ or $n + \pi^+$, corresponding to $W^2 \geq (M + m_\pi)^2$. We will be interested in the case of high hadronization, and therefore require $W > 2$ GeV.

Each of the hadrons generated in the process will carry some energy E_h in the laboratory frame. A commonly used variable is the fraction of the available energy carried by each hadron h,

$$Z_h = \frac{E_h}{\nu}. \tag{1.15}$$

As defined, Z_h should be a number between zero and one. Hadrons which carry most of the energy in the event ($Z_h > 0.5$) are of special interest as they come from a hard parton process.

1.2.3 Hadronization

We have mentioned only superficially the concept of hadronization. The idea is that in DIS, when a high enough energy electron scatters off a nucleon, it could interact directly with a valence quark through a virtual photon, resulting in the production of a varying amount of hadrons in order for the quarks to comply with color confinement, a process known as hadronization. If in our experiment we are able to measure only the scattered electron, we refer to this as *inclusive* scattering. If we are certain to measure all produced hadrons in addition to the electron, this is called *exclusive* scattering. In the case where we measure the electron and *some* hadrons, we refer to this as *semi-inclusive* (deep inelastic) scattering, or SIDIS.

A well-established model for the hadronization process is the Lund string model[20]. The model can be explained with an analogy in which the quarks which make up an hadron are joined by a massless string, which would represent the gluon field lines. As

the string stretches, the energy stored in it becomes so large that it is sufficient for the creation of a quark-antiquark pair. Then the string breaks, resulting in a larger number of hadrons, as shown in figure 1.7. This model was originally developed by members of the Department of Theoretical Physics at Lund University, and has been implemented in the PYTHIA[21] event generator, one of the most widely used Monte Carlo generators for pp collisions at high energies (10 GeV - 100 TeV).

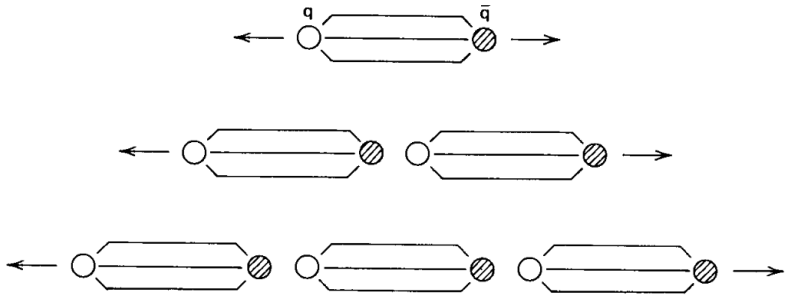


Figure 1.7: Hadronization as understood via string breaking. An initial highly energetic quark-antiquark pair on the top becomes three at the bottom as the string breaks and new quark-antiquark pairs are generated. The three lines in between particle pairs represent field lines. Source: ref. [13]

The hadronization of very high energy partons will result in a large particle shower formed in the direction of the original parton. These are called particle jets, and will be discussed in more detail in the future section 1.3.2.

1.2.4 Medium effects and collective behavior

So far we have discussed the scattering of electrons on the quark constituents of protons and neutrons by considering greatly simplified models in which they exist "in a vaccuum". However, another way to study hadronization is by introducing a medium with which partons might interact with, and study the effects of this medium on hadronization. For example, by scattering electrons off (heavy) nuclei ($e + A$) instead of a single nucleon ($e + p, n$) the "cold QCD matter" medium would be the nuclear target itself[22]. This is precisely one of the main goals of the EG2 experiment we will be exploring in this thesis, in which electrons are scattered off a range of targets of different nuclear size.

Other examples of medium are "hot QCD matter", which is produced in $A + A$ collisions, and the more extreme *quark-gluon plasma* (QGP)[23], a state of matter in which the quarks and gluons of QCD are no longer confined to the dimensions of the nucleon, but free to move in a volume with (extremely) high temperature or density. These conditions are believed to have existed naturally up to a few instants (10^{-12} to 10^{-6} seconds) after the big bang. Experimentally, circumstancial evidence for the creation of this new state was announced by CERN scientists in 2000[24] using Pb+Pb collisions, but its real discovery came in 2005 when four collaborations at the Berkeley National Laboratory Relativistic Heavy Ion Collider (BNL RHIC) announced the results of their first 5 years of measurements of Au+Au collisions[25–28].

In recent years there has been an enormous push to explain intriguing results obtained in proton-proton and proton-nucleus collisions at the Large Hadron Collider (LHC) and RHIC, as new long-range correlations have been observed in high multiplicity events. In particular we will study the phenomenon refered to often as "the ridge", a novel structure seen in two particle correlations extending over a large pseudorapidity range in the near

azimuthal side ($\Delta\phi \sim 0$), pictured in figure 1.8 [29, 30]. Such long range structures are typically interpreted in terms of collective behavior due to hydrodynamic flow in the strongly interacting medium produced in collisions [31], and characterized by the Fourier decomposition of the azimuthal particle distributions [32]. What is noteworthy is that, while this type of collective behavior was anticipated in heavy ion collisions, it was not for the smaller systems of proton-proton and proton-nucleus collisions. A natural follow up is to explore the existence of similar phenomena in even smaller (colder) systems such as electron-nucleus or electron-proton collisions, which motivates the search presented in this thesis.

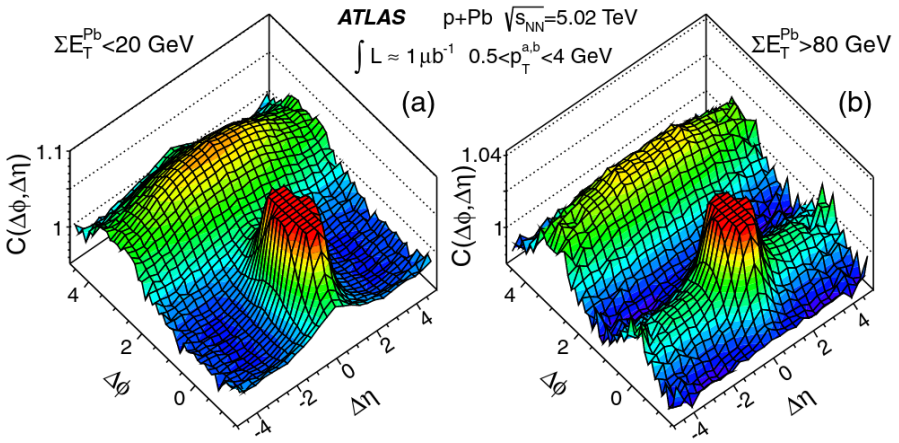


Figure 1.8: Two dimensional correlation plots in azimuthal angle and pseudorapidity difference for proton-lead collisions in the ATLAS experiment. (a) shows only a long-range structure in $\Delta\phi \sim \pi$ (far side) while (b) shows the $\Delta\phi \sim 0$ (near side) long-range in $\Delta\eta$ structure known as "the ridge". Figure from ref. [33]

1.3 Double Higgs production at LHC

The Large Hadron Collider has been a tremendous experimental success, as it already achieved its main goal of the discovery of the Higgs boson [10, 11]. With the existence of the Higgs established, the challenge now becomes the precision measurement of its properties and couplings, and the search for rarer processes involving this particle.

Double Higgs production is one such rare process. Its observation is pushing the limits of experimental measurements of the LHC and one of the main objectives of its near-future upgrade, the High Luminosity Large Hadron Collider (HL-LHC) [7]. It has also tremendous theoretical importance as it can uniquely provide information on the trilinear self-coupling of the Higgs λ_3 in equation (1.11) and the (b) diagram in 1.9 (the quadrilinear term is thought to be inaccessible for the current and near-future generation of experiments due to the rarity of the process and thus very large required luminosity).

The cross section of double Higgs production via gluon fusion, such as shown in 1.9, is $\sigma_{\text{SM}} = 31.05$ fb at $\sqrt{s} = 13$ TeV [34]. Gluon fusion accounts for more than 90% of the total production, with the subleading production channel, vector boson fusion, accounting for only 1.739 fb at $\sqrt{s} = 13$ TeV.

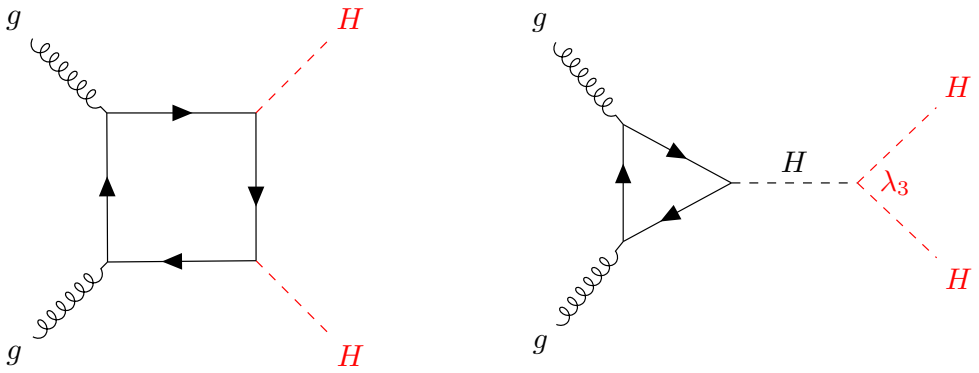


Figure 1.9: Higgs boson pair production at tree level in the Standard Model via gluon fusion. (a) production via either top or bottom quark loop (b) production via trilinear self coupling of the Higgs boson

1.3.1 Combined searches in ATLAS

The search for Higgs boson pairs is a very active area of research at ATLAS, with different groups working in parallel in the different decay channels. Searches have been carried out for gluon fusion production of HH events with a center of mass energy of $\sqrt{s}$ = 13 TeV in the $b\bar{b}b\bar{b}$ [35], $b\bar{b}W^+W^-$ [36], resolved $b\bar{b}\tau^+\tau^-$ [37], $W^+W^-W^+W^-$ [38], $b\bar{b}\gamma\gamma$ [39] and $W^+W^-\gamma\gamma$ [40] channels, along with a combination analysis, using all of the above searches [41]. Most of these analyses were carried out using only 2015-2016 data. The published analyses using the full Run 2 of the LHC (2015-2018 data) are the leptonic $b\bar{b}WW^*$ [42], the vector-boson-fusion production based $bbbb$ [43] and the boosted $b\bar{b}\tau\tau$ [44] in which a part of this thesis focuses on. The branching ratios can be seen in table 1.1.

Table 1.1: Branching ratios of HH decays [45]. The $HH \to b\bar{b}\tau^+\tau^-$ branching ratio is highlighted

	$H \to b\bar{b}$	$H \to W^+W^-$	$H \to \tau^+\tau^-$	$H \to ZZ$	$H \to \gamma\gamma$
$H \to b\bar{b}$	33%	25%	**7.4%**	3.1%	0.26%
$H \to W^+W^-$		4.6%	2.5%	1.2%	0.10%
$H \to \tau^+\tau^-$			0.39%	0.34%	0.029%
$H \to ZZ$				0.076%	0.013%
$H \to \gamma\gamma$					0.0005%

The results for Higgs pairs searches focus on obtraining the upper limits of pair production in terms of the SM prediction, and on the value of the λ_3 trilinear coupling in terms of its deviation from the SM value $\kappa_\lambda = \lambda_3/\lambda_3^{SM}$. The most stringent limits in both cases are given by the combination search and represent the observed (expected) limits at 95% confidence of 6.9 (10) times for the cross section and -5.0 $< \kappa_\lambda <$ 12.0 (-5.8 $< \kappa_\lambda <$ 12.0) [41]. It is expected that these values will be further constrained once results from the full Run 2 period are analysed and combined for the different channels.

Beyond the Standard Model resonances

Most of the publications cited in the previous subsection also present searches for BSM resonant decays of heavy particles into a Higgs pair. The two main candidates appearing

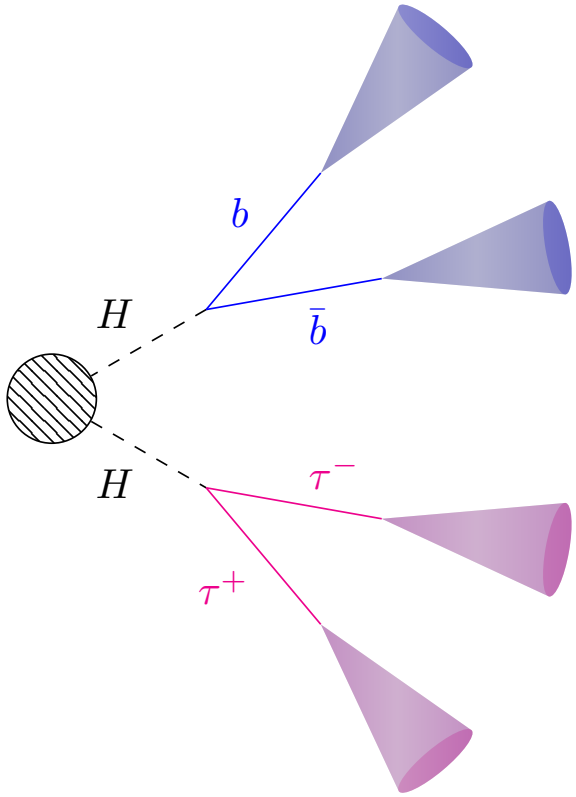


Figure 1.10: The $HH \to b\bar{b}\tau^+\tau^-$ decay channel. The further decay of b quarks (in blue) and τ leptons (in magenta) into hadrons will seed the particle jets (represented as cones) which will be ultimately reconstructed in the ATLAS detector. This picture represents the "resolved" regime, where jets occur without much overlap and standard reconstruction methods can be used

in these are spin-0 and spin-2 heavy resonances. The first type are usually based around a family of theories known as two Higgs-doublet models[46], in which extra Higgs bosons exist, including a charged pair ($h^{\pm}$) and a CP even scalar (H) heavier than the one found in 2012 (h), thus H might decay into a hh pair. The second kind is the decay of a spin-2 Kaluza-Klein graviton (G_{KK}), predicted by the Randall-Sundrum model[47]. In particular, in the *bulk* or *warped* versions of this 5-dimensional model[48, 49], the graviton couples with the weak scale (as opposed to the much weaker gravitational scale), making it accessible to accelerators in the TeV scale, the LHC in particular.

Resonant decay of a heavier BSM particle into Higgs boson pairs, whatever its origin, would lead to an enhancement of double Higgs production, which should be measurable as an increase in double Higgs events. A diagram of the decay of such a heavy resonance as generated from gluon fusion in pp collisions is shown in figure 1.11. Once again, the current best limits are set by the ATLAS combination search, ref. [41] for both a spin-0 and spin-2 resonance.

1.3.2 Jet physics

In collider experiments at high energies such as the LHC, detected particles originating from QCD decays are often clustered together in a narrow region rather than uniformly scattered across the detector surface, thus forming jets. The study of these particle jets

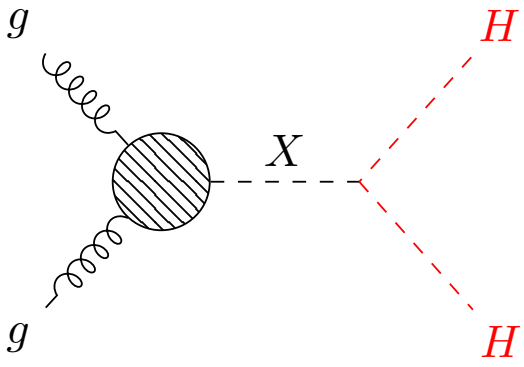


Figure 1.11: Beyond the Sandard Model Higgs pair production stemming from a particle X

not only simplifies the analysis by allowing us to study a few jets instead of the sometimes hundreds of particles composing them individually, but also gives us insight about the nature of the collision itself.

The formation of QCD jets is predominantly dictated by hadronization due to color confinement, as was described in section 1.2.3. In fact the narrow nature of jets is due to the collinear divergence of QCD, and so the existence of jets was an important factor in establishing QCD as the theory of strong interactions.

Formally a jet is defined by a jet algorithm (basically a selection of which of the detected particles belong to which jet) and a recombination scheme (which specifies how the kinematic properties of the jet are obtained from its constituent particles). In practice, most implementations use the E-scheme, in which the resulting jet mass, p_T, etc. are obtained from the sum of the constituent particles four-vectors.

Jet algorithms

The jet algorithm is the mechanism for deciding whether a given particle belongs to a jet or not, whatever this mechanism might be. In practice, the jet algorithm is defined using a set of parameters from which a selection based on measured quantities can be realized. A typical parameter is the *jet radius*, which is the maximum distance in the rapidity-azimuth $(y - \phi)$ plane for which two particles may be considered as part of the same jet.

There are two main types of algorithms: sequential recombination algorithms, and cone algorithms. Cone algorithms follow the idea that jets are due to the flow of energy in an event while sequential recombination algorithms try to trace back the parton branchings that gave rise to the observed particles by recombining two particles into one. The recombination is based on minimizing the distance between particles, as to replicate the QCD dynamics of the parton shower, which indicates that low distance processes are enhanced.

For example the *Generalised-k_t algorithm*, from which the recombination algorithms most used in hadronic collisions derive, is defined by the following steps, which are copied almost verbatim from ref. [50]:

1. Take the final state particles in the event as our initial list of objects.

2. From the list of objects, build two sets of distances: an *inter-particle distance*

$$d_{ij} = \min\left(p_{t,i}^{2q}, p_{t,j}^{2q}\right) \Delta R_{ij}^2, \tag{1.16}$$

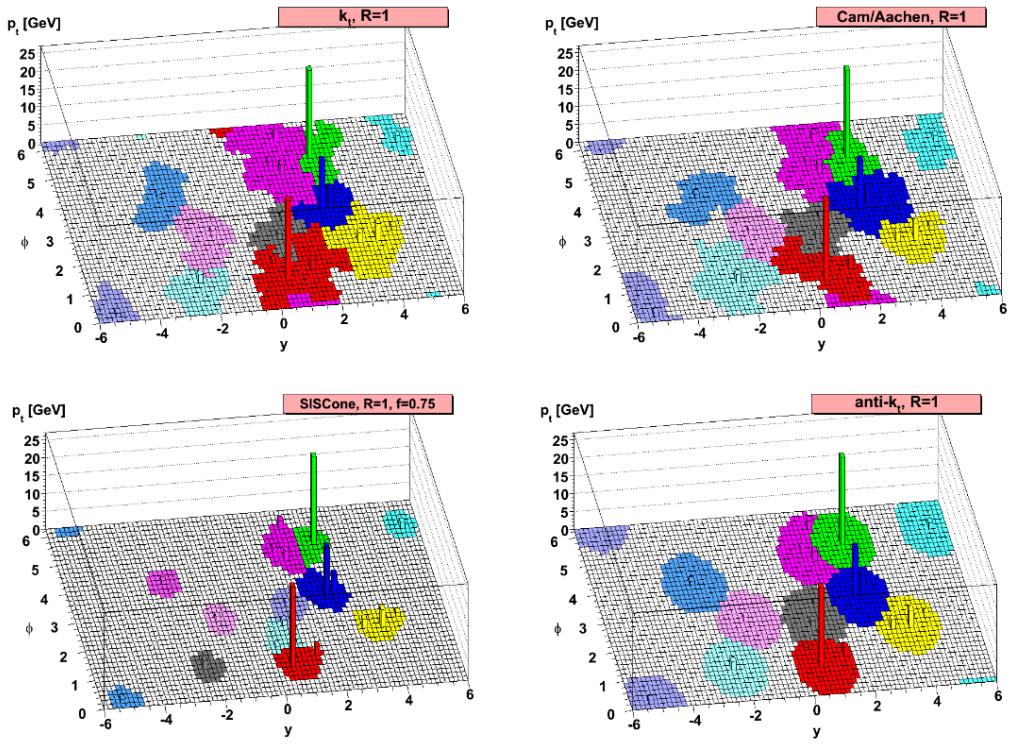


Figure 1.12: A comparison of the clustering of different jet algorithms. Source: ref. [52]

where the momentums are to the power of $2q$, a free parameter, and ΔR_{ij}^2 is the distance of particles i and j in the rapidity-azimuth angle plane

$$\Delta R_{12} = \sqrt{\Delta y_{12}^2 + \Delta \phi_{12}^2}, \tag{1.17}$$

and the *beam distance*

$$d_{iB} = p_{t,i}^{2q} R^2 \tag{1.18}$$

with R a free parameter usually called the *jet radius*.

3. Iteratively find the smallest distance among all the d_{ij} and d_{iB}
   - If the smallest distance is a d_{ij} then objects i and j are removed from the list and recombined into a new object k (using the recombination scheme) which is itself added to the list.
   - if the smallest is a d_{iB}, object i is called a *jet* and removed from the list.

   Go back to step 2 until all objects in the list have been exhausted.

Some famous examples of algorithms based on this scheme can be obtained by selecting different values for q in equation (1.16). These are the k_t algorithm [51] for which $q = 1$ and the Cambridge/Aachen algorithm with $q = 0$. The default algorithm for LHC however is the *anti-k_t* algorithm [52], for which $q = -1$. This choice will favor the clustering of hard particles first and produces a circular shape for hard jets. An illustration of the different kinds of results given by different jet algorithms can be seen in figure 1.12.

Dijets and boosted regime

The decay of a high mass particle X (1 to 3 TeV) into a back-to-back Higgs boson pair would result in highly energetic and high p_T Higgs bosons, the decays of which will result in the so called *boosted* regime, for which $p_T \gg m$. In this regime the four jets produced by the decay products $b\bar{b}$ and $\tau^+\tau^-$ would be highly collimated and overlapping, and it would not be possible to resolve them into four separate objects using the standard jet size of $R = 0.4$. The strategy we will discuss in a future section is to seed large-R jets which will contain two smaller subjets within. We will refer to these objects as *dijets*, the two kinds being the bb–jet and di–τ jet. A representation of decay into the boosted regime is presented in figure 1.13, which might be contrasted to the resolved case 1.10.

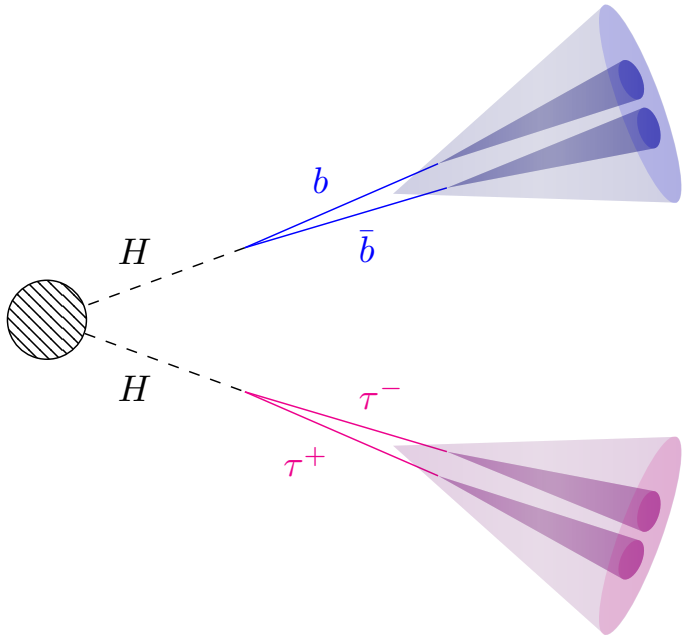


Figure 1.13: Boosted $HH \to b\bar{b}\tau^+\tau^-$. The decay products of the Higgs are highly collimated, and a careful selection of jet sizes must be made. The strategy is reconstructing large-R jets and small subjets within them. The decay of a heavy BSM resonance might result in its decay products being in the boosted regime ($p_t \gg m$)

2 CLAS

2.1 CLAS detector at Jefferson Lab

The Thomas Jefferson National Accelerator Facility (Jefferson Lab) is one of seventeen United States national research facilities overseen by the United States Department of Energy (DOE). Located in Newport News, Virginia, it specializes in studying the structure of nuclear matter, to which effect it hosts the Continuous Electron Beam Accelerator Facility (CEBAF), an accelerator which for the experimental period considered in this thesis delivered an electron beam of 5.014 GeV with an energy resolution of 10^{-4} and a typical diameter of 50–100 μm.

This chapter is structured as follows: section 2.1.1 describes the CLAS detector and its component sections, briefly describing its functions. Section 2.1.2 describes the experiments performed in Hall B of Jefferson Lab in which this thesis is based on.

2.1.1 The CLAS detector

CEBAF [53] is capable of delivering beams to four experimental Halls simultaneously. Each Hall is designed to target different physics. While Halls A and C focus on high resolution studies, Hall B hosts the CEBAF Large Acceptance Spectrometer (CLAS) [54], a detector with a very wide solid angle range of nearly 4π, capable of detecting charged particles with polar angles from 8° to 140° (the fourth Hall, D, started operating in 2014 after the experimental period this thesis describes). CLAS is well suited for experiments that require the detection of two or more particles in the final state as it connects the polar angle range with the full azimuthal coverage.

CLAS is composed of several sets of detectors separated in six essentially independent magnetic spectrometers. In it, drift chambers are used to determine the trajectories of charged particles, Cherenkov counters for electron identification, scintillation counters for measuring time-of-flight, and electromagnetic calorimeters for detecting electrons and photons. Neutrons can be also detected by combining information from the different systems. These measurements allow the identification of mass, charge and momentum of each particle.

The toroidal magnet bends charged particles toward or away from the beam axis. The six coils naturally separate the detector into six sectors (see fig. 2.2) where 18 separate **drift chambers** (DC) are located in three radial positions for each one. Region One chambers surround the target in an area of low magnetic field. Region Two chambers are placed in between the magnet coils in an area of high magnetic field, while Region Three chambers are located outside of the magnet coils. A schematic can be viewed in figure 2.1.

The **Cherenkov Counter** (CC) has the dual function of triggering using electrons as well as allowing the distinction of electrons from negative pions. It was designed in such a way as to maximize the solid-angle coverage with the least amount of material. The readout instrumentation (such as photomultiplier tubes) are placed in the regions already blocked by the magnet coils.

The **time-of-flight system** (TOFS) is composed of 206 m^2 of scintillator counters, 5.08 cm thick, 15 and 22 cm wide and lengths varying from 32 cm at the forward angle to

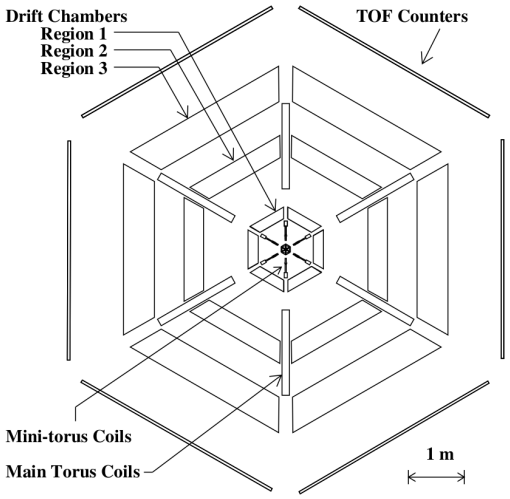


Figure 2.1: Transverse view of the CLAS detector showing the regions of the Drift Chambers as well as the Torus Coils and TOF counters

450 cm at large angles. It has been designed to optimize time resolution, achieving 80 ps for the short counters to 160 ps for the longer counters. The system is located outside of the tracking system and Cherenkov Counters but in front of the calorimeters. For more information see ref. [55]

The **electromagnetic calorimeter** (EC) is able to detect electrons at energies above 0.5 GeV and photons at energies above 0.2 GeV. Acceptance for electrons covers the θ angles from 12° to 70°. It is made of 39 alternating layers of lead sheets and scintillator in a 0.24 thickness ratio for a total thickness of 39 cm of scintillator and 8.4 cm of lead, for a total of 16 radiation lengths. It uses a projective geometry in which the area of each successive layer increases. The scintillator layers consist of 36 parallel strips which are rotated 120° after each layer, thus defining three *views*, which provide stereo information on the location of the energy deposits. In addition to this, the EC has its own timing system, independent of the TOFS. It is used to differentiate photons and neutrons, and is sufficient to provide a start time in case any channels in the TOFS are inoperative in the forward region. A complete description of the EC can be found in ref. [56].

Neutron detection can be achieved by combining information from the electromagnetic calorimeter, drift chambers and time-of-flight. While the EC can't be used to directly measure the energy deposited by the neutron, hits are recorded in all three views for more than 50% of cases. Neutron energy is obtained using the EC TOF information, and their detection efficiency is measured using the missing mass technique for the $p(e, e'\pi^+)$ reaction. The missing mass momentum must match the direction of a neutral hit in the calorimeter, i.e. a hit registered in the EC but not the DC. The efficiency is around 60% for neutrons of momentum above 1.6 GeV and falls linearly below that, as described in ref. [56].

2.1.2 The EG2 experiment and Double Target System

A new double-target system [57] was developed for precision measurements of nuclear medium effects in unpolarized electron scattering in the E02-104 and E02-110 experiments. One cryo-target and one solid target were simultaneously located in the beam separated by 4 cm to minimize acceptance correction differences while maintaining the ability to identify the target event by event. With this system the two targets are put in

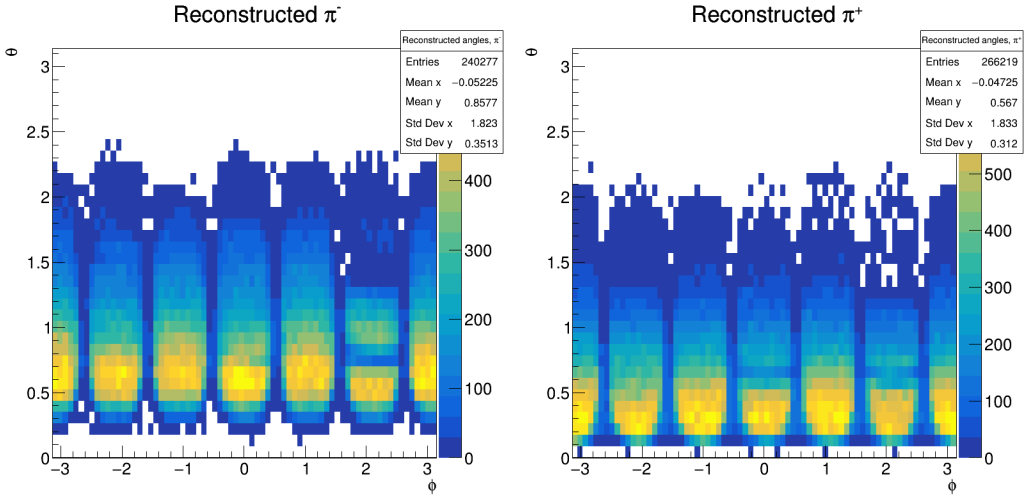


Figure 2.2: Left:Reconstructed negative pions. Right: positive pions in the laboratory frame angular space. The six sectors of the detector can be seen in the ϕ axis. The tendency of negative particles to bend away from, and positive particles towards the beam axis (located at $\theta = 0°$) can be seen comparing the θ axis in these two plots. The low acceptance regions in between the sectors are where the magnet coils are located

the electron beam at the same time, hence having the same luminosity and thus reducing the errors in the estimation of the ratio of the cross sections. A picture of the double-target system setup can be seen in figure 2.3.

The E02-104 experiment sought to measure hadron attenuation and transverse momentum broadening in order to study quark propagation through the nuclear medium and hadronization. The E02-110 on the other hand proposed to measure the attenuation of rho mesons in nuclei, looking for color transparency. Both of these experiments were carried out during the EG2 run period, from January 9th 2004 to March 5th 2004. As both of these experiments sought to measure subtle effects, the double-target design also helps in minimizing the systematic uncertainties.

The design of the target contemplated some key features, for example, a large acceptance for semi-inclusive and exclusive kinematics, a good match to the CLAS detector acceptance, minimal mass for low-energy particles at large angles as well as forward particles, similar scattering rates for the two targets in the beam, rapid target changes for the heavy nuclear targets, among others. One of the targets needed to be a stable deuterium cryotarget, while the second was a solid, heavy nuclei target. The cryotarget was located upstream of the solid targets, limiting the effects of secondary electromagnetic processes contributing to the flux incident on either target. A full description of the double target system can be found in [57]. A GEANT3 simulation of the double target system was added to the existing simulations of the full CLAS detector.

The details of the target data considered for this thesis are shown in table 2.1.2

Target	Longitudinal dimension	Transverse thickness areal density (g/cm^3)
Deuterium		
Carbon	1.7 mm	0.33
Iron	0.40 mm	1.2
Lead	0.14 mm	1.7

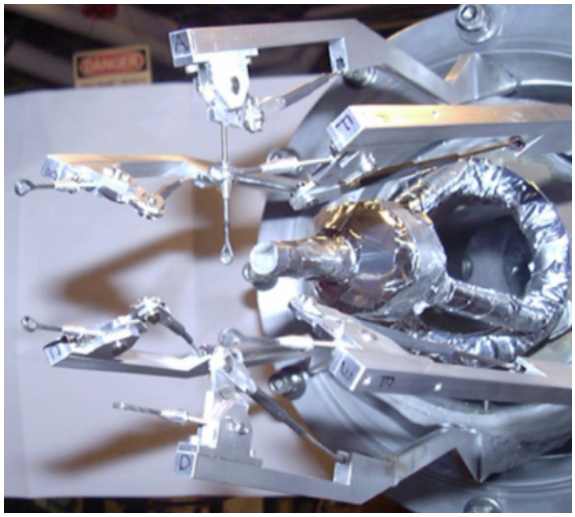


Figure 2.3: Experimental setup of the double-target system. Six arms hold different solid targets; 5 arms are retracted while the top arm is extended, placing its target in front of the cryotarget which holds deuterium (covered in foil). Both are placed in the beamline, which exits the page.

2.2 EG2 data analysis

In this section we will describe how the EG2 data was filtered and prepared for the studies and why these choices were made.

2.2.1 Compact File Format

The EG2 experimental data is stored in a format based around ClasTool, a collection of C++ classes meant to be used as a ROOT package for analyzing CLAS data. We refer to this as ClassTool format. This format is meant to provide easy access while still retaining full detector information, and facilitate the migration from previous file formats such as BOS.

In practice however, the ClasTool format proved to be too cumbersome for performing end user analyses, by retaining too much information and only "raw" detector variables. Running a script over all 4 TB of data is also slow, but more importantly the UTFSM group has for a long time performed particle identification outside of the existing ClasTool files via a software called Analyser.

Smaller files with selected events were used, but the formats and variables were not standarized and their origins couldn't be traced back to the original run numbers and events. A significant problem was that event synchronization was lost for simulation files, so for example a study of bin migration was not possible.

As part of this thesis a file format was designed, which would be able to maintain event synchronization in simulations, reduce the file size through skimming and thinning of the full dataset, storing end-analysis variables, and maintain traceability for transparency in the analysis.

2.2.2 Event and particle pre-selection

The selection of events for the files were simple

1. Full reconstruction of the scattered electron

2. At least one π^+ or π^- in the event

3. There had to be certainty on whether the scattered particle belonged to Deuterium or the heavy nuclei target.

Information about the scattered electron is fundamental for semi-inclusive deep inelastic scattering (SIDIS) studies, as it is used in triggering, and many of the variables of the event are calculated from the virtual photon which is directly correlated with this electron. Pions are the main source of information for hadronization studies, due to the quantity of them (being the most readily produced hadron) and how well they can be reconstructed. The double target system used in EG2 has many benefits but these can only be reaped if you can differentiate which target the electron is hitting. In the case of simulation files these three conditions must be satisfied both for *thrown* and *accepted* particles.

Once an event passes these criteria, only information about charged pions, photons and the scattered electron is kept. Unidentified particles are discarded.

The total event yield is detailed in table 2.1

Table 2.1: Yield of skimmed data files generated

Target	Number of events	Total Filesize
C	24,112,917	4.8 GB
Pb	24,421,879	4.9 GB
Fe	41,734,758	8.3 GB

2.2.3 Virtual photon frame of reference

Many of these studies need to be performed in the center of mass of the *virtual photon – initial proton* system, or just virtual photon frame for short. Assuming that we can model the interaction as a single-photon exchange, we can compute the four-momentum of the virtual photon $(\gamma^*)^\mu$ using the information we have for the four-momenta of scattered electron e'_μ and the electron beam e^μ. By 4-momentum conservation at the vertex,

$$(\gamma^*)^\mu = e^\mu - e'^\mu = \begin{pmatrix} E - E' \\ -p_x \\ -p_y \\ p_{\text{beam}} - p_z \end{pmatrix}. \tag{2.1}$$

The virtual photon frame is defined as the one in which the virtual photon lies in the z axis, with the x axis in the plane of the scattered electron, and boosted in the z direction. The coordinate transformation for a particle of four-momentum P^μ in the lab frame and $(P')^\mu$ in the virtual photon frame is obtained by two rotations and a boost,

$$\left(P'\right)^\mu = R_3 R_2 R_1 P^\mu \tag{2.2}$$

where

$$R_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\varphi_{e'}) & -\sin(\varphi_{e'}) & 0 \\ 0 & \sin(\varphi_{e'}) & \cos(\varphi_{e'}) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.3}$$

$$R_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta_{e'}) & 0 & -\sin(\theta_{e'}) \\ 0 & 0 & 1 & 0 \\ 0 & \sin(\theta_{e'}) & 0 & \cos(\theta_{e'}) \end{pmatrix}, \tag{2.4}$$

$$R_3 = \begin{pmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{pmatrix}. \tag{2.5}$$

In R_3, $\gamma = 1/\sqrt{1-\beta^2}$ is the usual Lorentz factor (not to be confused with a photon). The boost β has to be such that the momentum of the γ^*–nucleon system is zero. Applying the boost (2.5) to the virtual photon and the nucleon at rest (mass M_n),

$$R_3 \begin{pmatrix} E_\gamma \\ 0 \\ 0 \\ p_\gamma \end{pmatrix} = \begin{pmatrix} \gamma E_\gamma - \beta\gamma p_\gamma \\ 0 \\ 0 \\ -\beta\gamma E_\gamma + \gamma p_\gamma \end{pmatrix}, \qquad R_3 \begin{pmatrix} M_n \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma M_n \\ 0 \\ 0 \\ -\beta\gamma M_n \end{pmatrix} \tag{2.6}$$

we obtain adding the z components and equating to zero,

$$\beta = \frac{p_\gamma}{E_\gamma + M_n}, \tag{2.7}$$

where we note p_γ and E_γ are the virtual photon momentum and energy after the two rotations (and is now entirely in the z axis). This matrix representation assumes that the 4-momentum is a row vector with energy as its first component.

A diagram representing the main steps in the frame change can be seen in figure 2.4. After performing the frame change, we end up with a collider experiment setup, as shown in picture (c) of figure 2.4, in which the virtual photon is opposite to the nucleon along some "beam" axis which needs to be calculated event by event. This allows us to compare more directly our results with those of collider experiments and to use the appropriate variables, such as rapidity y, which is defined as

$$y = \frac{1}{2}\log\left(\frac{E + p_z}{E - p_z}\right), \tag{2.8}$$

where E is the energy and p_z the z-axis momentum of the particle, both in the virtual photon frame. The rapidity represents the boost an observer would need to make to see the particle as moving perpendicular to the beam, thus one can define a reference system using y, the azimuth angle ϕ and the transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$. More importantly however is the fact that rapidity differences Δy are Lorentz invariant under boosts in the beam axis. It should be noted that in experiments such as ATLAS, an approximate variable called *pseudorapidity* (η) is used instead of rapidity as it is a purely angular variable and does not depend on the particle type. As our data consists of well-identified pions, rapidity will be preferred.

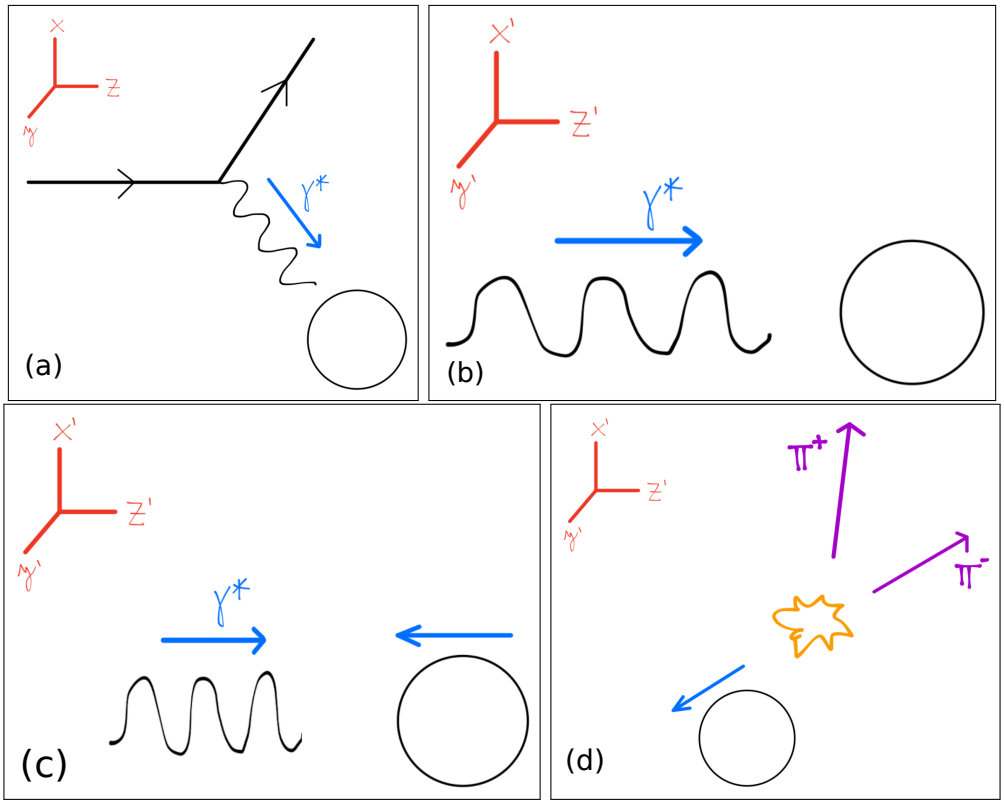


Figure 2.4: The black solid line represents the electron before and after scattering, the sinusoidal line the virtual photon and the big circle a nucleon. (a) In the lab. frame the electron interacts with the nucleon via a virtual photon γ^*. (b) γ^* is left in z' after rotations. (c) The frame is boosted in the z' direction to the center-of-mass frame. We have transformed the fixed target experiment into a collider one. (d) All of the resulting particles (pictured: two pions and a nucleon) are in the center of mass frame.

2.3 Two particle correlations

In a physics scattering event, two particles may be correlated in a variable v. This may occur for example if the particles originate from the decay of a common parent particle. To study this, we consider pairs of particles in an event, denominated *trigger* and *associated* particle. These can be chosen by many criteria, for example by their transverse momentum p_T, Z_h value, electric charge, or particle species. A physical correlation in variable v can be studied by taking the differences $\Delta v = v_1 - v_2$ where 1 correspond to the trigger and 2 to the associated particle.

This measurement of Δv may contain a physics correlation as well as other types of correlation such as those due to detector geometry. A physics-correlation-free sample may be obtained by selecting triggers and associated particles from two different events (mixed events), where physics correlations would be impossible, but detector correlations are unaffected. By then taking the ratio of these distributions all correlations other than physical ones will cancel out, and we obtain a correlation index N_c as a measure of the strength of the correlation,

$$N_c(\Delta v) = \frac{N_s(\Delta v)}{N_m(\Delta v)}. \tag{2.9}$$

Here N_s and N_m correspond to the yield of pairs taken from the same or from mixed events. This kind of analysis can be generalized to consider a higher number of variables

$$N_c(\Delta v_1, \Delta v_2) = \frac{N_s(\Delta v_1, \Delta v_2)}{N_m(\Delta v_1, \Delta v_2)}. \tag{2.10}$$

The N_s and N_m distributions are binned and filled event by event by the number of pairs N_{same} and N_{mixed} in the same and mixed events respectively, then normalized by the total number of trigger particles to obtain an average over all events,

$$N_s = \frac{1}{N_{\text{trig}}} \frac{d^2 N_{\text{same}}}{d\Delta v_1 d\Delta v_2} \qquad N_m = \frac{1}{N_{\text{trig}}} \frac{d^2 N_{\text{mixed}}}{d\Delta v_1 d\Delta v_2}. \tag{2.11}$$

2.3.1 Event selection and analysis

We will consider first as triggers all pions with high $Z_h > 0.4$. These pions carry most of the available energy of the event after hadronization. We will define three bins for the associated particles (which will be all of the remaining pions): first, low Z_h pions of $Z_h < 0.12$, medium Z_h of $0.12 < Z_h < 0.25$, and pions with Z_h comparable to the trigger, $0.25 < Z_h < 0.4$. The mixed events are obtained by selecting the associated particles from the previous 10 events with respect to the current triggers, within the same Z_h bin. Using 10 events ensures sufficient statistics. The non-overlapping cuts ensure that triggers and associated particles are distinct, and the high Z_h requirement for triggers will result in events having only one trigger particle.

The variables we will study are the ϕ angle in the virtual-photon nucleon frame, and rapidity y. The $\Delta\phi$ and Δy variables are binned as absolute values in the ranges (0,180) [degrees] and (0,2.1) respectively, as outlined in ref. [58] in order to increase the statistics in each bin. However, by using Z_h bins and different ranges for trigger and associated particles, we introduce a bias in Δy, as Z_h is correlated with rapidity. We can obtain the values of these biases by examining the Δy distribution in a wider range (-4.8,4.8), and extract the bias value by taking the mean value of rapidity; we then subtract this bias from the 2D distributions, so that they are centered around 0 in rapidity difference. The values obtained for the rapidity difference bias y_{bias} for each target and associated particle Z_h bin are presented in table 2.2. It can be seen that the bias is directly correlated with the Z_h bin of the associated particle as compared to the triggers.

Table 2.2: Values obtained for the rapidity bias, when using as triggers pions with $Z_h^{\text{trig.}} > 0.4$

Target \ $Z_h^{\text{assoc.}}$ bin	$Z_h^{\text{assoc.}} < 0.12$	$0.12 < Z_h^{\text{assoc.}} < 0.25$	$0.25 < Z_h^{\text{assoc.}} < 0.4$
Deuterium	1.578	0.9759	0.5042
Carbon	1.606	1.003	0.4939
Iron	1.510	0.969	0.4806
Lead	1.496	0.9623	0.5376

The 2D correlation plot for the different targets in all Z_h bins with subtracted biases can be seen in figure 2.5. Despite the mentioned bias, the rapidity distribution was found to be nearly symmetrical, so the correlation is reflected around (0,0) in both $\Delta\phi$ and Δy in order to better appreciate any noteworthy structures at $\phi \sim \pi$ and $\phi \sim 0$ and increase bin statistics. A far-side ($\Delta\phi \sim \pi$) long-range correlation can be appreciated in rapidity

for the lower bins in associated Z_h, but this correlation becomes short-range in rapidity for the highest bin, mostly centered around $\Delta y \sim 0$. No near-side ($\Delta\phi \sim 0$) correlation that can be compared to the ridge phenomena is observed. The peak at $(0, 0)$ observed in most literature is absent also, due to the low multiplicity of the events and is indicative of the lack of jets in the events.

To quantify the azimuthal correlation, a one-dimensional distribution in $\Delta\phi$ is obtained from the 2D histograms by integrating in rapidity for $|y| > 1.0$. By doing this we ensure that any effect by structures at the origin (0,0) do not affect the distributions and we can obtain a *ridge yield*. The 1D data along a third order Fourier fit can be seen in figure 2.6. The third order Fourier decomposition in its harmonics can be seen in figure 2.7. The Fourier decomposition shows the appearance of the second-order term, known as v_2 or *elliptic flow*, for the higher Z_h bins (center and right plots) in all targets, while the lowest bin (plots on the left) only shows significant contribution of the first term (v_1 or *directed flow*). The v_2 term appearance for high associated Z_h is stronger for deuterium, which indicates a suppression of elliptic flow due to the nuclear medium. The values of v_2 can be seen in table 2.3.

Table 2.3: Values obtained for the second Fourier coefficient for events with $|\Delta y - y_{\text{bias}}| > 1.0$

Target \ bin	$Z_h^{\text{assoc.}} < 0.12$	$0.12 < Z_h^{\text{assoc.}} < 0.25$	$0.25 < Z_h^{\text{assoc.}} < 0.4$
Deuterium	$1.86 \cdot 10^{-03}$	$7.73 \cdot 10^{-03}$	$1.42 \cdot 10^{-02}$
Carbon	$-1.58 \cdot 10^{-03}$	$4.89 \cdot 10^{-03}$	$7.77 \cdot 10^{-03}$
Iron	$-1.40 \cdot 10^{-03}$	$3.05 \cdot 10^{-03}$	$6.64 \cdot 10^{-03}$
Lead	$-7.67 \cdot 10^{-04}$	$4.57 \cdot 10^{-03}$	$3.91 \cdot 10^{-03}$

2.3.2 Nuclear medium dependence

To fully appreciate the nuclear medium dependence of the azimuthal correlation we can plot simultaneously the distributions for the different targets. First, we plot the results integrating in rapidity in figure 2.8. We can see a pattern of the correlation being more pronounced the lightest the target, with a tendency toward less extreme behavior the heavier the target. By introducing the rapidity cut we obtain the distributions shown in figure 2.9.

It was found that a different particle pair selection enhances the nuclear medium effects. By using as triggers particles with high transverse momentum $p_T^{\text{trig.}} > 0.4$ GeV and using transverse momentum bins for the associated particle of $p_T^{\text{assoc.}} < 0.2$ GeV, 0.2 GeV $< p_T^{\text{assoc.}} < 0.3$ GeV and $p_T^{\text{assoc.}} < 0.4$ GeV, we obtain the distributions in figure 2.10. The separation between targets in the plots is greatly enhanced as the associated particle has higher transverse momentum, being the maximum in the highest bin, where both trigger and associated particles have similar p_T. It should be noted that unlike for the Z_h binning, p_T bins introduce no bias in the rapidity variable. Now, when performing a cut for high rapidity difference pairs ($|\Delta y| > 1.0$) we obtain the distribution shown in figure 2.11. This corresponds to a significant reduction in transverse momentum for pairs with high azimuthal difference for heavy nuclei.

2.3.3 Q^2 dependence of v_2

The second Fourier coefficient, v_2 is of particular interest in anisotropic flow studies, as in heavy ion collisions it is considered to be sensitive to the early stages of system

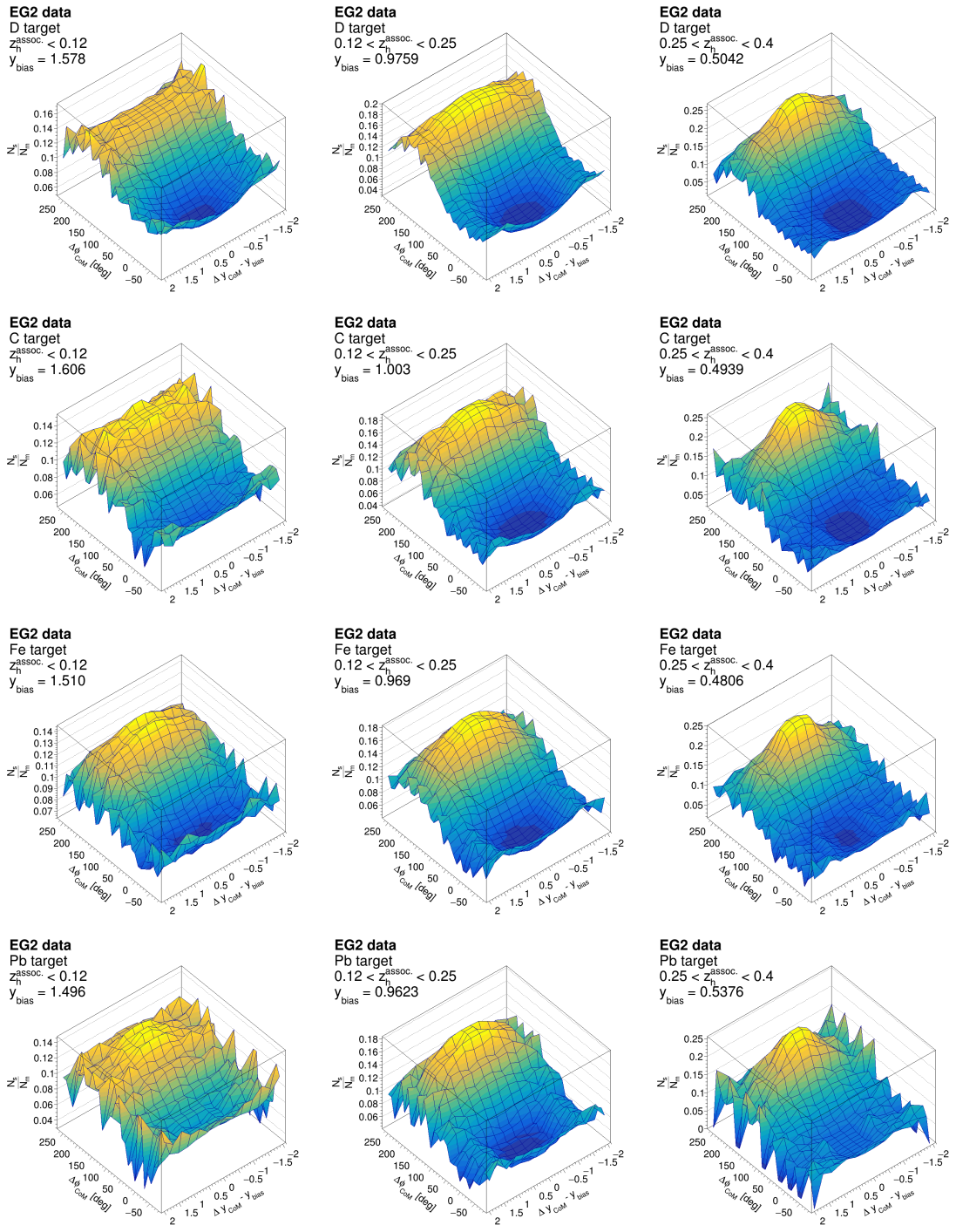


Figure 2.5: 2D correlation plots for the different targets. From top to bottom: D, C, Fe and Pb targets (increasing nuclear size). From left to right, cuts of increasing Z_h for the associated particle

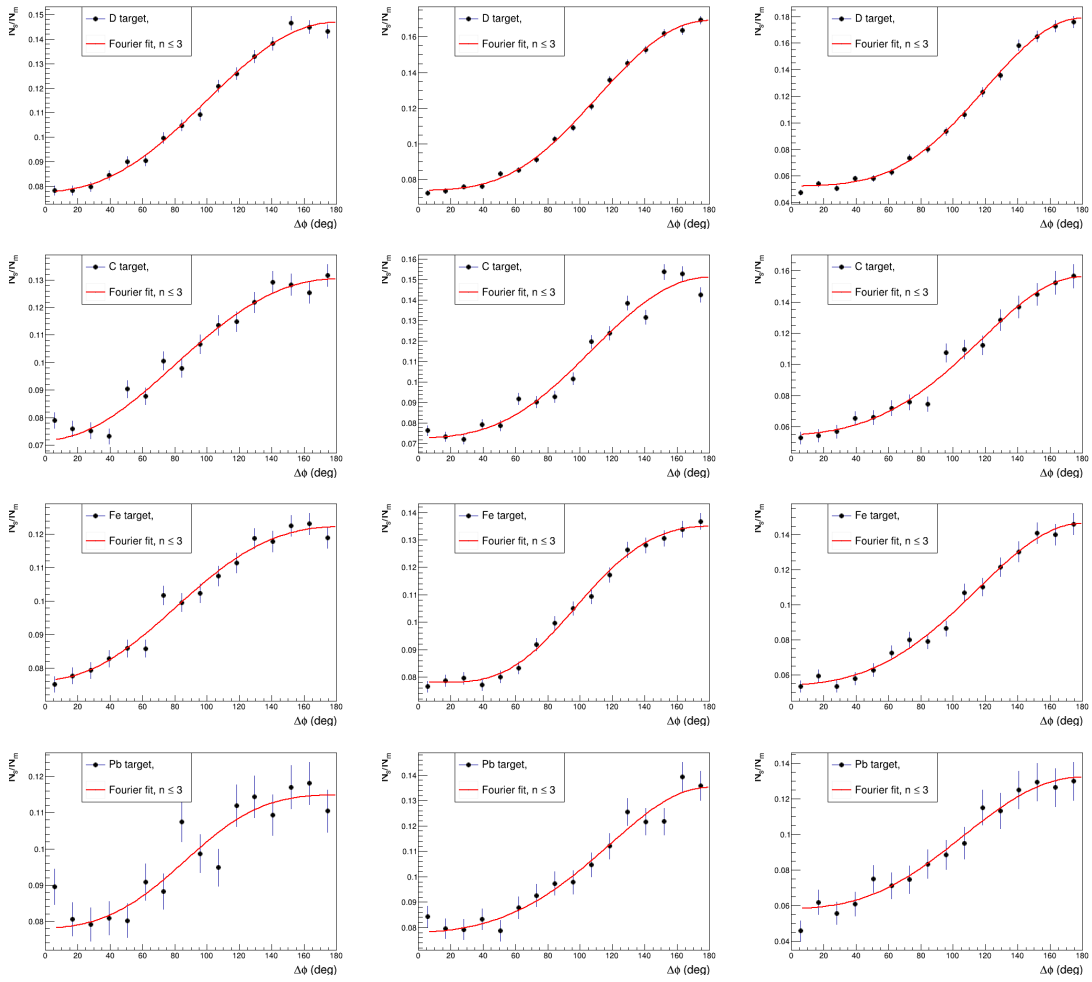


Figure 2.6: Third order Fourier fits in the $\Delta\phi$ projection using a $|y| > 1.0$ cut for the different targets and associated particle bins. From top to bottom: D, C, Fe and Pb targets (increasing nuclear size). From left to right, cuts of increasing Z_h for the associated particle

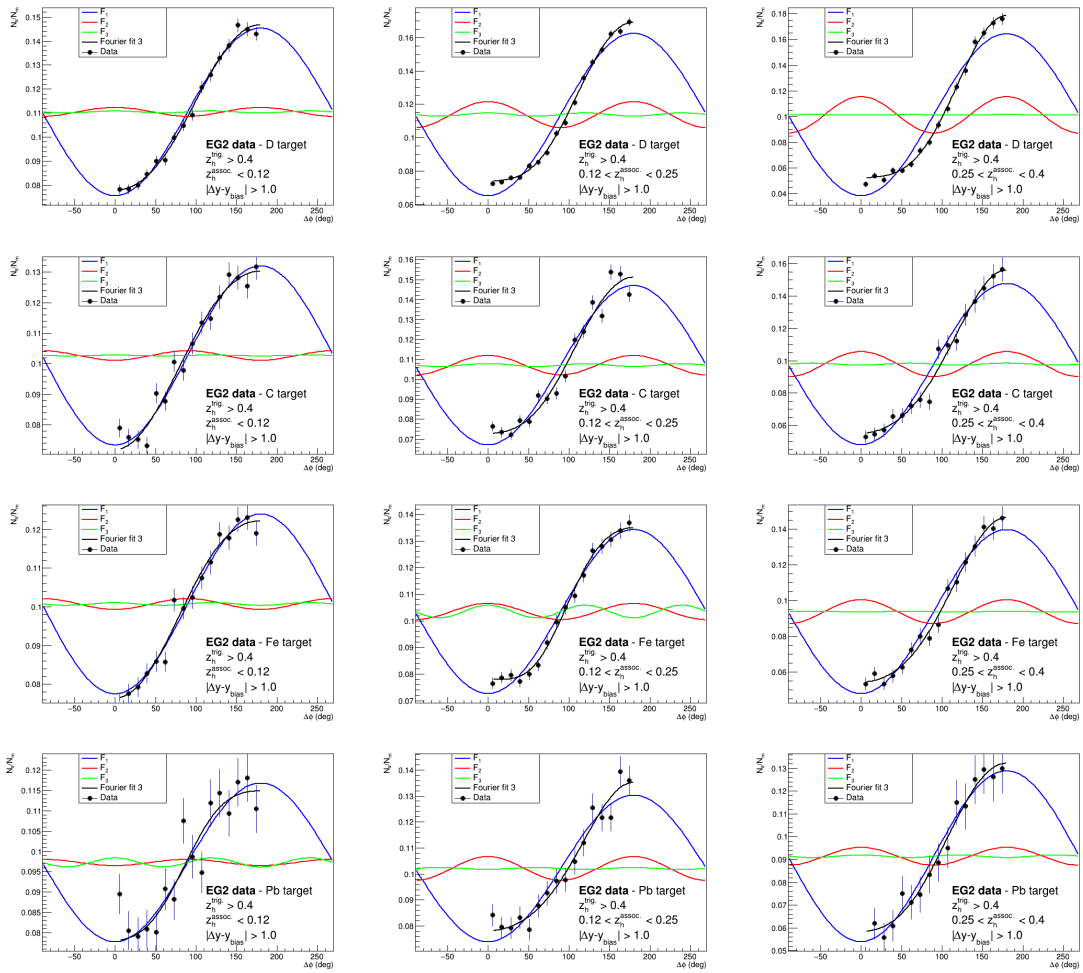


Figure 2.7: Fourier decomposition in 1D for $\Delta\phi$ for different targets. From top to bottom: D, C, Fe and Pb targets (increasing nuclear size). From left to right, cuts of increasing Z_h for the associated particle

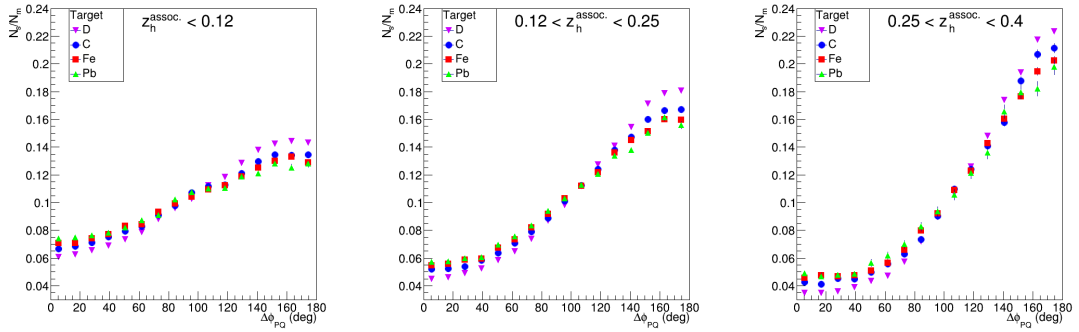


Figure 2.8: Azimuthal correlations for the different targets in the Z_h bins for the associated particle, inclusive in rapidity

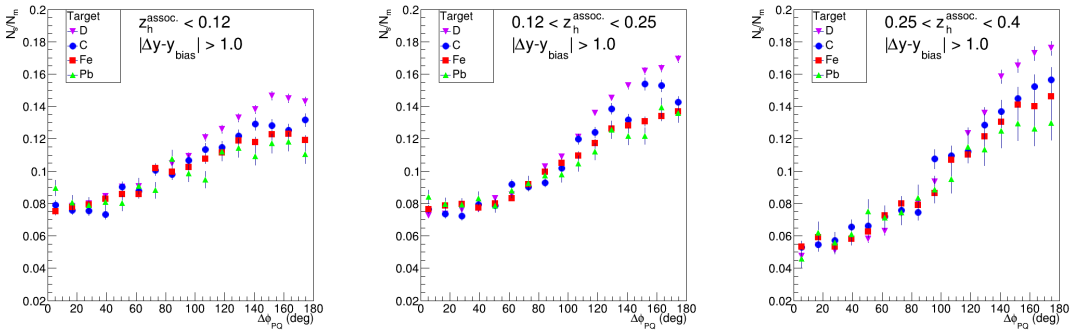


Figure 2.9: Azimuthal correlations for the different targets in the Z_h bins for the associated particle

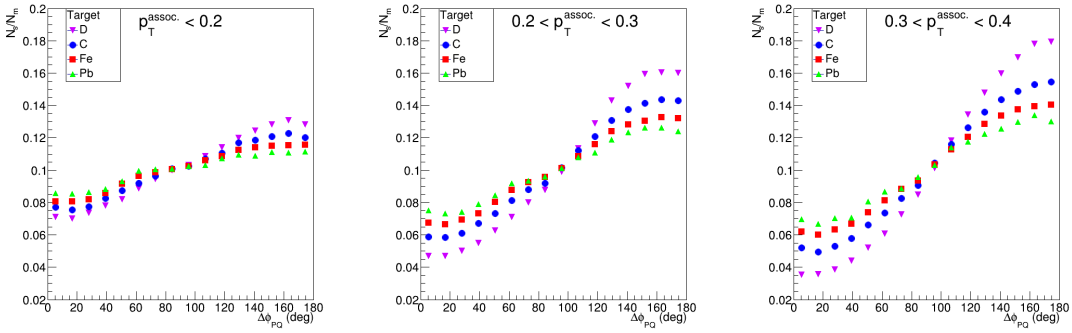


Figure 2.10: Azimuthal correlations for the different targets in the p_T bins for the associated particle, inclusive in rapidity.

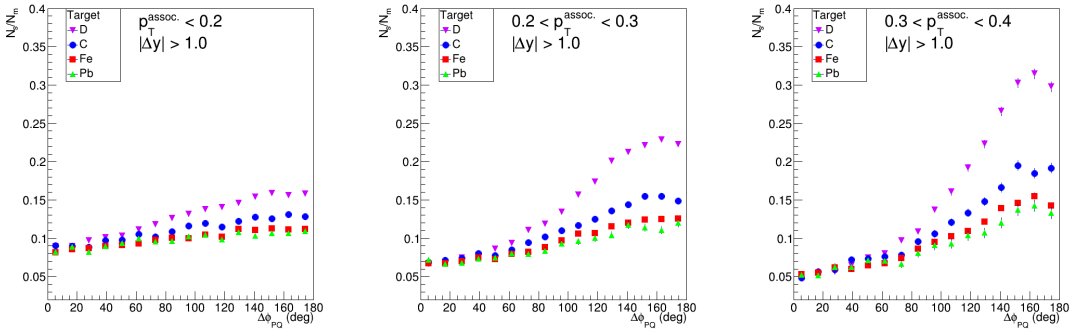


Figure 2.11: Azimuthal correlations for the different targets in the p_T bins for the associated particle.

evolution [59]. The usual way to characterize v_2 is in terms of either collision centrality, the closely associated particle multiplicity, and transverse momentum p_T. Since centrality is harder to measure in electron collisions, and the particle multiplicities we obtain in our experiment are low, we can characterize v_2 using an equivalent variable to p_T. In this case it would be Q^2 as they are both a measure of interaction "hardness". In this case, no shifting in rapidity is performed, and the two-dimensional distribution is left asymmetric (i.e. we fill the histograms with Δy as opposed to $|\Delta y|$).

We will use triggers with $Z_h > 0.4$ and partners with $Z_h < 0.4$, while using four bins in Q^2 chosen to (roughly) evenly distribute the event pairs in each. The bins are $1.0 < Q^2 < 1.25$, $1.25 < Q^2 < 1.52$, $1.52 < Q^2 < 1.92$ and $1.92 < Q^2 < 5.0$.

In addition to this, we will consider a restriction on the fraction of the lepton's energy lost, ν/E, in order to minimize contamination by radiative processes. If the electron radiates a photon before or after the main scattering, it would affect the calculation of variables such as Q^2 or W. We will therefore consider only events for which $\nu/E_{\text{beam}} < 0.85$. The results of v_2 for the different Q^2 bins can be appreciated in figure 2.12. The full 2-dimensional correlation functions are shown in figure 2.13, and the Fourier fits in figure 2.14.

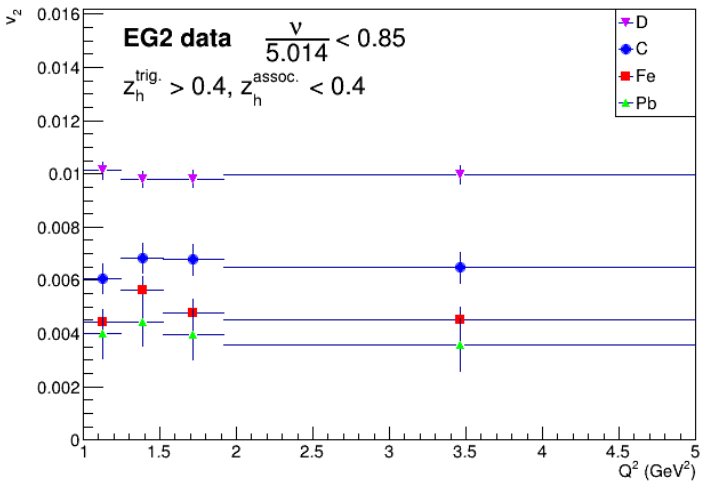


Figure 2.12: Q^2 dependence of v_2, for the different targets

Unlike what has been observed for heavy ions at high energies [60], v_2 doesn't seem to increase with Q^2 and is consistently flat, as it does for high p_T in the aforementioned systems. This lack of dependance on Q^2 could be due to the low energy or multiplicity of the system, and a future study using CLAS12 [61] data could be used to improve this measurement.

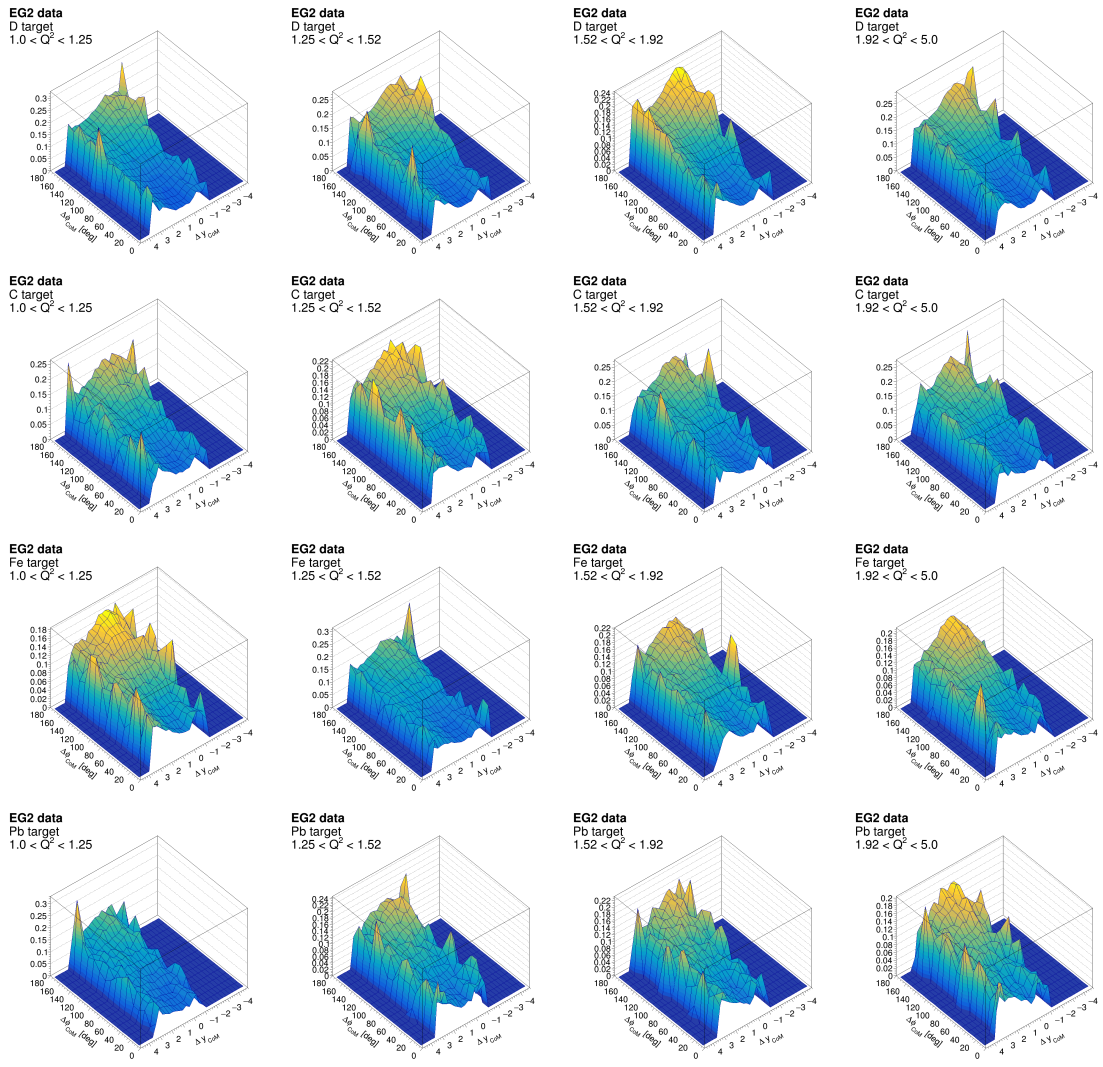


Figure 2.13: 2D correlation plots for the different targets. From top to bottom: D, C, Fe and Pb targets (increasing nuclear size). From left to right, cuts of increasing Q^2 for the event

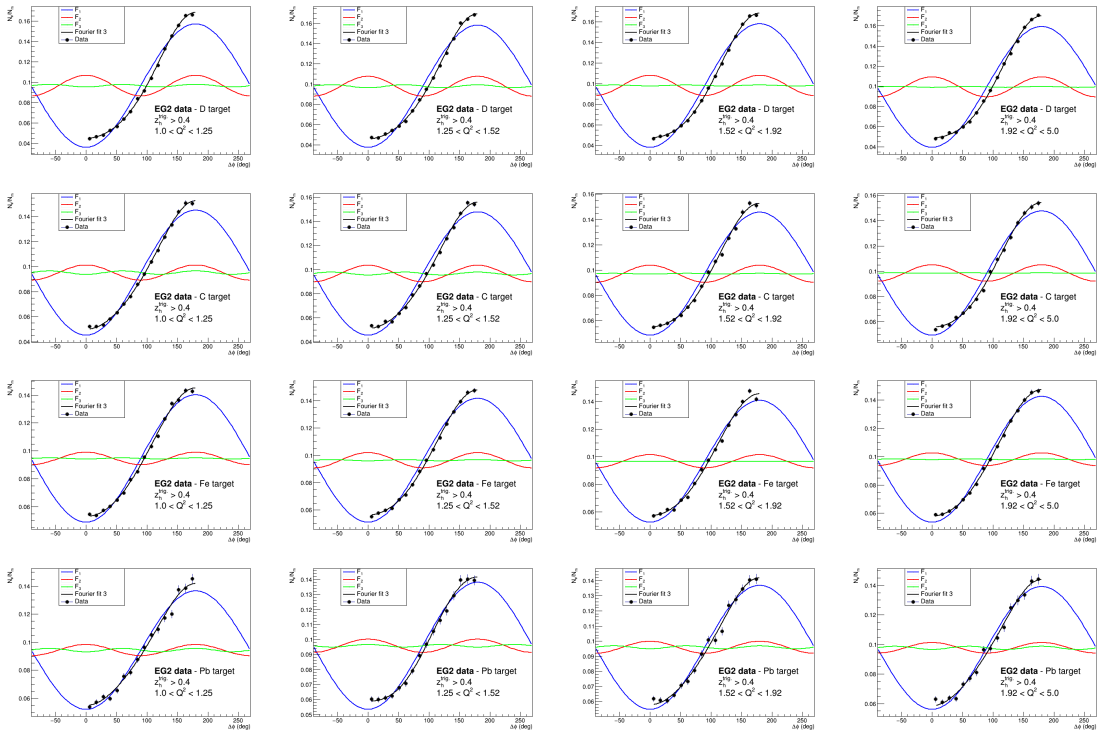


Figure 2.14: Fourier decomposition in 1D for $\Delta\phi$ for different targets. From top to bottom: D, C, Fe and Pb targets (increasing nuclear size). From left to right, cuts of increasing Q^2 for the event

3 ATLAS

3.1 The ATLAS experiment at CERN

The Large Hadron Collider (LHC) is the biggest, highest energy particle collider ever made. It is built in a 27 km circumference tunnel at the border between France and Switzerland, near Geneva, and it is administered by the European Organization for Nuclear Research (CERN). Using this collider the Higgs boson was discovered in 2012 by the ATLAS and CMS collaborations, delivering one of the last fundamental pieces for the validation of the Standard Model.

The LHC serves as the basis for many particle physics experiments at CERN. This thesis will focus on proton–proton collisions at a center of mass energy of $\sqrt{s} = 13$ TeV, as measured by the ATLAS experiment in its second experimental run, from 2015 to 2018 (known as Run 2).

This chapter is structured as follows. Section 3.1.1 will describe with some detail relevant technical aspects of proton–proton collisions at the LHC. Section 3.1.2 will describe the ATLAS detector.

3.1.1 LHC

The LHC[62] was built between 1998 and 2008 in the tunnel which was previously used for the Large Electron-Positron Collider. The collider tunnel has two parallel beamlines, each containing a beam of particles traveling in opposite directions respect to each other in an ultrahigh vacuum in the order of 10^{-10} to 10^{-11} mbar. The collider is used most of the time for proton–proton collisions, although it is also capable of handling heavy ions, which is typically done one month a year. We will describe the proton–proton mode in what follows.

Protons are accelerated to their peak energy in various stages. Before entering the LHC, they go through the injection chain Linac 2 — Proton Synchrotron Booster (PSB) — Proton Synchrotron (PS) — Super Proton Synchrotron (SPS), which is illustrated in figure 3.1. First, a bottle of hydrogen gas at one end of the Linac 2 is passed through an electric field to strip the gas of its electrons. The now free protons enter the linear accelerator and through the use of radiofrequency (RF) cavities reach an energy of 50 MeV at the end of it. The protons are then injected into the PSB, which consists of four synchrotron rings with a radius of 25 meters. In it, protons are accelerated to an energy of 1.4 GeV.

Synchrotrons are ring-shaped accelerators that work by using sections with RF cavities for particle acceleration, dipole magnets for beam guiding, and quadrupole magnets for beam focusing. As the name implies, the magnetic field guiding the particles increases in synchronization with the increasing particle energy, so that the particles can continue accelerating in the same ring of constant radius when passing through the RF cavities section. Beam focusing is used to avoid radial or transverse deviations (transverse or betatron oscillations) from the beamline, and is typically achieved using dedicated quadrupole magnets, which squeeze the beam either horizontally or vertically. Another important feature of synchrotrons are synchrotron oscillations, which arise when a particle arrives earlier or

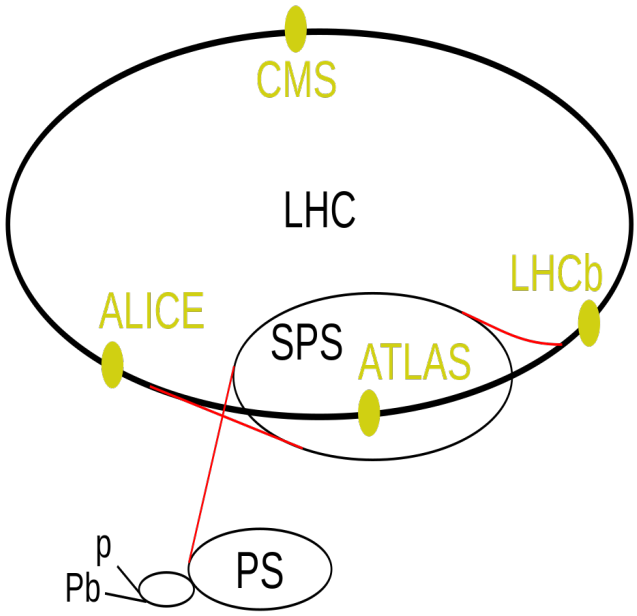


Figure 3.1: Main accelerator stages in the LHC along with the main four detectors

later than the ideal phase with respect for the RF acceleration. When particles in the beam reach its target energy, RF cavities will not further accelerate them, and will in fact correct these oscillations by decelerating early particles and accelerating lagging ones. A very important consequence of this is the formation of packs of protons in the positions of optimal RF acceleration, called *bunches*.

The 1.4 GeV protons from the PSB are fed into the PS, which has a circumference of 628 meters and uses 277 room-temperature magnets, resulting in their acceleration into 25 GeV. These are injected into the SPS, the second-largest accelerator at CERN. With a circumference of 7 km and 1317 room-temperature magnets, it provides particle beams for the LHC, as well as the NA61/SHINE, NA62, NA64 and COMPASS experiments, among others. The SPS injects the LHC with 450 GeV protons in both clockwise and anticlockwise directions.

The LHC has two parallel beamlines with protons traveling in opposite directions respect to each other in an ultrahigh vacuum. As mentioned in the introduction this hadron collider is 27 km long, making it the biggest proton collider ever built. The beams collide at four interaction points where the main detectors (ATLAS, CMS, LHCb and ALICE) are located. Unlike other synchrotrons, the LHC uses 1232 *superconducting* dipole electromagnets for bending the particles, which can reach high magnetic fields of up to 8.3 tesla. In order to maintain superconductivity in the electromagnets, the accelerator is connected to a distribution system of 96 tonnes of superfluid helium-4 at a temperature of 1.7 K, making the LHC also the largest cryogenic facility in the world. Without superconducting magnets the LHC ring would need to be much larger to accommodate its target 7 TeV proton beams. During Run 2, the LHC accelerated proton beams to an energy of 6.5 TeV, for a center of mass collision energy of 13 TeV.

For particle acceleration the LHC uses 8 superconducting RF cavities, each delivering 2 MV at a frequency of 400 MHz and operating at 4.5 K. They accelerate the protons with a maximum of 16 MeV of energy each lap (in practice the acceleration stage takes

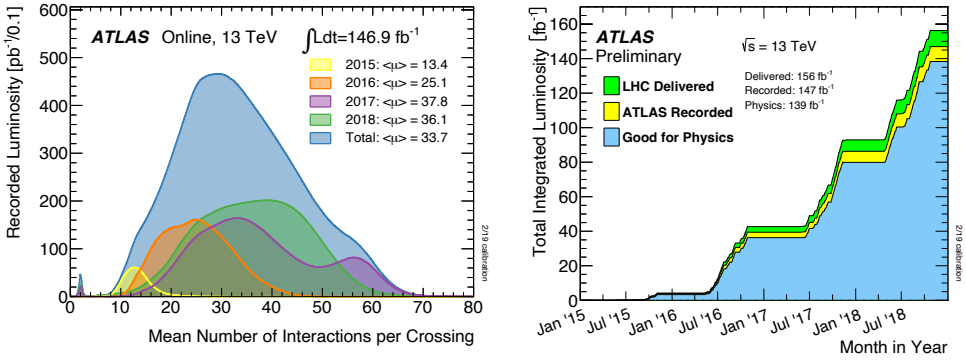


Figure 3.2: (a): Interactions per bunch crossing for Run 2 (b): Integrated luminosity for Run 2

around 20 minutes), and perhaps more importantly they are used to maintain a compact bunching of the protons, which will increase the probability of collisions at the interaction points.

Luminosity and *pile-up*

Each beamline contains around 2800 bunches spaced 25 ns apart, with bunches containing $1.15 \cdot 10^{11}$ protons. Bunches are further squeezed near the interaction points using a system of three quadrupole magnets called an inner triplet (two for each of the four experiments) which tighten the beam to a cross-sectional size of $16 \times 16\,\mu\text{m}^2$. The bunches cross at interaction points with a frequency of 40 MHz, and an average of 33.7 collisions occurred every time the bunches crossed for Run 2 (beyond the design operation of 20-25). Interactions per bunch crossing are referred to as *pile-up*, and the values of pile-up in Run 2 as measured by the ATLAS experiment can be seen in figure 3.2.

3.1.2 The ATLAS detector

ATLAS (A Toroidal LHC AparatuS) is one of the two multi-purpose detectors at the LHC (alongside CMS). It is comprised of various subsystems which allow it to detect many different kinds of particles and accurately measure their physical properties, all of this while handling the unprecedented rate of collision data generated at the LHC. In July 2012 it reported evidence of a new particle with a mass of 125 GeV consistent with the Standard Model description of the Higgs boson [10], thus achieving one of its main physics goals. Since then, the Higgs boson properties along searches for new physics have been largely studied in ATLAS by the over 3000 scientists from 183 institutions in 38 countries comprising the scientific collaboration.

The detector is shaped like a barrel with end caps around the beam axis, centered around the interaction point. The dimensions of the detector are 25 m in height and 44 m in length, with an approximate weight of 7000 tonnes. It has an Inner Detector used for tracking, a system of electromagnetic and hadronic calorimeters for measuring the energy of jets, electrons and photons, and a muon spectrometer system for detecting muons. The main subsystems will be described in some detail in the following subsections, with an emphasis on improvements made for Run 2. A full description of the original design of the detector can be found in [63], from which most of the following subsections are based upon.

Inner detector

Approximately 1000 particles emerge from the ATLAS collision point every 25 ns, thus the highest-precision measurements are required to reconstruct the high density of tracks and vertices at the innermost layer. The layout of the ATLAS Inner Detector can be seen in figure 3.3 and a cross-sectional view of the barrel section at 3.4. It is comprised of a semiconductor tracker (SCT), a Pixel Detector, and a Transition Radiation Tracker (TRT).

The high granularity near the interaction point is obtained using silicon pixel detectors in both barrel and endcaps. With a pixel size of 50×400 μm^2, in Run 1 the ATLAS Pixel Detector consisted of three barrel layers at distances of 50.5, 88.5 and 122.5 mm centered around the beam axis, and 2 endcaps with 3 disc layers each. An additional layer of pixel detectors, the Insertable-B-Layer (IBL) was implemented for Run 2 as the innermost layer of pixel detectors at 33.2 mm around the beam pipe (which was reduced in size from 29 to 25 mm to accommodate for it). The IBL improves tracking by adding an additional measurement point, and mitigates the losses of hits in the other 3 layers for the higher instantaneous luminosities of Run 2. In addition to planar sensors, it also uses 3D sensors (the first large scale application of its kind), allowing particles to interact with the bulk of the detector rather than just the surface. The improved vertexing increases b-jet identification (b-tagging), which will be further discussed in section 3.2.2. The endcaps of the Pixel Detector consist of three layers of pixels each.

In the next layer, the SCT consists of 4 barrel strip layers with an angular coverage of $|\eta| < 2.5$, and 9 layers in each endcap. When a charged particle passes through a pixel in the Pixel Detector or SCT, a "hit" is recorded. These hits are then used along those from the TRT to reconstruct the particle tracks. The inner detector is immersed in a 2 tesla solenoidal magnetic field, which allows to identify the charge and some momentum information of the particle using the bending of the tracks.

The TRT is located outside of the silicon detectors as the outermost layer in the barrel region of the Inner Detector. It consists of a series of 4 mm diameter *straw tube* detectors. This type of detector consists in a hollow tube filled with gas and a wire at the center, which is maintained at a potential difference with the walls. When a particle passes through the straw, the gas ionizes resulting in electrons and ions traveling in opposite directions, producing a current. In ATLAS the straws are made of polymide coated with aluminum and filled with a mixture of 70% Xe, 27% CO_2 and 3% O_2. The wire anodes are 31 μm diameter tungsten wires plated with 0.5–0.7 μm gold. The TRT provides a large number of hits per track, with a minimum of 36 crossed straws for a particle of $p_T > 0.5$ GeV at $|\eta| < 2.0$ (except for the barrel-endcaps transition region at 22 hits). In the barrel region the straws are parallel to the beam axis with a length of 114 cm, and in the endcaps they are arranged radially with an uniform 37 cm long. The straw hits at larger radius contribute significantly to momentum measurement, as their lower precision is compensated by the number of hits measured.

Calorimeters

The ATLAS calorimetry system covers the range $|\eta| < 4.9$ using different techniques according to the physics needs of the region. It can be divided in two subsystems; the Electromagnetic Calorimeter (EM calorimeter) is in a region matched to the inner detector and has a higher granularity suited for precision measurements of electrons and photons; the Hadronic Calorimeter has a coarser granularity, sufficient for jet reconstruction and missing transverse energy (MET) measurements[63].

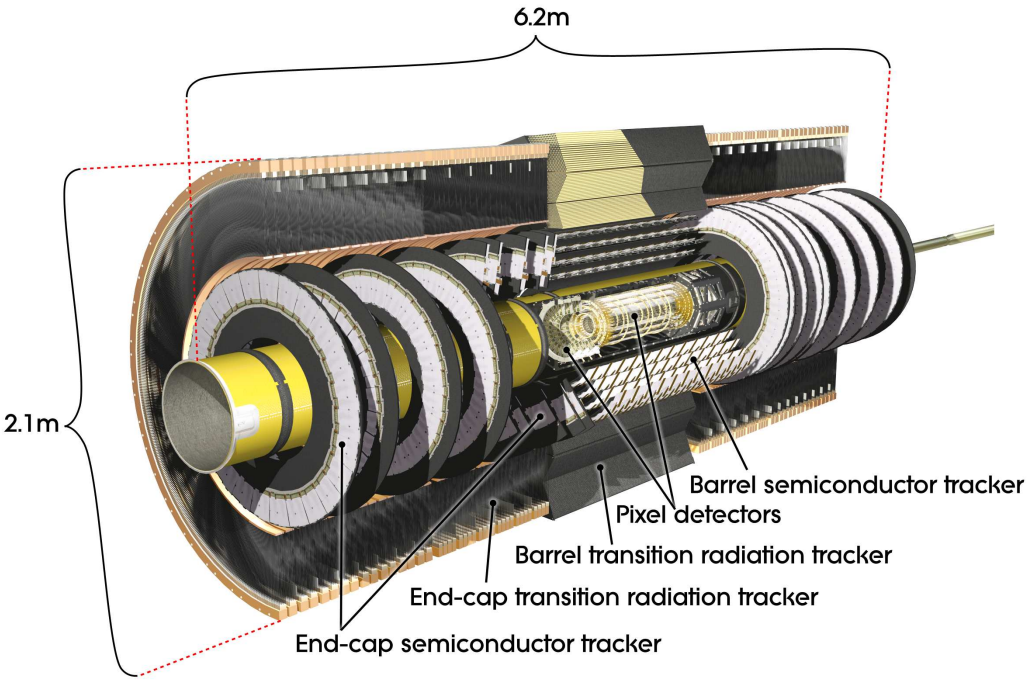


Figure 3.3: The ATLAS inner detector.

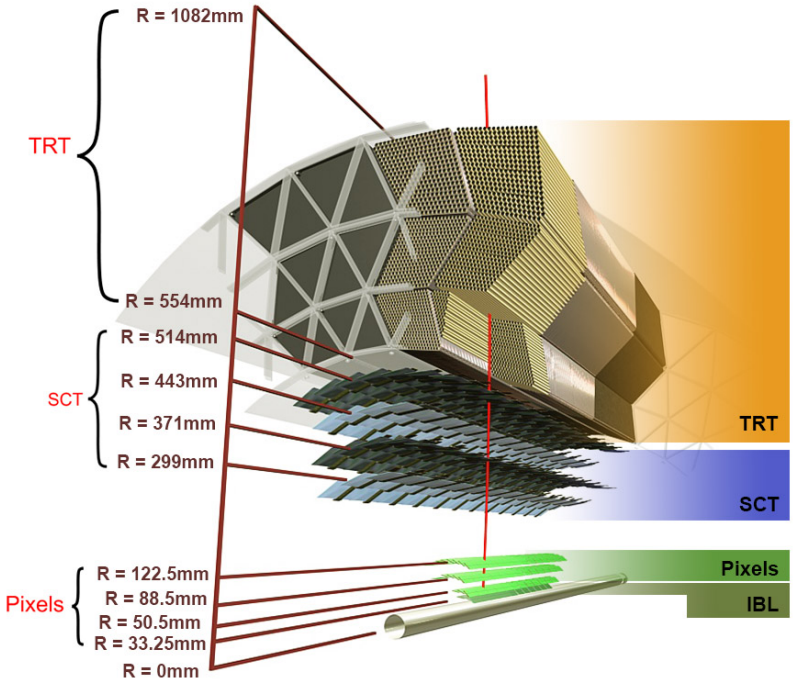


Figure 3.4: Transverse view of the barrel section of the ATLAS inner detector, including the IBL added for Run 2.

The goals of a calorimeter are measuring the energy of particles and the spatial location of the energy deposits. They do this by fully containing the incoming particles until they lose all their energy through interactions with the calorimeter medium. This means they are designed to stop and fully contain the particles they measure. Both calorimeter systems in ATLAS are fully ϕ-symmetrical *sampling* calorimeters, which alternate layers of an active material used for the measurement itself, and an absorbing material used for facilitating particle showers. This effectively stop particles by converting the relatively few high-energy particles which come in into a shower of lower-energy ones.

In the EM calorimeter the energy loss is determined by the *radiation length* X_0, defined as the mean path length over which a high-energy electron energy is reduced by a factor $1/e$ due to Bremsstrahlung, and is a property of the material. In the case of ATLAS, the EM calorimeter uses lead as absorber material and possesses a total thickness of over 22 radiation lengths (X_0) in the barrel and over 24 in the end-caps. It uses liquid argon (LAr) as its sampling material, and in the region $|\eta| < 2.5$ it is segmented in three sections of decreasing granularity from the beam pipe. The EM calorimeter uses an accordion shape for minimizing the dead-zones.

The Hadronic Calorimeter subsystem is actually three calorimeters. The outer hadronic calorimeter (TileCal) uses scintillating tiles as sampling medium (thus the name) and steel as absorber. It is composed of a central and two extended barrels, and provides maximum radial depth while keeping costs low for ATLAS. The LAr hadronic end-cap calorimeter (HEC) consists of two wheels per end-cap, located behind the end-cap EM calorimeter. The LAr forward calorimeter (FCal) is approximately 10 interaction lengths deep and consists of three modules.

The calorimeters are designed to provide good containment of electromagnetic and hadronic showers respectively, which also limits punch-through into the muon detection system[63].

3.2 Boosted $X \to HH \to bb\tau\tau$ analysis

This chapter will describe the search of a Beyond the Standard Model heavy scalar resonance X (in the range 1–3 TeV) decaying into a Higgs boson pair HH, with one Higgs boson decaying into a $b\bar{b}$ pair and the other decaying into a $\tau^+\tau^-$ pair, with both decays in the boosted regime (which means $p_T \gg m$ for the Higgs pair), and with both τ particles decay into hadrons (more succinctly, $X \to HH \to b\bar{b}\tau^+_{\text{had.}}\tau^-_{\text{had}}$). The search was performed using the full Run 2 dataset, with an integrated luminosity of 139 fb^{-1}. The analysis was carried out by members of the ATLAS collaboration during 2019 and 2020 and has resulted in an internal analysis note and a publication[43].

This publication presents a novel method for boosted di-τ reconstruction, and makes use of a data driven method for estimating the multi-jet background, and simulations for other backgrounds. In the next section, 3.3, we will present an alternative method of studying the multi-jet background using two-dimensional sidebands. As the object reconstruction and event selection remain almost identical in both analyses, we will present them in detail in sections 3.2.2 and 3.2.3, and in the following chapter we will only make mentions to the departures from these descriptions. We will not discuss the results obtained in this analysis as pertains to the search for Higgs pairs or a heavy scalar resonance, as we are interested only in the background estimation method.

As mentioned in section 1.3.1 in the introduction, the $b\bar{b}\tau^+\tau^-$ decay channel has a 7.4% of the branching ratio for Higgs boson pairs; the third highest, after $b\bar{b}b\bar{b}$ (33%) and $b\bar{b}W^+W^-$ (25%), but has a much cleaner signature than both, with the reconstruction of

the W^+W^- Higgs being particularly difficult.

3.2.1 Simulated events

In addition to the Run 2 data, samples of Monte Carlo (MC) simulated events are used to perform calibrations as well as to model the signal and SM background processes.

The simulated signal of a heavy spin-0 resonance via gluon fusion and subsequent decay into a Higgs boson pair $ggF \rightarrow X \rightarrow HH$ was simulated using MadGraph5_aMC v2.6.1. The parton shower and hadronization was modeled using Herwig v7.1.3. A total of 8 resonance masses are used in this thesis, corresponding to 1, 1.2, 1.4, 1.6, 1.8, 2.0, 2.5 and 3.0 TeV, although the analysis group later expanded this number to 9 for the publication.

For the backgorund simulations, a number of methods were used. For W and Z boson production in association with jets (V + jets), Sherpa v2.2.1 was used. For diboson processes, WW, WZ and ZZ, the Powheg-Box v2 generator was used, then interfaced with Pythia v8.186 for modelling the parton shower, hadronization and underlying event. Finally, single-top-quark processes and $t\bar{t}$ events were simulated using Powheg-Box v2 and interfaced using Pythia v8.230. Full details including the *tunes* (highly specific program configurations) used, and the extensive bibliography justifying the software choices, can be found in the full publication [43].

3.2.2 Object reconstruction and identification

Both $H \rightarrow b\bar{b}$ and $H \rightarrow \tau^+\tau^-$ decays are reconstructed separately. The further decay of these four particles should give rise to four separate jets, however in the boosted regime each pair is merged due to overlap, so it is more convenient to reconstruct the decays as two larger jets containing two thin subjets and use identification methods specific to each regime, as introduced in section 1.3.2.

Di-τ jets

Typically, a τ-lepton decaying hadronically will produce a neutrino and, most of the time, either one or three charged pions (known as *one-prong* or *three-prong* decays, 72% and 22% of all cases respectively), and a variable number of neutral pions. Together, the charged and neutral hadrons originating from this decay would be reconstructed and identified as a $\tau_{\text{had-vis}}$ object according to ATLAS standard procedure [64]. However, in the search for a high-mass resonance, a large fraction of the $\tau^+\tau^-$ pairs would fail this test, in part due to their high collimation (separation $\Delta R < 0.4$ between the visible jets), as is shown by using a simulation for the decay of a 2 TeV particle, shown in figure 3.6. Therefore a new reconstruction procedure was developed by the analysis team for these types of highly collimated pairs, one specially focused in identifying a di-τ dijet [65].

Starting with an untrimmed large-radius jet ($R = 1.0$) with $p_T > 300$ GeV, the constituents are reclustered into anti-k_t subjets with $R = 0.2$. The seed is required to have at least two sub-jets, and the pair leading in p_T will be used to construct the di-τ system, both of which must have $p_T > 10$ GeV. The same selection criteria for track selection and track-vertex matching as for the standard procedure [64], and it is required that both leading subjets have at least one associated track. Passing these, it is also required for di-τs to have either one or three tracks matched in each subjet, with a distance $\Delta R < 0.1$ from its axis. A visual representation of a di-τ jet can be seen in figure 3.5. The efficiency improvements over the standard "resolved" selection can be seen in figure 3.6 where a

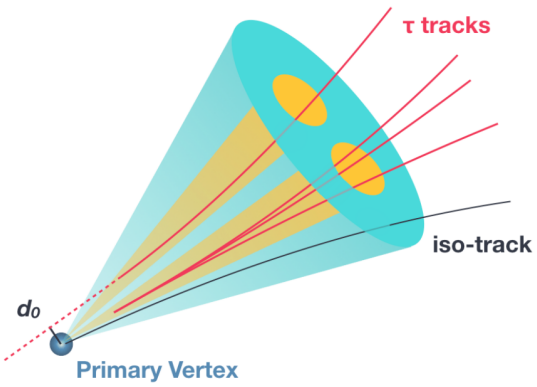


Figure 3.5: Schematic representation of a di-τ system consisting in two jets of $R = 0.2$ embedded in a large-radius jet of $R = 1.0$. Examples of one-prong and three-prong decay are represented in red, along with an external track denominated isotrack. Source: ref. [43]

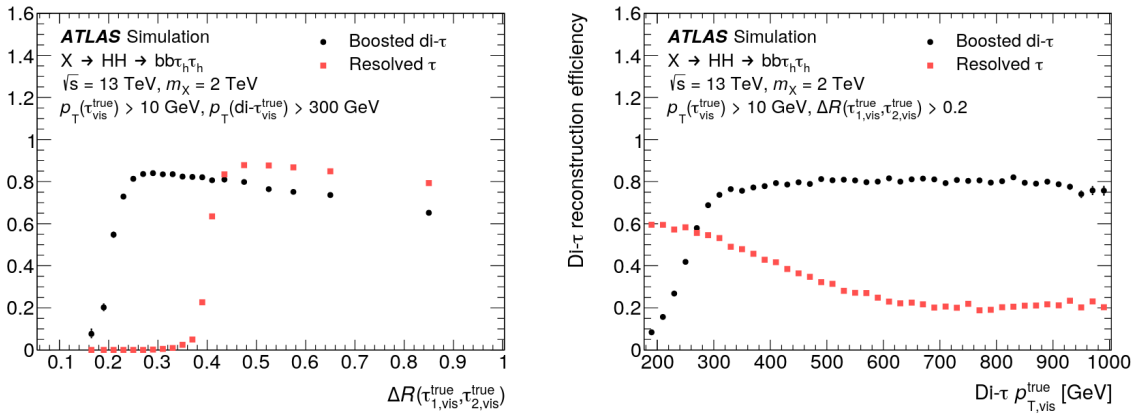
simulation was used. As expected di-τs with lower ΔR can now be reconstructed, with only a small loss of efficiency at higher ΔR values; however the reconstruction efficiency is greatly enhanced for high p_T di-τs, which proves particularly useful for finding the decay products of a heavy resonance.

In parallel to this, a separate identification using multivariate analysis is made, in order to reduce the background from quark and gluon initiated jets. A BDT is trained with information from calorimeter clusters, tracks, and vertices, using the AdaBoost algorithm [66], in the TMVA toolkit [67]. The di-τ objects used for training are taken from graviton $G \to hh \to (\tau^+\tau^-)(\tau^+\tau^-)$ simulations, with samples of mass of 1.5, 1.75, 2, 2.25 and 2.5 TeV combined for the signal. The spin of the resonance is found to have no impact on the di-τ identification. The BDT output score is used to define a cut in the di-τ selection, and is used to separate the events in different regions which will be used for the background estimation, as will be explained in 3.2.3.

bb–jets

For the reconstruction of bb–jets, candidates are trimmed [68] by reclustering the constituents of the original large-R jet into smaller $R = 0.2$ sub-jets using the k_t algorithm [51], and removing those sub-jets with only a small fraction of the transverse momentum of the untrimmed large-R jet. The trimmed jets are then calibrated using factors derived from simulation, and a $p_t > 300$ GeV cut is made on the trimmed large-R jets, whose mass is now calculated using the combined mass technique, taking information from the tracking and the calorimeter [69].

In order to determine the b-hadron (or $\bar{b}$) content of the large-R jets, track-jets within it undergo a flavor-determining algorithm in a process called b-tagging. In this analysis, track-jets with a variable radius parameter $R(p_T) = \rho/p_T$ [70] are reconstructed from inner-detector tracks using the anti-k_t algorithm. The factor ρ determines how fast the effective size of the jet decreases with its transverse momentum, and lower and upper cut-offs in $R(p_T)$ of $R_{\text{min}} = 0.02$ and $R_{\text{max}} = 0.4$ are used. These track-jets are then matched to an untrimmed large-R jet with the ghost-association method [52], and only those with $p_T > 10$ GeV and $|\eta| < 2.5$ are considered. The MV2c10 algorithm [71] combines information from various b-jet identification algorithms into a boosted decision tree (BDT) which is trained



(a) (b)

Figure 3.6: Efficiency of di-τ reconstruction using the novel boosted di-τ method (circles) compared to the standard $\tau_{\text{had-vis}}$ reconstruction (squares). Improvements can be seen at (a) highly collimated regimes and (b) high system p_T. Simulation of a 2 TeV resonance is used, where τ leptons decay hadronically. Source: ref. [43]

to discriminate b-jets from a simulation background composed of 93% light-flavored jets and 7% c-jets. A track-jet is labeled as b-tagged based on the output score of the MV2c10 algorithm.

As a result, the trimmed large-R jets we fed into the system are tagged as either 0-tag (containing zero identified b-hadrons), 1-tag (containing exactly one) or 2-tag, (containing exactly two b hadrons). Dijets with more than 2-tags are discarded in this analysis.

3.2.3 Event selection and data driven approach to background estimation

Although further cut selections are applied to the data in order to remove background events, some of this data removed by cuts is stored and used to estimate the multijet background and apply corrections to the efficiency of the di-τ tagging process. For this end, the data is separated in different regions according to the cuts they pass. Table 3.1 details most of the cuts used in the analysis (for the full table see ref. [43]).

Table 3.1: Definition of regions and their uses. Q is the product of the electric charge of the two leading di-τ subjets (OS is opposite sign and SS same sign). $N_{b\text{-tags}}$ is the number of b-tagged track-jets in the large-R jet. The azimuthal angle $\Delta\phi$ between the di-τ and the missing transverse momentum. Cut m_J affects the large-R jet mass. Source: [43]

Name	Usage	Q	$N_{b\text{-tags}}$	$\Delta\phi$	m_J [GeV]
FF SS	Fake-factor computation	+1	0	-	-
MJ OS 0-tag	Closure test multi-jet 0-b-tag	-1	0	>1	-
MJ OS 1-tag	Closure test multi-jet 1-b-tag	-1	0	>1	50-60 or > 160
$Z_{\tau\tau}$ 0-tag	Di-τ tagger correction	-1	0	<1	-
$Z_{\tau\tau}$ 1-tag	Di-τ tagger closure test	-1	1	<1	50-60 or > 160
HH SR	$X \to HH \to b\bar{b}\tau^+\tau^-$ search	-1	2	<1	> 60 and < 160

The di-τ objects are required to have no more than three subjets, with the leading and

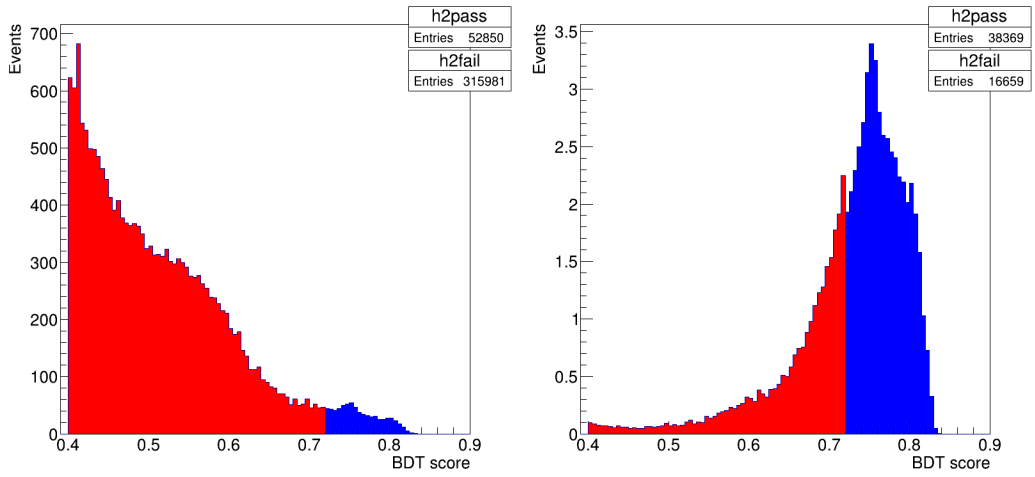


Figure 3.7: BDT score for simulated background (left) and a scalar resonance of mass 2 TeV (right). In red, the events which are below the 0.72 threshold, and in blue those above. These correspond to all simulated events, with all other cuts removed, and are weighted by their cross sections. Events with BDT score below 0.4 do not pass preselection.

subleading having a separation of no more than $\Delta R = 0.8$ so that both are contained in a large-radius jet. Also the electric charge of the two leading subjets must satisfy $Q = q_{\text{lead}} \times q_{\text{sub-lead}} = \pm 1$, with the charge of a subjet defined as the sum of the charges of the associated tracks. Both the di-τ and the bb-jet must have $p_T > 300$ GeV and $|\eta| < 2.0$, excluding the barrel-endcap transition region.

The event must also have $E_T^{\text{miss}} > 10$ GeV, as neutrinos are expected in the decay and also a clear direction is needed for the classification of the event. The angle between the E_T^{miss} and the di-τ is used to separate regions in the analysis. Finally a separation of $\Delta R > 1.0$ between bb-jets and di-τ is required. If there is more than one bb-jet satisfying all of these conditions, the one with highest p_T is used.

The events in which both subjets of the di-τ have the same charge ($Q = 1$) define the region in which *fake factors* binned in p_T are calculated

$$\text{FF} = \frac{N_{\text{pass}}}{N_{\text{loose}}} \tag{3.1}$$

depending on whether the di-τ passed the full BDT-based identification requirement, or passed a much looser version of it (denominator), corresponding to scores of 0.72 and 0.4 respectively. To illustrate the efficiency of the BDT score selection, the scores obtained for simulations of background and a signal of mass 2 TeV are shown in figure 3.7. The fake factors are then used to estimate the yield of events with misidentified di-τ objects $N_{\text{mis-ID}}$ in regions closer to the signal region where the di-τ object has subjets of opposite sign ($Q = -1$). The estimation uses the number of events which failed the BDT test N_{fail}, with

$$N_{\text{mis-ID}} = N_{\text{fail}} \times \text{FF}. \tag{3.2}$$

The regions labeled $Z_{\tau\tau}$ in table 3.1 are used to correct the efficiency of the di-τ tagger, they represent a region enriched in di-τ objects originated in $Z \to \tau^+\tau-$ decays. The details of how this correction is applied are not necessary for the following, but they can be found in the original paper, ref. [43].

3.3 Data driven estimation of QCD background using sidebands

The fake factor based method of background estimation presented in the previous chapter that the ATLAS analysis team decided to use is not the only method for estimating the multi-jet background. A method based on 2 dimensional sidebands was used in the past for double Higgs searches in the $bbbb$ decay channel using Run 1 data in both boosted and resolved regimes [35]. In this chapter we will adapt such a method for the $bb\tau\tau$ decay channel.

A sideband method of background estimation consists in the extrapolation of the data points from a signal-free region into the region in which signal is expected (signal region). The chosen signal-free region (known as sideband) must be close enough to the signal region, to accurately model the background within it. Data driven methods such as this might be preferable if the simulations do not describe the data sufficiently well, as is the case in multi-jet background.

First we will need to define the signal region, for which we will search for an optimal two-dimensional mass window (2DMW) in the bb and $\tau\tau$ 2D mass space, detailed in section 3.3.1. We will do so using simulations for the different masses of scalar resonances and find a one-size-fits-all cut that will define our signal region.

In the following section 3.3.2, we will define the sideband and present the results.

3.3.1 2D mass window cut

The idea of defining a mass window for a signal region is to create a compromise between signal efficiency and background rejection. The parameter we are interested in optimizing is the significance

$$Z = \sqrt{2(s+b)\log\left(1+\frac{s}{b}\right) - 2s}, \tag{3.3}$$

this formula works better than the more common $s/\sqrt{b}$ in low background situations. Here s is the number of signal events and b the number of background events. The background events consist in the MC simulations used in the previous chapter, subsection 3.2.1, in addition to the *fakes* multi-jet background obtained from data, as described in section 3.2.3.

The 2DMW was obtained in two stages. First, a separate optimization was made for each mass signal (1 TeV, 1.2 TeV, etc.) completely independent of each other, resulting in 8 separate 2DMW regions. Then, taking these results, a single 2DMW that would work for all signals was chosen. These two steps are detailed in the next subsections.

Per signal optimization

The starting point of the analysis is to take a look at the background and signal samples after all the main analysis cuts and selections on figure 3.8.

The analysis framework used in the previous chapter delivers a set of histograms as output, but this is restrictive as information is lost during the binning process. This is specially a problem when trying to optimize for a smooth 2d function using blocky bins; the unbinned data falling within or outside the defined region might depend on the binning, therefore unbinned data was used unless absolutely necessary. When working with binned data, 5 GeV bins were used for compatibility with the analysis of the previous chapter. The binned background samples can be seen in figure 3.8.

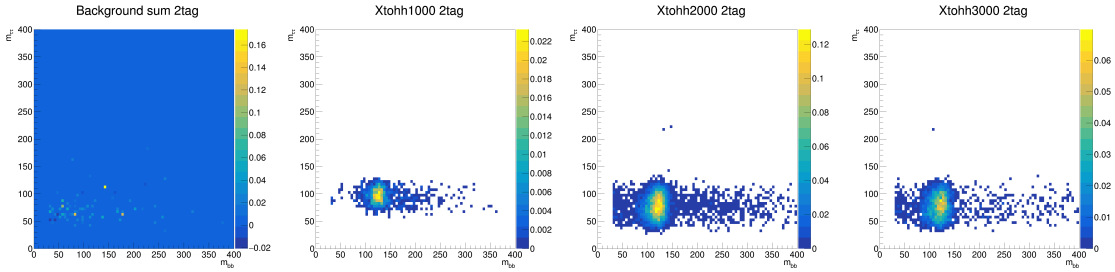
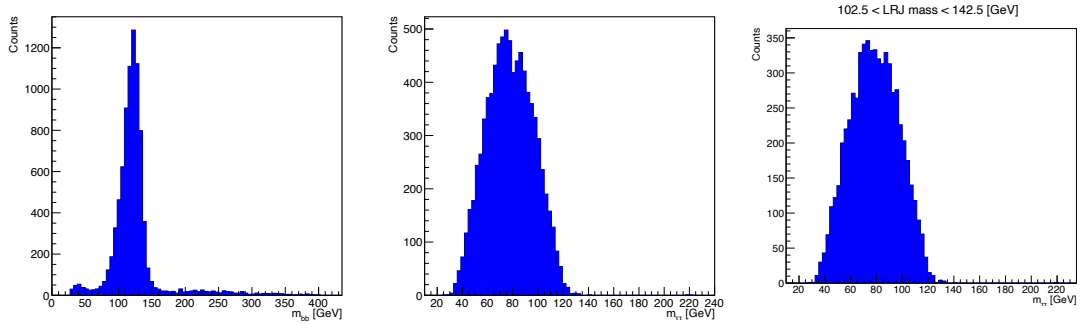


Figure 3.8: Sum of all backgrounds and signals for X mass of 1, 2 and 3 TeV for 2-tag cut and $\Delta\phi < 1$, where $\Delta\phi$ is the angle difference between the di-τ and the missing transverse momentum. Masses are on GeV

For m_{bb} the mass center was chosen as the bin center in the maximum value in the mass distribution, as a peak can be seen unambiguously in figure 3.9 (a). For the $m_{\tau\tau}$ center, as two peaks appear when integrating over m_{bb} due to the long tails present in m_{bb}. The tails are removed using a 20 GeV cut around the m_{bb} maximum value resulting in a cleaner shape for the $m_{\tau\tau}$ shown in 3.9 (c). The $m_{\tau\tau}$ center is obtained by taking the mean value, with the results shown in 3.10 (b). As a result of obtaining the m_{bb} center from bin centers, discrete values are observed in subfigure (a) of 3.10, while a continuous shape in subfigure (b) of 3.10 is obtained for $m_{\tau\tau}$ as it is the result of a mean.



(a) (b) (c)

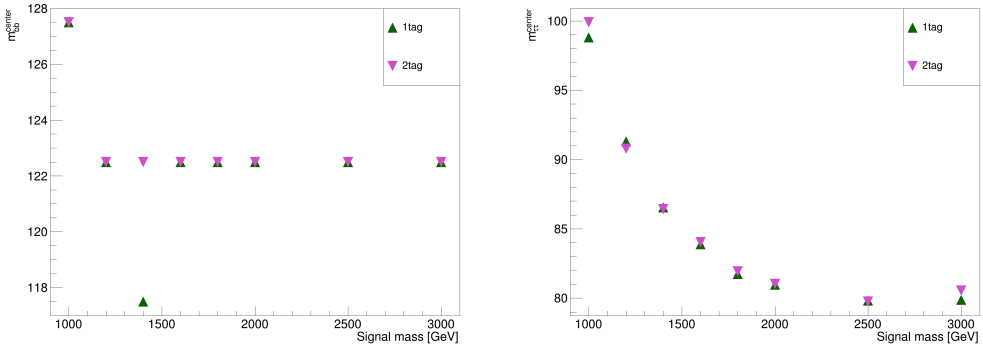
Figure 3.9: Integrated mass distributions for a 2 TeV signal in the (a) bb and (b) and (c) $\tau\tau$ systems. A cut of 20 GeV around the chosen m_{bb} center of 122.5 GeV is applied in (c) to remove the long tails visible in (a)

A selection of the shape of the 2DMW must be made. This is a problem that doesn't arise at all for a 1D mass window, where the window is simply defined by two points in the mass axis. The starting point considered was the one used in the *bbbb* analysis [35],

$$X_{HH} = \sqrt{\left(\frac{m_{bb}^{\text{center}} - m_{bb}}{\sigma(m_{bb})}\right)^2 + \left(\frac{m_{\tau\tau}^{\text{center}} - m_{\tau\tau}}{\sigma(m_{\tau\tau})}\right)^2} \tag{3.4}$$

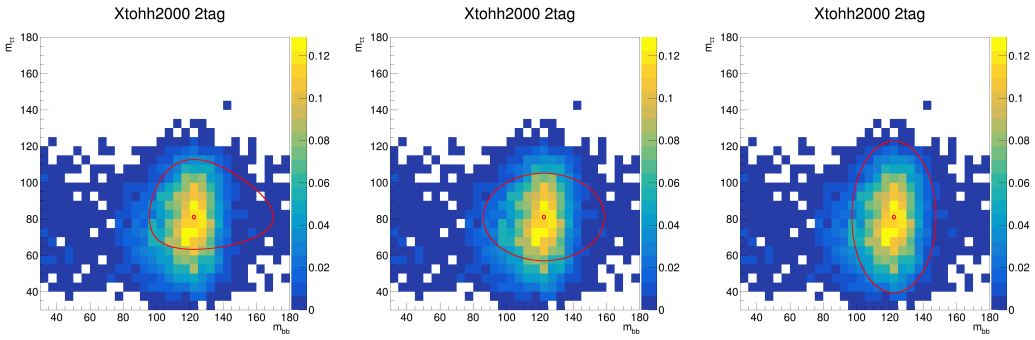
Here m_{bb} and $m_{\tau\tau}$ are the mass variables, whereas m_{bb}^{center}, $m_{\tau\tau}^{\text{center}}$ are the mass centers already obtained for each signal, and X_{HH} is the radius variable yet to be determined. Also $\sigma(m_{xx})$ is the uncertainty of the mass in either bb or $\tau\tau$. The uncertainty can in principle be a complicated function of the mass, in the *bbbb* analysis it was simply $\sigma(m_{xx}) = 0.1 \times m_{xx}$.

Three shapes shown in figure 3.11 were tested. The first one is the same shape as in the *bbbb* analysis, which doesn't look like a good fit except for the very low signal masses,



(a) (b)

Figure 3.10: Mass centers for each MC signal in the (a) m_{bb} space and (b) $m_{\tau\tau}$ space. Both 1-btag and 2-btag reconstructed samples are shown.



(a) (b) (c)

Figure 3.11: Shapes consideref over a 2 TeV signal for 2-btag. (a) shape used in the *bbbb* publication (b) horizontal ellipse (c) vertical ellipse. The center for each shape is fixed as is detailed in the text.

as it doesn't translate well for a channel with a large spread in the mass axes. A vertical ellipse of equation (3.5) was found to be the best fit, and finally a horizontal ellipse was used as a control. The results of the significance having optimized all other variables is shown in 3.12. The vertical ellipse gives the largest expected significance except for the 1 TeV mass. The low p_T cuts for di-τ jets affect disproportionately the lower signals. This signal is an outlier sample, as it also has the lowest acceptance times efficiency, and will weight less at the time of making a decision for a single 2DMW that fits all signals.

$$X_{HH}^{\text{V. ellipse}} = \sqrt{\left(\frac{m_{bb}^{\text{center}} - m_{bb}}{0.1 \times m_{bb}^{\text{center}}}\right)^2 + \left(\frac{m_{\tau\tau}^{\text{center}} - m_{\tau\tau}}{0.25 \times m_{\tau\tau}^{\text{center}}}\right)^2} \tag{3.5}$$

The radius optimization is obtained by sweeping different values of X_{HH} from $X_{HH} =$ 0.5 to 5.0 in intervals of size 0.01, and keeping the value which delivers the largest significance. Convergence was ensured by inspecting the values of Z and noting that they peak within the sweeping interval, and shown in figure 3.13 for three different sample signals. The results of the optimized regions are detailed in table 3.2.

The results in table 3.2 represent the maximum possible values for Z for each signal and the parameters needed to obtain them. One could in principle do 8 different analyses,

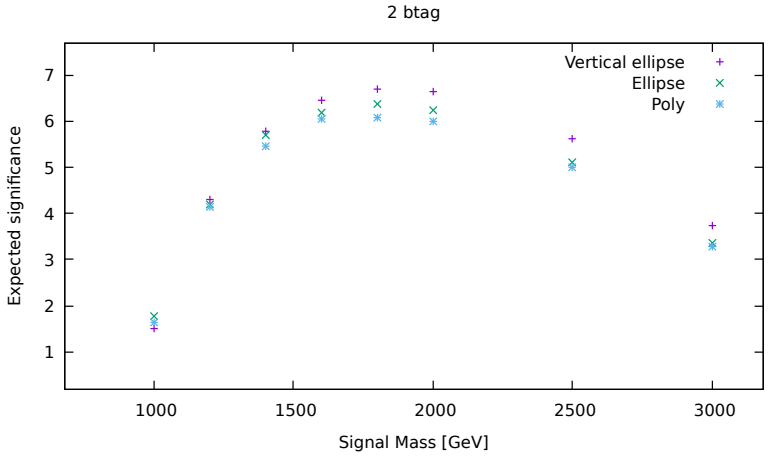


Figure 3.12: Maximum significance for each shape and signal when all other parameters are optimized.

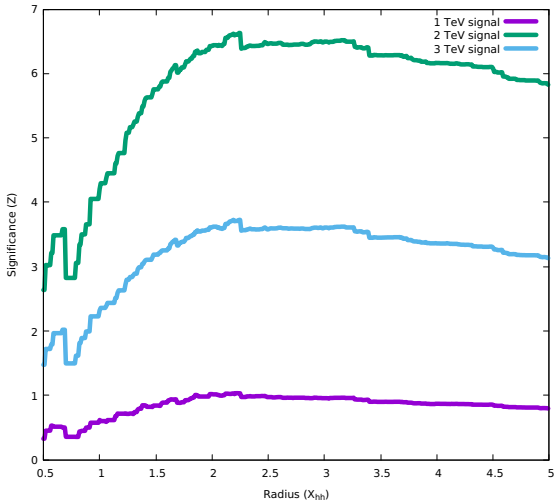


Figure 3.13: Values for the significance Z for set values of the mass centers for a vertical ellipse shape. Convergence is ensured by obtaining a peak within the sweeping region

Table 3.2: Results per signal for 2-btag samples in the signal region, for vertical ellipse shape

Signal mass [GeV]	X_{HH}	Z	m_{bb}^{center}	$m_{\tau\tau}^{\text{center}}$
1000	1.04	1.406	127.50	99.20
1200	1.29	4.100	122.50	90.13
1400	1.74	5.303	122.50	86.01
1600	2.09	5.805	122.50	83.69
1800	2.08	6.003	122.50	81.76
2000	2.07	5.850	122.50	81.07
2500	2.28	4.982	122.50	79.99
3000	2.26	3.388	122.50	80.77

each optimized for a possible signal mass value, however in practice it is better to pick a single 2DMW that is a good compromise between these and focus on optimizing other parts of the analysis.

Single 2DMW

Now we would like to obtain one 2DMW which still has a good sensitivity for all possible signals, this is, which doesn't deviate much from the optimal values already found.

For the fixed m_{bb} mass center, it is clear from figure 3.10 that 122.5 GeV should be chosen as for the 2 b-tag signal it is the best value for all resonance masses except the 1 TeV sample. In the case of the $m_{\tau\tau}$ center, figure 3.10 is not so clear cut. The results in table 3.2 for the significance show that the analysis is intrinsically most sensitive to a 1.8 TeV signal, with a sensitivity four times greater than for the 1 TeV sample even under the best conditions. Therefore a $m_{\tau\tau}$ center of 81 GeV was decided intended to maintain high sensitivity at higher masses sacrificing the lower masses. The use of two working points for the $m_{\tau\tau}$ mass center was considered but it was found not to have much impact on the overall significance or expected limits.

For the radius optimization a simple equation to account for deviations D from the best possible significances Z_i^{best} shown in 3.2 is defined as

$$D(X_{HH}) = \sum_{i=\text{signals}} \left(Z_i^{\text{best}} - Z_i(X_{HH}) \right)^2 \tag{3.6}$$

By minimizing the deviation for the obtained values of Z_i using the unified criteria for a single 2DMW it was found that $Z = 2.06$ is the value that minimizes this deviation. The results are shown in table 3.3.

Table 3.3: Results for a single (fixed) 2DMW of $X_{HH} = 2.06$, vertical shape and $m_{bb}, m_{\tau\tau}$ centers at 122.5, 81.0 GeV

Signal mass [GeV]	Z (signal optimized)	Z (fixed)	Reduction in sensitivity (%)
1000	1.406	0.981	30.221
1200	4.100	3.818	06.884
1400	5.303	5.211	01.735
1600	5.805	5.751	00.933
1800	6.003	5.979	00.396
2000	5.850	5.839	00.186
2500	4.982	4.972	00.201
3000	3.388	3.374	00.432

With this, the final 2DMW cut to be applied to event selection is

$$\boxed{\sqrt{\left(\frac{122.5 - m_{bb}}{0.1 \times 122.5}\right)^2 + \left(\frac{81.0 - m_{\tau\tau}}{0.25 \times 81.0}\right)^2} < 2.06} \tag{3.7}$$

which defines the optimal signal region for this analysis.

Expected and observed limits

To test the expected exclusion limits that might be achievable in the signal region, single-bin counting experiments are performed for each signal mass hypothesis, based on a profile-likelihood fit [72] implemented in the HistFitter framework [73]. The 95% confidence levels

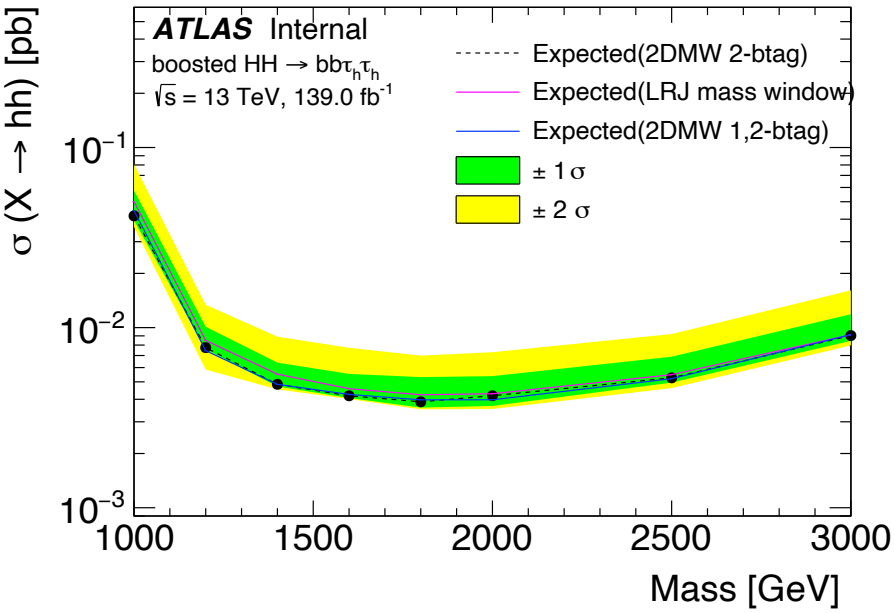


Figure 3.14: Expected limits for different signal masses for the presented 2DMW (dots and dotted line) and the mass window used in the published paper (pink)

upper limits on the $X \to HH$ cross section are calculated using the CL_s method [74]. Only the expected event yields are presented, and due to their low size, the distribution of the test statistic is sampled by generating pesudo-experiments. For each signal mass, 10000 background-only and 10000 signal+background pseudo-datasets are generated, obtaining the exclusion limits by scanning the signal strength and interpolating to find the value at which $CL_s = 0.05$. This is shown in figure 3.14 for different cuts corresponding to different definitions of the signal region. The one labeled LRJ mass window (pink line) is the one which ended up being used in the main analysis, while the 2DMW 2-btag (dotted line, with point markers) is the one obtained using the 2DMW described in the previous subsections. Table 3.4 shows the data plotted in the figure 3.14. It can be seen that the 2DMW presented is an improvement in the expected limits in the complete mass range, as it would exclude a larger region and place more stringent limits in the cross section.

There are a few reasons why this improvement was not implemented in the final analysis. First was the timing. By the time the 2DMW analysis was completed, the analysis group was already far advanced in the publication process, having already contacted the editorial board. Implementing such a non-straightforward change in the final signal region this late without further testing would not have been ideal, as if some error was discovered there wouldn't have been time to fix it. Second, the improvements, despite spanning the whole mass range, are not tremendously significant. Even at this stage the exclusion limits would not have been competitive with the other decay channels. Third, a study including systematic uncertainties was never carried out for this 2DMW, and including this would have further delayed the publication process. Despite our 2DMW cut not being included in the final publication, we will continue using it as our signal region in the following section.

Table 3.4: Expected limits for different signal masses for the presented 2DMW and the mass window used in the published paper, along with the limits using no extra mass-window cuts in the final event selection

m_X (GeV)	2DMW (fb)	LRJ MW (fb)	No cuts (fb)
1000	41.73	50.10	55.84
1200	7.74	8.49	9.16
1400	4.84	5.52	5.72
1600	4.18	4.58	4.97
1800	3.87	4.24	4.69
2000	4.19	4.29	4.70
2500	5.26	5.48	5.98
3000	9.03	9.13	9.82

3.3.2 Sideband and validation regions

The sideband region must be close enough to the signal region to accurately model its background. The shape of the SM background is assumed to be known, and its yield is derived from the sideband via extrapolation. An intermediate region may be placed in between the sideband and signal regions in order to validate this normalization.

In the case of our analysis, we will use the two dimensional signal region obtained in the previous section and create two dimensional validation and sideband regions around it. One problem that this analysis faces is the lack of sufficient statistics. The cuts performed in the main boosted $bb\tau\tau$ analysis are too tight, leaving only two data points in the signal region and insufficient data to obtain meaningful shapes in all regions.

There are a few ways in which more data points can be obtained in the regions we are interested in. The first one is expanding our regions of interest; we have the freedom to choose the size of our sideband region, provided the sideband remains reasonably close to the signal region. Widening the sideband doesn't increase the data in each bin however. The sideband and validation regions were thus selected as

Sideband:

$$\sqrt{\left(\frac{135.0 - m_{bb}}{0.25 \cdot 122.5 \cdot 2.06}\right)^2 + \left(\frac{81.0 - m_{\tau\tau}}{0.35 \cdot 81.0 \cdot 2.06}\right)^2} < 1.0 \tag{3.8}$$

Validation

$$\sqrt{\left(\frac{122.5 - m_{bb}}{0.15 \cdot 122.5 \cdot 2.06}\right)^2 + \left(\frac{81.0 - m_{\tau\tau}}{0.3 \cdot 81.0 \cdot 2.06}\right)^2} < 1.0 \tag{3.9}$$

Their shape is illustrated in figure 3.15

A Second way in which data points available may increase is by relaxing the cuts. It may be that in some cases the cuts are too tight, and a new analysis may benefit from identifying the points in which a lot of signal is cut. In the acceptance times efficiency plot 3.16, we can see that the most significant loss of signal events occurs in the di-τ selection step. Two cuts will be relaxed. First the 300 GeV cut for the di-τ jet p_T; the effects of this cut in the simulated background and simulated signal (described in 3.2.1) can be seen in figure 3.17. The second is a 50 GeV cut on the subleading-track within the di-jet p_T, the effects of which on simulated signal and background can be seen in figure 3.18.

Both of these cuts were found to carry out of the analysis an important fraction of signal events, with the subleading p_T cut not even removing much more background in the process, so both of these cuts were removed. It is important to note that despite

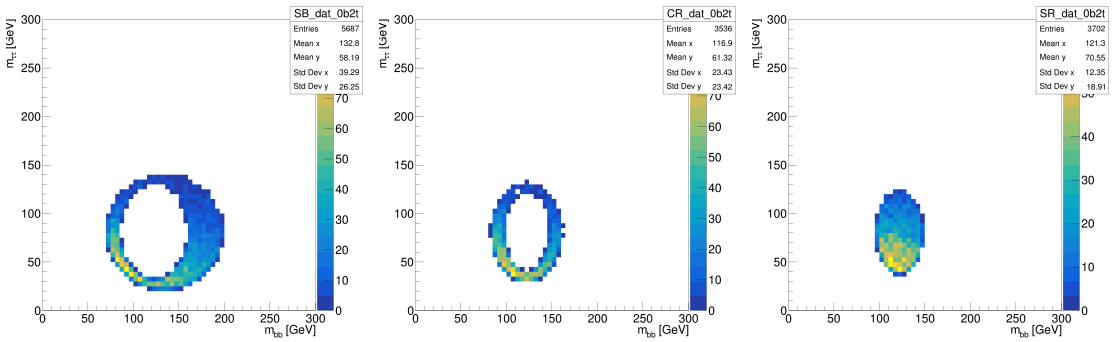


Figure 3.15: From left to right: sideband, validation, and signal regions

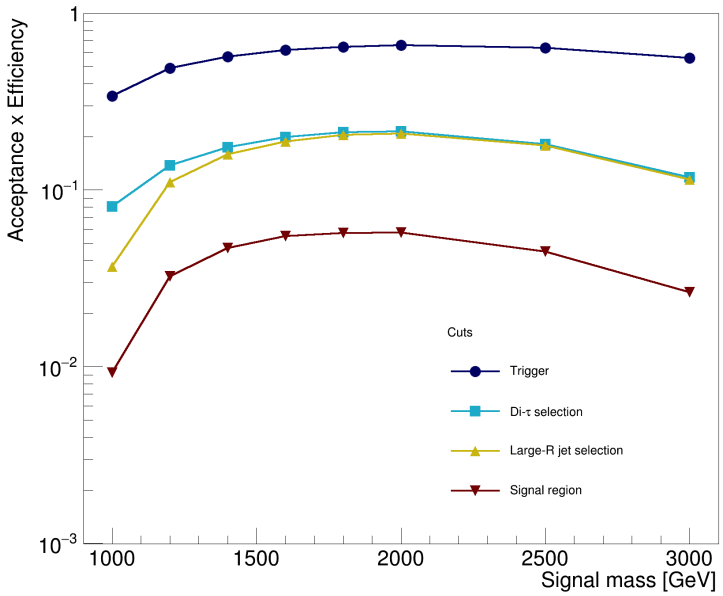


Figure 3.16: Acceptance times efficiency of the main cuts in the analysis

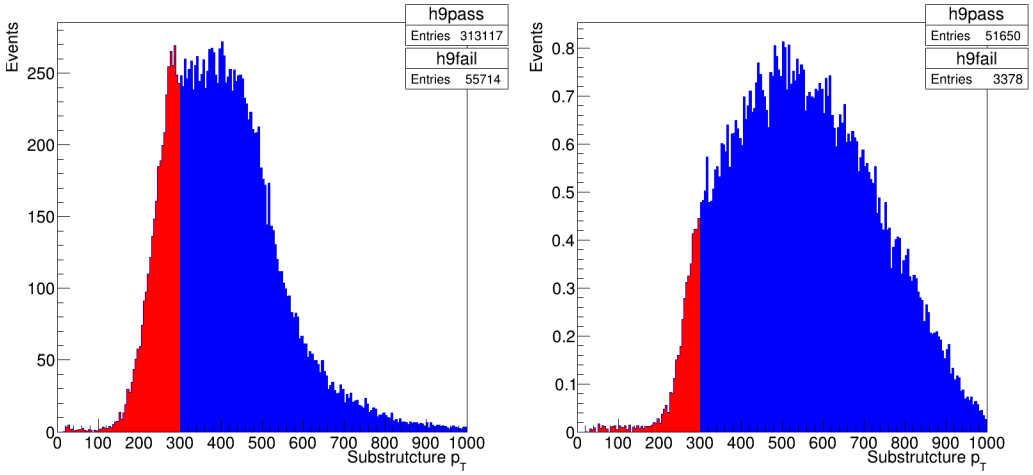


Figure 3.17: Cut on the di-tau p_T on Monte Carlo samples. On the left plot, simulated sum of backgrounds. On the right, simulated signal. The samples are weighted. In red are the events being cut with the default selection and in blue, the events passing, with all other cuts turned off.

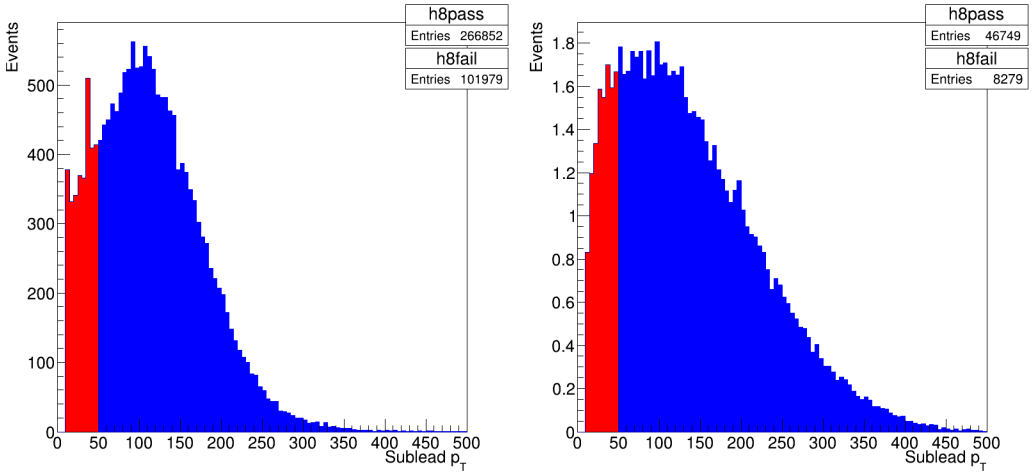


Figure 3.18: Cut on the di-tau p_T on Monte Carlo samples. On the left plot, simulated sum of backgrounds. On the right, simulated signal. The samples are weighted. In red are the events that being cut with the default selection and in blue, the events passing, with all other cuts turned off.

justifying the removal of these cuts using the MC samples, the background estimation method described in this section is completely data driven and uses no MC samples in its implementation.

The final statistics for each region are detailed in table 3.5

Table 3.5: Number of events in all kinematic and lower-tagged regions

	2b2t	2b0t	1b2t	1b0t	0b2t
SB	26	44	193	530	1883
CR	18	29	159	326	1201
SR	blinded	37	147	355	1284

Lower-tagged samples and results

The signal we are looking for is composed of one large-R jet containing two b quarks, and a di-tau jet. We will refer to the events in which the analysis code finds such jets as $2b2\tau$ events. In addition to these data samples, we can consider events in which the analysis code found a large-R jet containing two b quarks, but failed to identify a jet as corresponding to a di-tau and call this a $2b0\tau$ event. In a similar fashion we can consider events which failed b-tagging but where a di-tau jet is successfully reconstructed, in which case we talk about a $0b2\tau$ event. To complete the picture, we consider the cases where *one* b-quark is identified in the large-R jet and a di-tau was or was not reconstructed and refer to them as $1b2\tau$ and $1b0\tau$ events. We will refer to these collectively as *lower-tagged samples*. This is an adaptation of the lower-tagged regions used in the *bbbb* analysis[35], in which lower b-tagged jets were used.

The key point to consider is that for studying the mass distributions of the large-R jets in the m_{bb} axis the lower-tagged sample $2b0\tau$ should work reasonably well, as only the $2b$ part will be considered effectively integrating over the di-tau jet. Analogously, the $1b2\tau$ should be very effective for modeling the $m_{\tau\tau}$ distribution. These samples can also be supplemented with the $1b0\tau$ and $0b2\tau$ samples respectively, in order to garner more statistics, while needing at most minor corrections to account kinematic differences.

A maximum likelihood fit is performed in the sideband region on the m_{bb} projection of the $2b2\tau$ sample using as linear combination of the lower tagged samples $2b0\tau$ and $1b0\tau$, as shown in equation 3.10.

$$N^{SB}_{2b2\tau}(m_{bb}) = \mu_{2b0\tau} N^{SB}_{2b0\tau}(m_{bb}) + \mu_{1b0\tau} N^{SB}_{1b0\tau}(m_{bb}). \tag{3.10}$$

In this and the following equations, N^r_s represents the yield of sample s in region r. The weights μ_s are the factors to be extracted from the fit, and will be assumed to carry from the sideband region into the signal region, after validating them in the validation region. The mass in parenthesis reminds us which distribution we are modeling.

After the μ factors are obtained from the sideband region fit, we will obtain our guesses for the expected yield in the control and signal regions with

$$N^{CR,\ \text{exp.}}_{2b2\tau}(m_{bb}) = \mu_{2b0\tau} N^{CR}_{2b0\tau}(m_{bb}) + \mu_{1b0\tau} N^{CR}_{1b0\tau}(m_{bb}), \tag{3.11}$$

$$N^{SR,\ \text{exp.}}_{2b2\tau}(m_{bb}) = \mu_{2b0\tau} N^{SR}_{2b0\tau}(m_{bb}) + \mu_{1b0\tau} N^{SR}_{1b0\tau}(m_{bb}), \tag{3.12}$$

where, we emphasize, the $\mu_{2b0\tau}$ and $\mu_{1b0\tau}$ factors appearing in equations (3.10), (3.11) and (3.12) are the same.

The result of the linear combination of m_{bb} distributions according to equation 3.10 is shown in figure 3.19

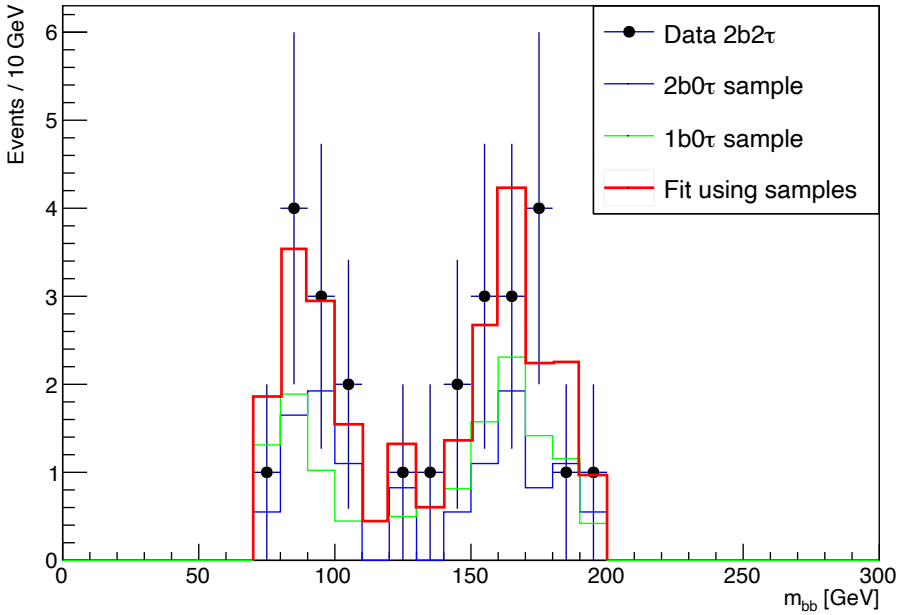


Figure 3.19: Fit in m_{bb} using lower tagged samples. The bins are 10 GeV wide

As can be seen, the fit adjusts within uncertainties to the data points. The values obtained for the scaling factors of equation (3.10) are

$$\mu_{2b0\tau} = 0.27 \pm 0.20, \qquad \mu_{1b0\tau} = 0.03 \pm 0.02. \tag{3.13}$$

By extrapolating these factors into the validation region using the values in table 3.5, we obtain a prediction of 16.52 ± 8.24 for the validation region yield. This value has a relative error of 8.22% when compared to the true value of 18, which we will consider sufficient taking into account the low statistics we have available.

Additionally to the obtained yield, we can perform a validation using the visible double Higgs mass m_{HH}, which is the invariant mass of the sum of the four-momenta of the di-τ jet and the bb jet. We obtain the distribution shown in 3.20. The lower end of the spectrum is poorly modeled, just as we had anticipated with the sensitivity measurements obtained in the previous sections.

Our final result then will be the extrapolation of the fit into the signal region using the same fit values (3.13) in equation (3.12) to obtain an expected yield of

$$N_{2b2\tau}^{SR,\,\text{exp.}} = 19.48 \pm 9.80. \tag{3.14}$$

It's only at this point in the analysis when we will unblind the $2b2\tau$ bin for the signal region in 3.5, which contained a value of $N_{2b2\tau}^{SR} = 17$. Comparing this to our guess, we see that it had a relative error of 14.59%. The distribution of the fit compared to the signal region points is shown in figure 3.21.

Again, we see that the lower end of the spectrum, in particular the second bin is not representin well the data. A further improvement that was not tried in this work is the

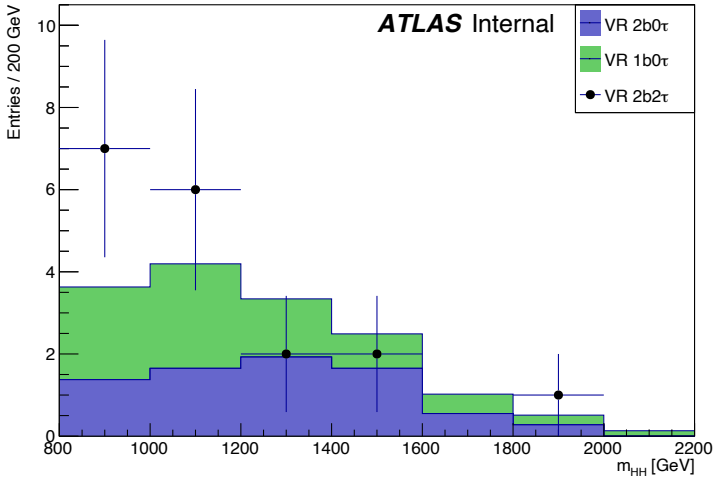


Figure 3.20: Visible double Higgs mass m_{HH} for the $2b2\tau$ in the validation region as modeled by the sum of the lower tagged samples $2b0\tau$ and $1b0\tau$

reweighting of the lower-tagged samples. The lower-tagged samples differ from the $2b2\tau$ region in the di-τ identification, and this identification process itself affects the kinematics of the event, as it involves requirements in p_T, to name one example. To compensate for this, the lower-tagged samples could be reweighted to mimic the distribution of $2b2\tau$ samples in one or more kinematic variables. This could work in redistributing kinematic distributions such as the visible double Higgs mass without heavily impacting the yield estimation. Another improvement that could be made is allowing the use of *some* MC samples to help modelling the data. For example the *bbbb* analysis team used $t\bar{t}$ MC samples to supplement their data driven method [35].

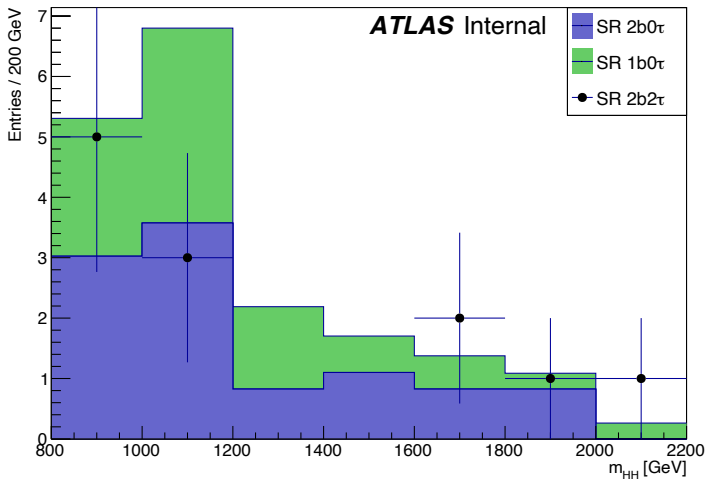


Figure 3.21: Visible double Higgs mass m_{HH} for the $2b2\tau$ in the control region as modeled by the sum of the lower tagged samples $2b0\tau$ and $1b0\tau$

Conclusions

A search for long-range structures in two-particle correlations was presented. By studying 1- and 2-dimensional correlation functions in azimuthal and rapidity space, a comparison between heavy ion collider experiments and the EG2 experiment at CLAS, a fixed target experiment at much lower energies. Despite the huge disparities between these systems, some common structures were observed, such as a far-side in azimuth correlation, extended in rapidity, such as usually associated with back-to-back pairs and momentum conservation. In particular a detailed study using Z_h bins was performed for the 1-dimensional azimuth variable, and a third order Fourier fit was applied, showing the apperance of non-neglible v_2 component. No near-side correlations consistent with the "ridge" phenomena observed in pp and pA collisions was observed.

The v_2 component was further analyzed using Q^2 bins, a variable which is more closely comparable to the transverse momentum p_T used in collider experiments. No dependence of v_2 on Q^2 was found in the range $1.0 < Q^2 < 5.0$. Future measurements using a similar system using a higher energy beam could expand this range, and such a system is the CLAS12 experiment.

For ATLAS, a full description of the boosted $X \to HH \to b\bar{b}\tau^+\tau^-$ search for a heavy scalar resonance was made, including the description of the decay channel, the method of reconstruction and identification of the di-jet objects, and the *fake factors* data driven approach to multijet background estimation. This ATLAS analysis is one of the first published double Higgs searches using the full Run 2 period (2015–2018) of the LHC.

Next, an optimization on the signal region using a two-dimensional mass window was presented. It was found that it improved the sensitivity by lowering the expected limits across all simulated heavy resonances considered in the range 1–3 TeV. This improved signal region was then used as the basis for the implementation of an alternative data driven approach to multijet estimation using two-dimensional sidebands, based on work done by the *bbbb* analysis team in a previous ATLAS analysis. The sideband method was validated in the yield of expected particles, but kinematic validation proved to be poor. A number of upgrades to this background estimation method were discussed, such as kinematic reweighting of the lower-tagged samples, or the use of Monte Carlo simulations to complement the data.

Bibliography

Bibliography

[1] J.J. Thomson. "Cathode Rays". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 44.269 (1897), pp. 293–316. DOI: 10.1080/14786449708621070. URL: https://doi.org/10.1080/14786449708621070.

[2] N. Bohr. "I. On the constitution of atoms and molecules". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 26.151 (1913), pp. 1–25. DOI: 10.1080/14786441308634955, eprint: https://doi.org/10.1080/14786441308634955. URL: https://doi.org/10.1080/14786441308634955.

[3] N P Samios. "Early baryon and meson spectroscopy culminating in the discovery of the Omega minus SU(3) and quarks". In: (June 1993). URL: https://www.osti.gov/biblio/10162083.

[4] M. Gell-Mann. "A schematic model of baryons and mesons". In: *Physics Letters* 8.3 (1964), pp. 214–215. ISSN: 0031-9163. DOI: https://doi.org/10.1016/S0031-9163(64)92001-3. URL: http://www.sciencedirect.com/science/article/pii/S0031916364920013.

[5] G Zweig. *An SU_3 model for strong interaction symmetry and its breaking; Version 1*. Tech. rep. CERN-TH-401. Geneva: CERN, Jan. 1964. URL: http://cds.cern.ch/record/352337.

[6] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV". In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91177-2. URL: http://www.sciencedirect.com/science/article/pii/0370269383911772.

[7] G. Arnison et al. "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider". In: *Phys. Lett. B* 126 (1983), pp. 398–410. DOI: 10.1016/0370-2693(83)90188-0.

[8] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[9] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[10] The ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020. URL: http://www.sciencedirect.com/science/article/pii/S037026931200857X.

[11] The CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL: http://www.sciencedirect.com/science/article/pii/S0370269312008581.

[12] G. 'tHooft. "Renormalization of massless Yang-Mills fields". In: *Nuclear Physics B* 33.1 (1971), pp. 173–199. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(71)90395-6. URL: http://www.sciencedirect.com/science/article/pii/0550321371903956.

[13] F. Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course In Modern Particle Physics*. Jan. 1984. ISBN: 978-0-471-88741-6.

[14] Paul Langacker. *The standard model and beyond (2nd ed.)* June 2017. ISBN: 9780367573447.

[15] The ATLAS Collaboration. "Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment". In: *Phys. Rev. D* 101 (1 Jan. 2020), p. 012002. DOI: 10.1103/PhysRevD.101.012002. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.012002.

[16] The ATLAS Collaboration. "Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment". In: *Phys. Rev. Lett.* 120 (21 May 2018), p. 211802. DOI: 10.1103/PhysRevLett.120.211802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.211802.

[17] The ATLAS Collaboration. "A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector". In: *Physics Letters B* 812 (2021), p. 135980. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2020.135980. URL: http://www.sciencedirect.com/science/article/pii/S0370269320307838.

[18] Richard P. Feynman. "The Behavior of Hadron Collisions at Extreme Energies". In: *Special Relativity and Quantum Theory: A Collection of Papers on the Poincaré Group*. Ed. by M. E. Noz and Y. S. Kim. Dordrecht: Springer Netherlands, 1988, pp. 289–304. ISBN: 978-94-009-3051-3. DOI: 10.1007/978-94-009-3051-3_25. URL: https://doi.org/10.1007/978-94-009-3051-3_25.

[19] J. D. Bjorken and E. A. Paschos. "Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon". In: *Phys. Rev.* 185 (5 Sept. 1969), pp. 1975–1982. DOI: 10.1103/PhysRev.185.1975. URL: https://link.aps.org/doi/10.1103/PhysRev.185.1975.

[20] B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7. URL: http://www.sciencedirect.com/science/article/pii/0370157383900807.

[21] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[22] A. Accardi et al. "Parton Propagation and Fragmentation in QCD Matter". In: *Riv. Nuovo Cim.* 32 (2010), pp. 439–553. DOI: 10.1393/ncr/i2009-10048-0. arXiv: 0907.3534 [nucl-th].

[23] Roman Pasechnik and Michal umbera. "Phenomenological Review on QuarkGluon Plasma: Concepts vs. Observations". In: *Universe* 3.1 (Jan. 2017), p. 7. ISSN: 2218-1997. DOI: 10.3390/universe3010007. URL: http://dx.doi.org/10.3390/universe3010007.

[24] Ulrich W. Heinz and Maurice Jacob. "Evidence for a new state of matter: An Assessment of the results from the CERN lead beam program". In: (Jan. 2000). arXiv: nucl-th/0002042.

[25] I. Arsene et al. “Quarkgluon plasma and color glass condensate at RHIC? The perspective from the BRAHMS experiment”. In: *Nuclear Physics A* 757.1 (2005). First Three Years of Operation of RHIC, pp. 1–27. ISSN: 0375-9474. DOI: https://doi.org/10.1016/j.nuclphysa.2005.02.130. URL: http://www.sciencedirect.com/science/article/pii/S0375947405002770.

[26] B.B. Back et al. “The PHOBOS perspective on discoveries at RHIC”. In: *Nuclear Physics A* 757.1 (2005). First Three Years of Operation of RHIC, pp. 28–101. ISSN: 0375-9474. DOI: https://doi.org/10.1016/j.nuclphysa.2005.03.084. URL: http://www.sciencedirect.com/science/article/pii/S0375947405005282.

[27] J. Adams et al. “Experimental and theoretical challenges in the search for the quarkgluon plasma: The STAR Collaboration’s critical assessment of the evidence from RHIC collisions”. In: *Nuclear Physics A* 757.1 (2005). First Three Years of Operation of RHIC, pp. 102–183. ISSN: 0375-9474. DOI: https://doi.org/10.1016/j.nuclphysa.2005.03.085. URL: http://www.sciencedirect.com/science/article/pii/S0375947405005294.

[28] K. Adcox et al. “Formation of dense partonic matter in relativistic nucleusnucleus collisions at RHIC: Experimental evaluation by the PHENIX Collaboration”. In: *Nuclear Physics A* 757.1 (2005). First Three Years of Operation of RHIC, pp. 184–283. ISSN: 0375-9474. DOI: https://doi.org/10.1016/j.nuclphysa.2005.03.086. URL: http://www.sciencedirect.com/science/article/pii/S0375947405005300.

[29] The CMS collaboration. “Observation of long-range, near-side angular correlations in proton-proton collisions at the LHC”. In: *Journal of High Energy Physics* 2010.9 (Sept. 2010), p. 91. ISSN: 1029-8479. DOI: 10.1007/JHEP09(2010)091. URL: https://doi.org/10.1007/JHEP09(2010)091.

[30] The ATLAS Collaboration. “Measurement of the azimuthal anisotropy for charged particle production in $\sqrt{s_{NN}} = 2.76$ TeV lead-lead collisions with the ATLAS detector”. In: *Phys. Rev. C* 86 (1 July 2012), p. 014907. DOI: 10.1103/PhysRevC.86.014907. URL: https://link.aps.org/doi/10.1103/PhysRevC.86.014907.

[31] Jean-Yves Ollitrault. “Anisotropy as a signature of transverse collective flow”. In: *Phys. Rev. D* 46 (1 July 1992), pp. 229–245. DOI: 10.1103/PhysRevD.46.229. URL: https://link.aps.org/doi/10.1103/PhysRevD.46.229.

[32] S. Voloshin and Y. Zhang. “Flow study in relativistic nuclear collisions by Fourier expansion of azimuthal particle distributions”. In: *Zeitschrift fur Physik C Particles and Fields* 70.4 (May 1996), pp. 665–671. ISSN: 1431-5858. DOI: 10.1007/s002880050141. URL: http://dx.doi.org/10.1007/s002880050141.

[33] The ATLAS Collaboration. “Observation of Associated Near-Side and Away-Side Long-Range Correlations in $\sqrt{s_{NN}}$=5.02 TeV Proton-Lead Collisions with the ATLAS Detector”. In: *Phys. Rev. Lett.* 110 (18 May 2013), p. 182302. DOI: 10.1103/PhysRevLett.110.182302. URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.182302.

[34] M. Grazzini et al. “Higgs boson pair production at NNLO with top quark mass effects”. In: *Journal of High Energy Physics* 2018.5 (May 2018). ISSN: 1029-8479. DOI: 10.1007/jhep05(2018)059. URL: http://dx.doi.org/10.1007/JHEP05(2018)059.

[35] The ATLAS collaboration. "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2019.1 (Jan. 2019), p. 30. ISSN: 1029-8479. DOI: 10.1007/JHEP01(2019)030. URL: https://doi.org/10.1007/JHEP01(2019)030.

[36] The ATLAS collaboration. "Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2019.4 (Apr. 2019), p. 92. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2019)092. URL: https://doi.org/10.1007/JHEP04(2019)092.

[37] The ATLAS Collaboration. "Search for Resonant and Nonresonant Higgs Boson Pair Production in the $b\bar{b}\tau^+\tau^-$ Decay Channel in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector". In: *Phys. Rev. Lett.* 121 (19 Nov. 2018), p. 191801. DOI: 10.1103/PhysRevLett.121.191801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.191801.

[38] The ATLAS collaboration. "Search for Higgs boson pair production in the WW^*WW^* decay mode at $\sqrt{s}$= 13 TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2019.5 (May 2019), p. 124. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2019)124. URL: https://doi.org/10.1007/JHEP05(2019)124.

[39] The ATLAS collaboration. "Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state with 13 TeV pp collision data collected by the ATLAS experiment". In: *Journal of High Energy Physics* 2018.11 (Nov. 2018), p. 40. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2018)040. URL: https://doi.org/10.1007/JHEP11(2018)040.

[40] The ATLAS Collaboration. "Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel using pp collision data recorded at $\sqrt{s} = 13$=13TeV with the ATLAS detector". In: *The European Physical Journal C* 78.12 (Dec. 2018), p. 1007. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-6457-x. URL: https://doi.org/10.1140/epjc/s10052-018-6457-x.

[41] The ATLAS Collaboration. "Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector". In: *Physics Letters B* 800 (2020), p. 135103. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2019.135103. URL: http://www.sciencedirect.com/science/article/pii/S0370269319308251.

[42] The ATLAS Collaboration. "Search for non-resonant Higgs boson pair production in the bb final state with the ATLAS detector in pp collisions at $\sqrt{s}$=13 TeV". In: *Physics Letters B* 801 (2020), p. 135145. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2019.135145. URL: http://www.sciencedirect.com/science/article/pii/S0370269319308676.

[43] The ATLAS collaboration. "Reconstruction and identification of boosted di-τ systems in a search for Higgs boson pairs using 13 TeV proton-proton collision data in ATLAS". In: *Journal of High Energy Physics* 2020.11 (Nov. 2020), p. 163. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2020)163. URL: https://doi.org/10.1007/JHEP11(2020)163.

[44] ATLAS Collaboration. *Reconstruction and identification of boosted di-τ systems in a search for Higgs boson pairs using 13 TeV proton−proton collision data in ATLAS*. 2020. arXiv: 2007.14811 [hep-ex]. URL: https://arxiv.org/abs/2007.14811.

[45] D. de Florian et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. Oct. 2016. DOI: 10.23731/CYRM-2017-002. URL: https://cds.cern.ch/record/2227475.

[46] G.C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Physics Reports* 516.1-2 (July 2012), pp. 1–102. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2012.02.002. URL: http://dx.doi.org/10.1016/j.physrep.2012.02.002.

[47] Lisa Randall and Raman Sundrum. "Large Mass Hierarchy from a Small Extra Dimension". In: *Physical Review Letters* 83.17 (Oct. 1999), pp. 3370–3373. ISSN: 1079-7114. DOI: 10.1103/physrevlett.83.3370. URL: http://dx.doi.org/10.1103/PhysRevLett.83.3370.

[48] Kaustubh Agashe et al. "Warped gravitons at the CERN LHC and beyond". In: *Phys. Rev. D* 76 (3 Aug. 2007), p. 036006. DOI: 10.1103/PhysRevD.76.036006. URL: https://link.aps.org/doi/10.1103/PhysRevD.76.036006.

[49] Liam Fitzpatrick et al. "Searching for the Kaluza-Klein graviton in bulk RS models". In: *Journal of High Energy Physics* 2007.09 (Sept. 2007), pp. 013–013. DOI: 10.1088/1126-6708/2007/09/013. URL: https://doi.org/10.1088/1126-6708/2007/09/013.

[50] Simone Marzani, Gregory Soyez, and Michael Spannowsky. "Looking Inside Jets". In: *Lecture Notes in Physics* (2019). ISSN: 1616-6361. DOI: 10.1007/978-3-030-15709-8. URL: http://dx.doi.org/10.1007/978-3-030-15709-8.

[51] S. Catani et al. "Longitudinally-invariant k-clustering algorithms for hadron-hadron collisions". In: *Nuclear Physics B* 406.1 (1993), pp. 187–224. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(93)90166-M. URL: http://www.sciencedirect.com/science/article/pii/055032139390166M.

[52] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[53] Christoph W. Leemann, David R. Douglas, and Geoffrey A. Krafft. "THE CONTINUOUS ELECTRON BEAM ACCELERATOR FACILITY: CEBAF at the Jefferson Laboratory". In: *Annual Review of Nuclear and Particle Science* 51.1 (2001), pp. 413–450. DOI: 10.1146/annurev.nucl.51.101701.132327. eprint: https://doi.org/10.1146/annurev.nucl.51.101701.132327. URL: https://doi.org/10.1146/annurev.nucl.51.101701.132327.

[54] B.A. Mecking et al. "The CEBAF Large Acceptance Spectrometer (CLAS)". In: *Nucl. Instrum. Meth. A* 503 (2003), pp. 513–553. DOI: 10.1016/S0168-9002(03)01001-5.

[55] E.S. Smith et al. "The time-of-flight system for CLAS". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 432.2 (1999), pp. 265–298. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(99)00484-2. URL: http://www.sciencedirect.com/science/article/pii/S0168900299004842.

[56] M. Amarian et al. "The CLAS forward electromagnetic calorimeter". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 460.2 (2001), pp. 239–265. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(00)00996-7. URL: http://www.sciencedirect.com/science/article/pii/S0168900200009967.

[57] H. Hakobyan et al. "A double-target system for precision measurements of nuclear medium effects". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 592.3 (2008), pp. 218–223. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.04.055. URL: http://www.sciencedirect.com/science/article/pii/S0168900208006098.

[58] The CMS Collaboration. "Centrality dependence of dihadron correlations and azimuthal anisotropy harmonics in PbPb collisions at $\sqrt{s_{NN}}$ = 2.76 TeV". In: *The European Physical Journal C* 72.5 (May 2012), p. 2012. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-2012-3. URL: https://doi.org/10.1140/epjc/s10052-012-2012-3.

[59] K. H. Ackermann et al. "Elliptic Flow inAu+AuCollisions atsNN=130GeV". In: *Physical Review Letters* 86.3 (Jan. 2001), pp. 402–407. ISSN: 1079-7114. DOI: 10.1103/physrevlett.86.402. URL: http://dx.doi.org/10.1103/PhysRevLett.86.402.

[60] You Zhou and. "Anisotropic flow of identified particles in Pb-Pb collisions at = 2.76 TeV with the ALICE detector". In: *Journal of Physics: Conference Series* 509 (May 2014), p. 012029. DOI: 10.1088/1742-6596/509/1/012029. URL: https://doi.org/10.1088/1742-6596/509/1/012029.

[61] V.D. Burkert et al. "The CLAS12 Spectrometer at Jefferson Laboratory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 959 (2020), p. 163419. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.163419. URL: http://www.sciencedirect.com/science/article/pii/S0168900220300243.

[62] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: https://doi.org/10.1088/1748-0221/3/08/s08001.

[63] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003.

[64] *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s} = 13$ *TeV*. Tech. rep. ATLAS-CONF-2017-029. Geneva: CERN, May 2017. URL: https://cds.cern.ch/record/2261772.

[65] David Kirchmeier. "Reconstruction and Identification of Boosted Tau Pair Topologies at ATLAS". Presented 2015. May 2015. URL: https://cds.cern.ch/record/2105592.

[66] Yoav Freund and Robert E Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139. ISSN: 0022-0000. DOI: https://doi.org/10.1006/jcss.1997.1504. URL: http://www.sciencedirect.com/science/article/pii/S002200009791504X.

[67] A. Hoecker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: (2007). URL: https://arxiv.org/abs/physics/0703039.

[68] David Krohn, Jesse Thaler, and Lian-Tao Wang. "Jet trimming". In: *Journal of High Energy Physics* 2010.2 (Feb. 2010), p. 84. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2010)084. URL: https://doi.org/10.1007/JHEP02(2010)084.

[69] *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*. Tech. rep. ATLAS-CONF-2016-035. Geneva: CERN, July 2016. URL: https://cds.cern.ch/record/2200211.

[70] David Krohn, Jesse Thaler, and Lian-Tao Wang. "Jets with variable R". In: *Journal of High Energy Physics* 2009.06 (June 2009), pp. 059–059. DOI: 10.1088/1126-6708/2009/06/059. URL: https://doi.org/10.1088/1126-6708/2009/06/059.

[71] *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*. Tech. rep. ATL-PHYS-PUB-2016-012. Geneva: CERN, June 2016. URL: https://cds.cern.ch/record/2160731.

[72] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011), p. 1554. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: https://doi.org/10.1140/epjc/s10052-011-1554-0.

[73] M. Baak et al. "HistFitter software framework for statistical data analysis". In: *The European Physical Journal C* 75.4 (Apr. 2015), p. 153. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3327-7. URL: https://doi.org/10.1140/epjc/s10052-015-3327-7.

[74] A L Read. "Presentation of search results: theCLstechnique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313. URL: https://doi.org/10.1088/0954-3899/28/10/313.