UNIVERSITÀ DEGLI STUDI DI PAVIA
DIPARTIMENTO DI FISICA

CORSO DI LAUREA MAGISTRALE IN SCIENZE FISICHE

Color sensitive variables for Higgs studies

Variabili sensibili al colore per studi sul bosone di Higgs

Tesi per la Laurea Magistrale di
Alberto Lorenzo Rescia

Relatore:
Prof.ssa Daniela M. Rebuzzi

Correlatore:
Dott. Giovanni Stagnitto

Anno Accademico 2020/2021

Abstract

The discovery of the Higgs boson in 2012 marked the beginning of a new era of precision physics. In recent years, $H \to b\bar{b}$ decays in association with the production of a vector boson have been observed at the LHC. For a Higgs boson mass of 125 GeV, this decay channel has by far the largest branching ratio. It is therefore imperative to be able to correctly identify the resulting b-jets from this decay as having originated from the Higgs boson, as opposed to one of the many processes constituting the QCD background, so as to conduct precision studies on this rare particle.

To this aim, we focus on the problem of distinguishing b-jets originating from the decay of a color singlet from those originating from a color octet by means of a combination of color sensitive variables previously introduced in literature. We simulate 13 TeV proton-proton collisions and evaluate color sensitive observables before and after a fast simulation of the detector response. These variables are used to feed several machine learning algorithms. Results are given in terms of performance of single variables and of their combination.

La scoperta del bosone di Higgs nel 2012 ha segnato l'inizio di una nuova era di fisica di precisione. Negli anni recenti, il decadimento $H \to b\bar{b}$, prodotto in associazione con un bosone vettore, è stato osservato al LHC. Per una massa del bosone di Higgs pari a 125 GeV, questo canale di decadimento ha di gran lunga il più grande branching ratio: è quindi importante essere in grado di identificare correttamente i b-jet risultanti da questo decadimento come originatisi dal bosone di Higgs, differentemente da uno dei tanti processi che costituiscono il fondo di QCD, in modo da poter condurre studi di precisione su questa particella.

A questo scopo, ci focalizziamo sul problema di distinguere b-jet originatisi da un decadimento di un singoletto di colore da quelli originatisi da un ottetto di colore, attraverso una combinazione di variabili sensibili al colore precedentemente introdotte in letteratura. Simuliamo collisioni protone-protone a 13 TeV e valutiamo le osservabili sensibili al colore prima e dopo una simulazione "fast" della risposta del rivelatore. Queste variabili sono poi utilizzate per allenare diversi algoritmi di machine learning. I risultati sono dati in termini della performance delle singole variabili e della loro combinazione.

Contents

Introduction

The discovery of the Higgs boson in 2012 represents a turning point in our understanding of the Standard Model of particle physics. This discovery was so remarkable that it earned Peter Higgs and François Englert the 2013 Nobel Prize in Physics. Finally, after years of searching, the missing degree of freedom of the Standard Model was found, and physicists could begin to undertake the task of rigorously testing all properties of this particle.

The Higgs boson not only completes the known theory, but it also could give us clues regarding the existence of new physics. For example, several models of physics beyond the Standard Model include extensions containing expanded Higgs sectors. In addition to this, the currently known Higgs sector could provide insight to the nature of dark matter, since, if dark matter is a massive particle, it is natural to assume that it could couple to the Higgs boson. Finally, it cannot be excluded that a new discovery could stem from an unexpected anomaly or deviation from the Standard Model expectations. It is therefore imperative to be able to identify as many events containing the Higgs boson as possible - with high purity - in order to have sufficient statistics to find and claim any eventual new discovery.

The task of identifying events containing the Higgs boson is rendered difficult by the overwhelming QCD background. To further complicate this matter, the most likely decay channel of the Higgs, $H \rightarrow b\bar{b}$, is characterized by a final state which is far more likely to be produced by a gluon-initiated shower.

In this thesis, we illustrate a framework that we have developed to try to overcome this barrier by recognizing and distinguishing the two different color configurations of the Higgs boson and the gluon. The outline of this work is the following.

In Chapter 1, we begin by contextualizing the role of the Higgs boson within the Standard Model. In particular, we describe from a theoretical perspective the principal interactions of the Standard Model and show how they come together to build a complete theory.

Chapter 2 focuses on the physics of colliders. Over the past 50 years, colliders have been one of our main tools in understanding and building the Standard Model, and they continue to play a fundamental part in our understanding. We describe, again from a theoretical viewpoint, the physics behind particle collisions, and focus in particular on the role of jets, which represent a crucial bridge between theory and experiment.

Chapter 3 discusses Higgs physics from an experimental point of view. We describe the main production and decay mechanisms, examine the renowned 2012 discovery, and briefly describe the role that the Higgs sector will play moving forward.

Finally, in Chapter 4, we discuss the framework that we have developed to distinguish decays of color singlets from those of color octects. We begin by describing eight high-level, color sensitive variables which can be measured at colliders. We proceed by describing our simulation of proton-proton collisions at 13 TeV, and the analysis used to extract these variables. We conclude by discussing the details of the architectures of the machine learning algorithms used to identify the two different color configurations and illustrate our findings.

Chapter 1

The Standard Model

Since 2009, the Large Hadron Collider (LHC) has been exploring particle physics at the highest energy frontier in the hope of discovering new physics, including, but not limited to, the Higgs boson, supersymmetry, flavor violations, and much more. At the basis of all physics at the LHC lies the Standard Model. This theory describes all fundamental interactions involving the electroweak and strong forces, as well as the fields which partake in these interactions. In this chapter, we will briefly illustrate the main features of this theory and show how together they paint a complete picture of our current understanding of elementary particle physics.

1.1   An Overview of the Theory

The Standard Model of particle physics is composed of two sectors: the matter fields and gauge fields. The matter fields are fermionic fields whose excitations lead to the particles which make up ordinary matter, i.e. quarks and leptons, and their corresponding antiparticles. The gauge fields, on the other hand, are bosonic fields which arise from symmetries of the model and describe the force-carrying particles, specifically the photon (γ), gluons (g), $W^\pm$ and Z, as well as the Higgs boson, H.

The matter fields are organized in three generations, as shown in Figure 1.1. The particles in the three generations all have the same properties, with the sole exception of mass which increases with each generation. Neutrinos might be the exception to this rule, since their masses are not yet known to a useful degree of precision.

The properties of the matter fields determine the interactions in which they participate. All fermions can interact via the weak force. The electrically charged particles participate in electromagnetic interactions, and quarks, the only particles carrying color charge, can also interact via the strong nuclear force. The interactions with these forces are mediated by the exchange of the gauge bosons. Photons are responsible for electromagnetic interactions, gluons for strong interactions, and the $W^\pm$ and Z for weak interactions.

The strength of the interactions are determined by the value of the coupling constants, α for electromagnetic interactions, α_W for weak interactions and α_S for strong interactions. Despite their name, the coupling constants are known to run, i.e. they depend on the energy scale Q^2 of the process. Figure 1.2 shows how the coupling constants vary with energy. As can be seen, $1/\alpha$ decreases as Q^2 increases, while $1/\alpha_W$ and $1/\alpha_S$ increase with increasing energy. In this case of α_S, this property is known as asymptotic freedom.

The running of the coupling constants is determined by the β-function, a renormalization group equation which can be solved in perturbation theory. At one-loop, the electromagnetic

Standard Model of Elementary Particles

	three generations of matter (fermions)			interactions / force carriers (bosons)	
	I	II	III		
mass ≃2.2 MeV/c² charge ⅔ spin ½ **u** up	≃1.28 GeV/c² ⅔ ½ **c** charm	≃173.1 GeV/c² ⅔ ½ **t** top	0 0 1 **g** gluon	≃124.97 GeV/c² 0 0 **H** higgs

Figure 1.1: The "Periodic Table" of the Standard Model, showing the division between matter fields and gauge bosons, as well as the three different generations of the matter fields. Properties such as mass, charge and spin are also shown [1].
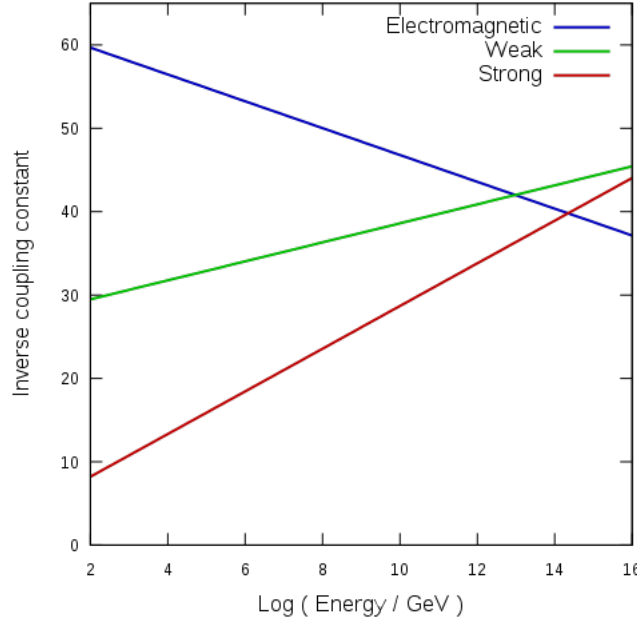
Figure 1.2: The β-functions for the inverses of the three coupling constants included in the Standard Model [2].

β-function is given by [3]

$$\beta(\alpha) = \frac{\alpha^2}{3\pi} + \mathcal{O}(\alpha^3), \tag{1.1}$$

while the β-function for the strong interaction can be written as

$$\beta(\alpha_S) = -\left(11 - \frac{2n_f}{3}\right)\frac{\alpha_S^2}{2\pi} + \mathcal{O}(\alpha_S^3) \tag{1.2}$$

The inverse dependence on energy of α_S is a direct consequence of the non-abelian nature of the strong nuclear force, as well as the number of flavors of quarks, n_f which enter the loops.

As mentioned, the gauge bosons originate from symmetries of the Standard Model. Specifically, the Standard Model Lagrangian is invariant under a local $SU_C(3) \otimes SU_L(2) \otimes U_Y(1)$ gauge symmetry. This means that, when a field ψ transforms under the action of this group, i.e. when

$$\psi \to \psi' = U(x)\psi, \tag{1.3}$$

the Lagrangian remains unchanged.

This is an example of invariance under a continuous symmetry. Per Nöther's theorem, these symmetries lead to conserved quantities, namely the conservation of color, weak isospin and hypercharge in every point of spacetime. There are a number of conserved quantities which arise from continuous symmetries: momentum from translational symmetries, angular momentum from rotational symmetries, and electric charge from a $U_{EM}(1)$ symmetry, just to name a few.

There are also a number of discrete symmetries present in the model. The most significant of these are *parity*, which inverts the spatial coordinates of a quantity

$$\hat{P}\psi(\vec{x}, t) = \psi(-\vec{x}, t), \tag{1.4}$$

charge symmetry, which exchanges all particles with the corresponding antiparticles,

$$\hat{C}\psi(x) = i\gamma^2\psi^*(x) \tag{1.5}$$

and *time-reversal symmetry* inverts the sign of the temporal coordinate of a quantity

$$\hat{T}\psi(\vec{x}, t) = \psi(\vec{x}, -t). \tag{1.6}$$

These symmetries are not individually present in all processes. For example, QED and QCD are both invariant under transformations of parity. These theories involve vector currents of the form $j^\mu = \overline{\psi}\gamma^\mu\psi$. These transform as

$$\begin{cases} \hat{P}j^0 = \overline{\psi}\gamma^0\gamma^0\gamma^0\psi = \overline{\psi}\gamma^0\psi \\ \hat{P}j^i = \overline{\psi}\gamma^0\gamma^i\gamma^0\psi = -\overline{\psi}\gamma^i\gamma^0\gamma^0\psi = -\overline{\psi}\gamma^i\psi = -j^i. \end{cases} \tag{1.7}$$

When we consider the matrix element $\mathcal{M}$ for one of these interactions, which is proportional to the scalar product of two currents, we find that

$$\hat{P}j_1 \cdot \hat{P}j_2 = j_1^0 j_2^0 - \sum_i \left(-j_1^i\right)\left(-j_2^i\right) = j_1 \cdot j_2, \tag{1.8}$$

i.e. the interaction is invariant under parity transformations. On the other hand, weak interactions famously show parity violations [4]. This is due to the fact that the currents involved are of type V-A, i.e. of the form $j^\mu \propto \overline{\psi}\gamma^\mu\left(1 - \gamma_5\right)\psi$. This form mixes a vector current with an axial vector current of the form $j_A^\mu = \overline{\psi}\gamma^\mu\gamma^5\psi$. In this case, when we consider the matrix element

$$\mathcal{M} \propto j_1 \cdot j_2 = g_V^2(j_V \cdot j_V) + g_A^2(j_A \cdot j_A) + g_V g_A(j_V \cdot j_A + j_A \cdot j_V) \tag{1.9}$$

we find three terms which conserve parity, and one parity-violating term, since

$$\hat{P}j_V \cdot \hat{P}j_A = -j_V \cdot j_A. \tag{1.10}$$

Overall, only the combination of charge, parity, and time-reversal symmetry is expected to be conserved in all Standard Model interactions.

1.2 Gauge Symmetries

1.2.1 QED Lagrangian

We shall now begin to construct the Standard Model Lagrangian. Let us start by considering the free Lagrangian for a massive fermion field, given by the Dirac Lagrangian

$$\mathcal{L}_D = \overline{\psi}(i\not{\partial} - m)\psi, \tag{1.11}$$

where ψ is the fermion field, $\overline{\psi}$ its Dirac adjoint, and, given the Dirac matrices γ^μ, $\not{\partial}$ is the del operator in Feynman slash notation. It is easy to show that this Lagrangian is invariant under transformations of the type

$$\psi \to \psi' = \exp(ie\alpha)\psi \tag{1.12}$$

where e is a parameter which represents the coupling constant and α is, for now, a parameter independent of the space-time coordinate x. In fact, the analogous transformation for the adjoint field $\overline{\psi}$ is

$$\overline{\psi} \to \overline{\psi'} = [\exp(ie\alpha)\psi]^\dagger\gamma^0 = \psi^\dagger\exp(-ie\alpha)\gamma^0 = \overline{\psi}\exp(-ie\alpha) \tag{1.13}$$

since the operator $\exp(-ie\alpha)$ commutes with γ^0. When applied to the whole Lagrangian, the transformation has the overall effect of leaving it unchanged:

$$\mathcal{L}_D \to \mathcal{L}'_D = \overline{\psi'}(i\slashed{\partial} - m)\psi' = \overline{\psi}(i\slashed{\partial} - m)\psi = \mathcal{L}_D. \tag{1.14}$$

Since the Lagrangian is unchanged, so too are the equations of motion. The transformed fields will therefore have the same dynamics.

We have just shown that the Dirac Lagrangian is invariant under a *global* $U(1)$ gauge symmetry in charge space.

At this point, if we want to construct the QED Lagrangian, we must add the kinetic term describing the free photon field

$$\mathcal{L}_{kin} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{1.15}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, as well as a term describing the interaction between the two fields. We can do this in two ways:

Minimal Coupling

The first, more direct, prescription calls for applying the minimal coupling rule. This requires substituting the four-momentum of the fermion, which we will take to be an electron, with an expression which includes the electromagnetic potential

$$p_\mu \to p_\mu - eA_\mu \tag{1.16}$$

and the coupling constant e. Quantum mechanically, this corresponds to substituting the del operator in the Lagrangian. The Lagrangian thus becomes

$$\mathcal{L}_{QED} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \overline{\psi}(i\slashed{\partial} - e\slashed{A} - m)\psi = \mathcal{L}_{kin} + \mathcal{L}_D - e\overline{\psi}\gamma^\mu\psi A_\mu. \tag{1.17}$$

This Lagrangian is still invariant under the same global gauge symmetry as before.

We would like, however, to impose a more stringent symmetry requirement: a *local* gauge symmetry dependent on the space-time coordinate x. Whereas a global symmetry establishes the conservation of a conserved quantity, e.g. electric charge, in any closed system, the local symmetry imposes the same requirement in each point x.

If we promote the gauge symmetry to a local symmetry, i.e. we apply the transformation

$$\psi \to \psi' = \exp[ie\alpha(x)]\psi, \tag{1.18}$$

we find that the Lagrangian is no longer invariant under this transformation due to the action of the derivative. In fact, ignoring the terms which remain invariant, we find that

$$\mathcal{L}' = i\overline{\psi'}\slashed{\partial}\psi' = i\overline{\psi'}\exp[-ie\alpha(x)]\gamma^\mu\partial_\mu\{\exp[ie\alpha(x)]\psi\} = i\overline{\psi}\slashed{\partial}\psi - e\overline{\psi}\gamma^\mu\psi\partial_\mu\alpha(x). \tag{1.19}$$

We can use a trick to re-obtain the gauge invariance. We know that the electromagnetic tensor $F^{\mu\nu}$ is gauge invariant. This means that if A_μ undergoes the transformation

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu f(x) \tag{1.20}$$

where $f(x)$ is a function such that $\Box f = 0$, then

$$F_{\mu\nu} \to F'_{\mu\nu} = \partial_\mu(A_\nu + \partial_\nu f) - \partial_\nu(A_\mu + \partial_\mu f) = F_{\mu\nu}. \tag{1.21}$$

If we choose f opportunely, we can cancel out the extra term which appears in (1.19) with the last term in (1.17). Specifically, the choice $f(x) = -\alpha(x)$ satisfies our request. Therefore, by combining the transformations (1.18) and (1.20), we can obtain an invariant Lagrangian.

Gauge Principle

The second, more general, way of adding an interaction term to the Lagrangian is via the gauge principle. This principle describes a protocol through which we can obtain the dynamics of QED, or any field theory, starting from the global gauge transformation (1.12).

We start, once again, from the Lagrangian (1.11). Having identified the global gauge symmetry of the Lagrangian and having promoted it to a local symmetry, we define the covariant derivative as

$$D_\mu \doteq \partial_\mu + ieA_\mu. \tag{1.22}$$

We then require that the term $D_\mu\psi$ transforms as the field ψ itself

$$D_\mu\psi \to (D_\mu\psi)' = D'_\mu\psi' = \{\partial_\mu + ieA'_\mu\}\psi' = \exp[ie\alpha(x)]D_\mu\psi. \tag{1.23}$$

By developing the equality, we find that A'_μ must be given by

$$A'_\mu = A_\mu - \partial_\mu\alpha(x) \tag{1.24}$$

in order for the Lagrangian to remain invariant.

We can then use the covariant derivative to build a term which describes the free propagation of the field A_μ. We do this by computing the commutator. With some basic algebra, we find that

$$[D_\mu, D_\nu] = ie\{\partial_\mu A_\nu - \partial_\nu A_\mu\} \equiv ieF_{\mu\nu}, \tag{1.25}$$

where $F_{\mu\nu}$ is now a generic tensor of the field A_μ. We thus have

$$F_{\mu\nu} = -\frac{i}{e}[D_\mu, D_\nu]. \tag{1.26}$$

We can then use the field tensor to construct a normalized, gauge invariant Lorentz scalar which will necessarily take the form (1.15).

We have thus arrived at the QED Lagrangian in a general way, without assuming any prior knowledge about the field A_μ.

1.2.2   QCD Lagrangian

Armed with the gauge principle, it is now straightforward to derive the QCD Lagrangian. We must note, however, that a few complications arise from the fact that we are now dealing with a non-abelian gauge theory, i.e. a theory whose symmetry group is non-commutative. For a general Yang-Mills theory, the gauge group is $SU(N)$, but in QCD we will be working with $N = 3$.

The Dirac field for the quark can be indicated as q_f^α where f is the flavor index and α is the color index. We know that each flavor comes in three colors, so we can group the fields for each flavor in a three-component vector

$$q_f = \begin{bmatrix} q_f^1 \\ q_f^2 \\ q_f^3 \end{bmatrix}. \tag{1.27}$$

We can thus write the free Lagrangian for the quarks as

$$\mathcal{L}_D = \sum_f \overline{q}_f \left(i\slashed{\partial} - m_f\right) q_f \tag{1.28}$$

where m_f is a parameter representing the quark mass and $(i\slashed{\partial} - m_f)$ is a three-dimensional diagonal matrix. The quark mass m_f must be understood as a free parameter of the Lagrangian since it is not directly measurable, as free quarks do not exist in nature.

The Lagrangian is invariant under the following global gauge transformations in color space:

$$q_f \to (q_f)' = \exp\left[i\theta_a \frac{\lambda^a}{2}\right] q_f \tag{1.29}$$

where θ_a is a parameter and $a = 1, \ldots, 8$ since, in general, the fundamental representation of $SU(N)$ has $N^2 - 1$ generators. λ^a represents the Gell-Mann matrices, which in the fundamental representation of $SU(3)$ can be written as

$$\lambda^1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \lambda^2 = \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \lambda^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\lambda^4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \qquad \lambda^5 = \begin{bmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{bmatrix}, \tag{1.30}$$

$$\lambda^6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad \lambda^7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix}, \qquad \lambda^8 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

The Gell-Mann matrices also allow us to define the structure constant of $SU(3)$, f_{abc}

$$\left[\frac{\lambda_a}{2}, \frac{\lambda_b}{2}\right] = i f_{abc} \frac{\lambda_c}{2}. \tag{1.31}$$

We can now proceed with the gauge principle. We define the covariant derivative as

$$D_\mu = \partial_\mu + i g_s \frac{\lambda_a}{2} G_\mu^a \tag{1.32}$$

where we have introduced the strong coupling constant g_s and eight spin-1 vector fields G_μ^a. These are the gluon fields. We now promote θ_a to $\theta_a(x)$ and require that $D_\mu q_f$ transform as q_f so as to fix the interaction term between the quarks and the gauge bosons. We find that

$$G_\mu^a \to (G_\mu^a)' = G_\mu^a - \frac{1}{g_s} \partial_\mu \theta^a(x) - f^{abc} \partial_\mu \theta_b(x) G_{\mu c}. \tag{1.33}$$

Last but not least, using the relation

$$-\frac{i}{g_s}[D_\mu, D_\nu] = \frac{\lambda_a}{2} G_{\mu\nu}^a \tag{1.34}$$

we can define the gluon tensor field

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_{\mu b} G_{\nu c} \tag{1.35}$$

which we use to construct the gauge-invariant kinetic term with proper normalization. We thus find that the complete Lagrangian takes the following form:

$$\mathcal{L}_{QCD} = -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} + \sum_f \bar{q}_f \left(i\slashed{D} - m_f\right) q_f. \tag{1.36}$$

1.3   Electroweak Unification

1.3.1   Lagrangian for Pure Weak Interactions

The gauge principle can also be applied to weak interactions. In this case, the fundamental symmetry is $SU_L(2)$, labelled as such because it only applies to left-handed particle states or right-handed anti-particle states. From here on we will consider only particle states, though analogous considerations hold for anti-particle states.

The symmetry acts on a weak-isospin doublet, e.g.

$$\psi_L = \begin{bmatrix} \nu_\ell \\ \ell^- \end{bmatrix}_L \tag{1.37}$$

composed of a left-handed neutrino and a lepton, or in general the left-handed states of any two fermions belonging to the same generation[1]. Corresponding leptonic right-handed states ℓ_R^- are placed in a singlet state, and right-handed neutrino states are not considered as they have not been observed in nature [5].

As before, by applying the local gauge transformations

$$\psi_L \rightarrow \psi_L' = \exp\left[i\frac{\tau_j}{2}\alpha^j(x)\right]\psi_L \tag{1.38}$$

$$\ell_R \rightarrow \ell_R' = \ell_R \tag{1.39}$$

where i=1,2,3 and τ^i are the generators of $SU(2)$, usually chosen to be the Pauli spin matrices

$$\tau^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \tau_2^2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \tau^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{1.40}$$

we can derive the full Lagrangian describing weak interactions

$$\mathcal{L}_W = -\frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} + \overline{\psi}_L(i\slashed{D} - m_i)\psi_L + \ell_R(i\slashed{\partial} - m_i)\ell_R \tag{1.41}$$

where $D_\mu = \partial_\mu + ig\frac{\tau_j}{2}W_\mu^j(x)$ and g is the weak coupling constant.

The fields $W_\mu^1(x), W_\mu^2(x)$, and $W_\mu^3(x)$ in their raw form are not sufficient to describe the observed phenomenology of weak interactions. By considering

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right) \tag{1.42}$$

we obtain charged currents which describe the transition from upper and lower components of the weak-isospin doublet observed in nature. Naturally, one would then attempt to identify W_μ^3 with the Z, however W_μ^3 only couples to left-handed particles or right-handed anti-particles, in contrast to what is observed for the physical Z.

1.3.2   Electroweak Unification

A more complete description is thus required to match the theory to the physical reality. The Z boson is not the only neutral boson observed in nature: there is also the γ. We can therefore attempt to include electromagnetic interactions in our description of weak interactions and derive the Z and γ fields from two neutral fields.

---

[1]Quark mixing slightly complicates this.

To this aim, we start by introducing *hypercharge*, defined as

$$Y = 2(Q - I_W^{(3)}) \tag{1.43}$$

This is a quantity meant to replace electric charge. It is a quantum number capable of distinguishing the states composing the left-handed doublet from the the one composing the right-handed singlet. The states considered in (1.37) both have hypercharge $Y = -1$ according to this definition, whereas the $SU_L(2)$ singlet state ℓ_R^- has hypercharge $Y = -2$.

We can now consider the full gauge symmetry for electroweak interactions, $SU_L(2) \otimes U_Y(1)$. Under this new gauge symmetry, the transformations (1.38) and (1.39) become

$$\psi_L \to \psi'_L = \exp\left[iy_1\beta(x)\right]\exp\left[i\frac{\tau_j}{2}\alpha^j(x)\right]\psi_L \tag{1.44}$$

$$\ell_R \to \ell'_R = \exp\left[iy_2\beta(x)\right]\ell_R \tag{1.45}$$

where y_1 and y_2 are the hypercharges of the weak isospin doublet and singlet, respectively. The covariant derivatives thus are

$$D_\mu\psi_L(x) = \left[\partial_\mu + ig\frac{\tau_j}{2}W_\mu^j(x) + ig'\frac{y_1}{2}B_\mu(x)\right]\psi_L(x) \tag{1.46}$$

$$D_\mu\ell_R = \left[\partial_\mu ig'\frac{y_2}{2}B_\mu(x)\right]\ell_R(x) \tag{1.47}$$

where g and g' are the two coupling constants, in general different from one another.

We now have four different gauge bosons, $W_\mu^j(x)$ and $B(x)$, which must be identified with the physical gauge bosons $W^\pm$, Z and γ. The physical W bosons can be identified through the relation in (1.42). The mapping from W_μ^3 and B_μ to Z_μ and A_μ can be achieved through a rotation in the neutral sector of the gauge bosons

$$\begin{bmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{bmatrix}\begin{bmatrix} B_\mu \\ W_\mu^3 \end{bmatrix} = \begin{bmatrix} A_\mu \\ Z_\mu \end{bmatrix}. \tag{1.48}$$

If we define the vector

$$\psi = \begin{bmatrix} \nu_{\ell L} \\ \ell_L \\ \ell_R \end{bmatrix} \tag{1.49}$$

and write out the full Lagrangian containing the neutral currents of the electroweak sector of Standard Model,

$$\begin{aligned}
\mathcal{L}_{NC} = &\overline{\psi}\gamma_\mu\left\{g\sin\theta_W\frac{\tau_3}{2} + g'\cos\theta_W\frac{Y(\psi)}{2}\right\}\psi A^\mu \\
&+ \overline{\psi}\gamma_\mu\left\{g\cos\theta_W\frac{\tau_3}{2} - g'\sin\theta_W\frac{Y(\psi)}{2}\right\}\psi Z^\mu,
\end{aligned} \tag{1.50}$$

we can see that we are required to impose a condition on the coefficients in (1.50) to re-obtain the physical currents that we are familiar with. Specifically, the first part of (1.50) corresponds to the interaction term of the Lagrangian (1.17) and the second term corresponds to an interaction term involving a second neutral boson. For the sake of simplicity, we can consider the case of an electron for the purpose of this matching.

The interaction term of (1.17), when specifying the left-handed and right-handed components, corresponds to

$$\mathcal{L} = -e\left[\overline{e}_L\gamma_\mu e_L + \overline{e}_R\gamma_\mu e_R\right]A^\mu. \tag{1.51}$$

fermion	Q	$I_W^{(3)}$	Y_L	Y_R
ν_ℓ	0	$+\frac{1}{2}$	-1	0
ℓ^-	-1	$-\frac{1}{2}$	-1	-2

Table 1.1: The quantum numbers associated to charged leptons and neutrinos.

By inspection, we find that

$$- e = g \sin \theta_W \frac{\tau_3}{2} + g' \cos \theta_W \frac{Y(\psi_e)}{2}. \tag{1.52}$$

By specifying τ_3 and $Y(\psi_e)$ to the appropriate component of ψ, in accordance with Table 1.1, we find that

$$g \sin \theta_W = g' \cos \theta_W = e. \tag{1.53}$$

θ_W is the *weak mixing angle*, corresponding to the angle of rotation in the neutral sector necessary to achieve the desired mapping. Experimentally, $\sin^2 \theta_W = 0.22290 \pm 0.00030$ [6].

We can also see that the second neutral current in (1.50) does indeed correspond to the Z current. It is easy to show that it can be written as

$$\mathcal{L}_{NC}^Z = \overline{\psi} \gamma_\mu \frac{e}{\sin \theta_W \cos \theta_W} Q_Z \psi Z^\mu. \tag{1.54}$$

where $Q_Z = \left\{ \frac{\tau_3}{2} - Q \sin^2 \theta_W \right\}$. Q_Z is a 3×3 diagonal matrix, allowing for access to each of the components of ψ. It needs to be specified in order to find the full coupling constant. This is straightforward for ν_{eL}: the coupling constant turns out to be $\frac{e}{2 \sin \theta_W \cos \theta_W}$.

For the electron, some additional manipulations must first be made since we must deal with the two components. The Lagrangian for the interaction can be written as

$$\mathcal{L}_{NC}^{Ze} = \frac{e}{\sin \theta_W \cos \theta_W} \left\{ \overline{e}_L \gamma_\mu Q_Z^L e_L + \overline{e}_R \gamma_\mu Q_Z^R e_R \right\} Z^\mu. \tag{1.55}$$

The projections can be obtained by considering the operators $P_{L/R} = \frac{1}{2} \left(1 \mp \gamma_5 \right)$, i.e.

$$\begin{cases} \overline{e}_L \gamma_\mu e_L = \overline{e} \gamma_\mu \frac{1}{2} \left(1 - \gamma_5 \right) e \\ \overline{e}_R \gamma_\mu e_R = \overline{e} \gamma_\mu \frac{1}{2} \left(1 + \gamma_5 \right) e. \end{cases} \tag{1.56}$$

The Lagrangian thus becomes

$$\mathcal{L}_{NC}^{Ze} = \frac{e}{\sin \theta_W \cos \theta_W} \left\{ \overline{e} \gamma_\mu \frac{1}{2} (Q_Z^L + Q_Z^R) e + \overline{e} \gamma_\mu \frac{1}{2} (Q_Z^L - Q_Z^R) e \right\} Z^\mu. \tag{1.57}$$

$Q_Z^{L/R}$ can be specified directly from (1.54) using the values from Table (1.1). We can use those values to specify the couplings which appear in (1.57). The Lagrangian simplifies to

$$\mathcal{L}_{NC}^{Ze} = \frac{e}{2 \sin \theta_W \cos \theta_W} \overline{e} \gamma_\mu \left\{ v_e - a_e \gamma_5 \right\} e Z^\mu \tag{1.58}$$

where

$$\begin{cases} v_e = I_W^{(3)}(e_L) \left(1 + 4 Q_e \sin^2 \theta_W \right) \\ a_e = I_W^{(3)}(e_L) \end{cases} \tag{1.59}$$

are the vector and axial components, respectively.

Thus, we correctly find that the interaction with the Z can involve both e_L and e_R and that the interaction is of the type V-A, with different coupling constants for the two chiral states of the electron.

The full Lagrangian for electroweak interactions can thus be written as

$$\mathcal{L}_{EW} = \overline{\psi}\left(i\slashed{D} - m\right)\psi - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^j_{\mu\nu}W^{\mu\nu}_j, \tag{1.60}$$

which, when combined with the rotation in (1.48), gives a correct description of the observed phenomenology.

1.4 Spontaneous Symmetry Breaking

So far, we have only considered Lagrangians which contain massless gauge bosons. This is for a very precise reason: mass terms vary under gauge transformations. If we consider, for example, the Proca action for a generic massive bosonic field in an abelian gauge theory,

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - m^2 A_\mu A^\mu \tag{1.61}$$

it is clear that when we apply the transformation (1.20) the Lagrangian is no longer invariant. This is a significant problem since it is known that the $W^\pm$ and Z bosons are massive.

To solve the problem of massive gauge bosons, it is necessary to introduce the *Brout-Englert-Higgs Mechanism* [7, 8], which induces the spontaneous breaking of the gauge symmetry.

This mechanism introduces a scalar field, known as the Higgs field, composed of a weak isospin doublet of two complex scalar fields, or equivalently four real scalar fields

$$\phi(x) = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}(\phi_3 + i\phi_4) \\ \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \end{bmatrix} \tag{1.62}$$

governed by a complex ϕ^4 Lagrangian

$$\mathcal{L} = (\partial_\mu\phi)^\dagger(\partial^\mu\phi) - V(\phi, \phi^\dagger), \tag{1.63}$$

where the potential $V(\phi, \phi^\dagger) = \frac{m^2}{2}\phi^\dagger\phi - \frac{\lambda}{4}(\phi^\dagger\phi)^2$. The components of the doublet, as part of the electroweak sector of the Standard Model, have quantum numbers

$$\begin{cases} I^{(3)}_W(\phi^+) = \frac{1}{2} \\ Y(\phi^+) = 1 \\ I^{(3)}_W(\phi^0) = \frac{1}{2} \\ Y(\phi^0) = 1 \end{cases} \tag{1.64}$$

This means that ϕ^+ carries electrical charge, based on (1.43).

The field ψ_L is added to the Lagrangian (1.60), leading to a Lagrangian which remains invariant under a global $SU_L(2) \otimes U_Y(1)$ gauge symmetry, as is easily verifiable. The symmetry can be promoted to a local gauge symmetry, resulting in the following Lagrangian for the Higgs field

$$\mathcal{L} = (D_\mu\phi)^\dagger(D_\mu\phi) - V(\phi, \phi^\dagger) \tag{1.65}$$

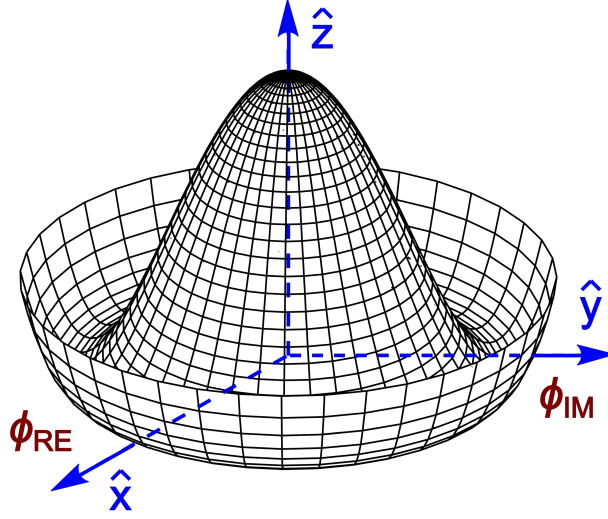where D_μ is the covariant derivative defined in (1.46).

Figure 1.3: The "Mexican hat" Higgs potential [9].

By minimizing the potential V, we can identify the ground state of the field

$$\frac{\partial V}{\partial |\phi|} = m^2 |\phi| + \lambda |\phi|^3 \tag{1.66}$$

If $m^2 > 0$ and $\lambda > 0$, the minimum occurs when $\phi = 0$. However, if we interpret m^2 as a parameter rather than as a mass and allow $m^2 < 0$, we find that there is a local maximum at $\phi = 0$ and a set of minimum value states at

$$\langle 0 | \phi^\dagger \phi | 0 \rangle \equiv (\phi^\dagger \phi)_0 = -\sqrt{\frac{m^2}{\lambda}}. \tag{1.67}$$

The quantum vacuum has thus shifted, as represented in Figure 1.3. The vacuum is *degenerate* since there are infinite values of $\phi^\dagger \phi$ which minimize V. The gauge symmetry is said to be spontaneously broken; the choice of parameters causes the vacuum state and the Lagrangian to no longer share the same symmetry. Since λ is an adimensional quantity, (1.67) has the dimensions of energy.

Without loss of generality, we can choose for the vacuum states

$$(\phi_1)_0 = -\sqrt{\frac{m^2}{\lambda}} \equiv \frac{v}{\sqrt{2}}, \ (\phi_i)_0 = 0 \tag{1.68}$$

where $i = 2, 3, 4$ and v is the vacuum expectation value of the Higgs field. Our doublet is thus

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v \end{bmatrix}. \tag{1.69}$$

The Lagrangian (1.60) along with (1.63) remain invariant under the local gauge transformation, however the vacuum state is no longer invariant under neither the $SU_L(2)$ nor the $U_Y(1)$ local gauge symmetries. For example,

$$\phi_0 \rightarrow \phi_0' = \exp\left[ig\frac{\tau_j}{2}\alpha^j(x)\right] \phi \approx \left\{1 + ig\frac{\tau_j}{2}\alpha^j(x) + \dots\right\} \phi_0 \neq \phi_0 \tag{1.70}$$

Specifically, the invariance is lost due to the action of the generators τ_j. For this reason, these generators are said to be *broken*. Likewise, Y is a broken generator.

Before continuing the discussion, we must first introduce an important theorem involving broken generators.

Theorem (Goldstone Theorem). *For all continuous global symmetries which do not leave the vacuum state unchanged, there exist corresponding massless particles equal in number to the number of broken generators.*

The theorem holds for global symmetries, however it is also relevant when dealing with local symmetries. In this case, the massless bosons which appear cannot be interpreted as physical particles. They can be gauged away, leading to Higgs-Kibble ghosts which result in *massive* bosons and allow for a physical interpretation of the theory. These ghosts are crucial for our ultimate goal: to give mass to the $W^\pm$ and Z while leaving γ massless.

In order to get to our desired result, we must make sure to have one unbroken generator. To obtain it, we can consider the following linear combinations

$$\begin{cases} Q = \frac{\tau_3}{2} - \frac{Y}{2} \\ Q' = \frac{\tau_3}{2} + \frac{Y}{2}. \end{cases} \tag{1.71}$$

It is easy to show that Q is an unbroken generator and Q' is broken. In this way we have obtained three broken generators (τ_1, τ_2, Q') and one unbroken (Q).

We can now proceed to study the vacuum fluctuations of ϕ. Naively, these can be written as

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_3(x) + i\phi_4(x) \\ v + \phi_1(x) + i\phi_2(x) \end{bmatrix}, \tag{1.72}$$

though, equivalently, we can write

$$\phi(x) = \frac{1}{\sqrt{2}} \exp\left[\frac{iT_j \xi_j(x)}{2}\right] \begin{bmatrix} 0 \\ v + H(x) \end{bmatrix} \tag{1.73}$$

where T_j are the three $SU(2)$ generators. In the latter form, we have merely parametrized the "naive" expression, as can be seen by expanding the exponential term to first order. The fields $\xi_j(x)$ are the ghosts, which we shall gauge away by choosing the unitary gauge

$$\phi(x) \to \phi'(x) = \exp\left[-\frac{iT_j \xi_j(x)}{2}\right] \phi. \tag{1.74}$$

In accordance with the gauge protocol, this requires a subsequent modification of D_μ, which leads to a modification of the gauge fields, which are said to "eat" the ghosts. After having done so, if we go on to calculate the first term in (1.63) and use (1.42), we find

$$(D_\mu \phi)^\dagger (D_\mu \phi) = \frac{g^2}{4} \left(v^2 + 2vH + H^2\right) W_\mu^+ W^{-\mu} + \frac{1}{8}(v+H)^2 \left(g^2 W_\mu^3 W^{3\mu} - 2gg' W_\mu^3 B^\mu + g'^2 B_\mu B^\mu\right) \tag{1.75}$$

which contains all the physically meaningful terms. In particular, we can see that (1.75) contains Proca mass terms such as

$$\frac{g^2}{4} v^2 W_\mu^+ W^{-\mu} \equiv M_W^2 W_\mu^+ W^{-\mu}. \tag{1.76}$$

We can then follow the same logic as in that used to derive the unified electroweak theory and map B_μ and W_μ^3 to A^μ and Z_μ, respectively. After doing so, the second term in (1.75) becomes

$$\frac{1}{8}(g^2 + g'^2)(v+H)^2 Z_\mu Z^\mu. \tag{1.77}$$

We can see that there is no mass term for A_μ, no interaction term involving H and A_μ, and that the Z acquires the mass $M_Z^2 = \frac{v^2}{4}(g^2 + g'^2)$. We can also see that the coupling of the Higgs boson to the W and Z is proportional the square of their masses.

We have succeeded in giving mass to the three weak gauge bosons. The choice of parameters λ and m^2 spontaneously breaks the gauge symmetry, and the interaction of the field ϕ with the potential generates would-be Goldstone bosons which manifest as ghosts. The choice of unitary gauge allows for the gauge bosons to eat the ghosts, thus gaining mass. The energy scale at which the gauge symmetry is spontaneously broken, known as the vacuum expectation value, is given by v, which in numerical terms corresponds to 246 GeV. Above this energy, the electromagnetic force and the weak nuclear force become one unified electroweak force.

1.5   Yukawa Lagrangian

A similar problem occurs when considering the mass terms for fermions. In this case, the mass term appearing in the Dirac Lagrangian (1.11) does not respect the $SU_L(2) \otimes U_Y(1)$ gauge symmetry due to the fact that the left and right-handed components of the spinor transform differently

$$- m\overline{\psi}\psi = -m \left(\overline{\psi}_R \psi_L + \overline{\psi}_L \psi_R \right). \tag{1.78}$$

This problem can be solved by introducing an interaction with the Higgs field. An infinitesimal $SU(2)$ local gauge transformation has the following effect on the Higgs

$$\phi \rightarrow \phi' = \left\{ 1 + ig\frac{\tau_j}{2}\alpha^j(x) \right\} \phi. \tag{1.79}$$

On the other hand, the same transformation has the opposite effect on $\overline{\psi}_L$

$$\overline{\psi}_L \rightarrow \overline{\psi}'_L = \overline{\psi}_L \left\{ 1 - ig\frac{\tau_j}{2}\alpha^j(x) \right\}. \tag{1.80}$$

Therefore, if we consider the combination $\overline{\psi}_L \phi$, we find that this is a gauge invariant quantity. The same holds true for the $U(1)$ gauge symmetry. Since ℓ_R transforms independently from ψ_L, we can add it to the combination so as to account for the right-handed component as well. Thus the Lagrangian

$$\mathcal{L}_Y = -k \left(\overline{\psi}_L \phi \ell_R + \overline{\ell}_R \overline{\phi} \psi_L \right), \tag{1.81}$$

where k is a coupling constant, is invariant under a $SU_L(2) \otimes U_Y(1)$ local gauge transformation. We can take once again the electron as an example and specify the terms in (1.81). We find that

$$\mathcal{L}_Y = -\frac{kev}{\sqrt{2}} \left(\overline{e}_L e_R + \overline{e}_R e_L \right) - \frac{keH}{\sqrt{2}} \left(\overline{e}_L e_R + \overline{e}_R e_L \right). \tag{1.82}$$

We thus find the Dirac mass term

$$m_e = \frac{kev}{\sqrt{2}} \tag{1.83}$$

as well as a term which couples the Higgs field to the fermion field. This term is proportional to the fermion mass.

In contrast to the derivation of the masses of the gauge bosons, the derivation of the fermionic masses is ad-hoc. The fermionic mass terms depends on the coupling k which must be measured from experiment. There is no explanation for the observed mass hierarchy of the fermions.

It is generally assumed that a different mechanism is responsible for the masses of the neutrinos. Although there is no specific reason as to why neutrinos should not have a Yukawa coupling

like the other elementary fermions, given their comparatively small mass, assumed to be of the order of ∼meV, they would have to have a coupling orders of magnitude smaller than that of the other fermions.

Instead, one possible explanation for the mass of the neutrinos is given by the *see-saw mechanism* [10]. If the neutrino is a Majorana particle and a sterile, supermassive (∼PeV), right-handed fourth flavor of neutrino exists, it is possible that the large mass of this hypothetical neutrino causes the small mass of the others. Several experiments are currently for evidence which supports this theory.

1.6   The Standard Model Lagrangian

We are now ready to put all the ingredients discussed together and bake the cake that is the Standard Model. The full Lagrangian for the model, written in compact form, is given by

$$
\begin{aligned}
\mathcal{L}_{SM} = &-\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \\
&+ \bar{\psi}\slashed{D}\psi + h.c. \\
&+ \psi_i y_{ij} \psi_j \phi + h.c. \\
&+ |D_\mu \phi|^2 - V(\phi).
\end{aligned}
\tag{1.84}
$$

This Lagrangian describes all possible interactions between all particles, and includes the building blocks described in this chapter, $\mathcal{L}_{QCD}$, $\mathcal{L}_{EW}$, $\mathcal{L}_{SSB}$ and $\mathcal{L}_Y$ gives mass to the fermions. The elementary vertices are summarized in Figure 1.4.

We should mention that, in this form, the Lagrangian is purely classical: it must then be quantized and renormalized [12] in order to fully describe our quantum world.
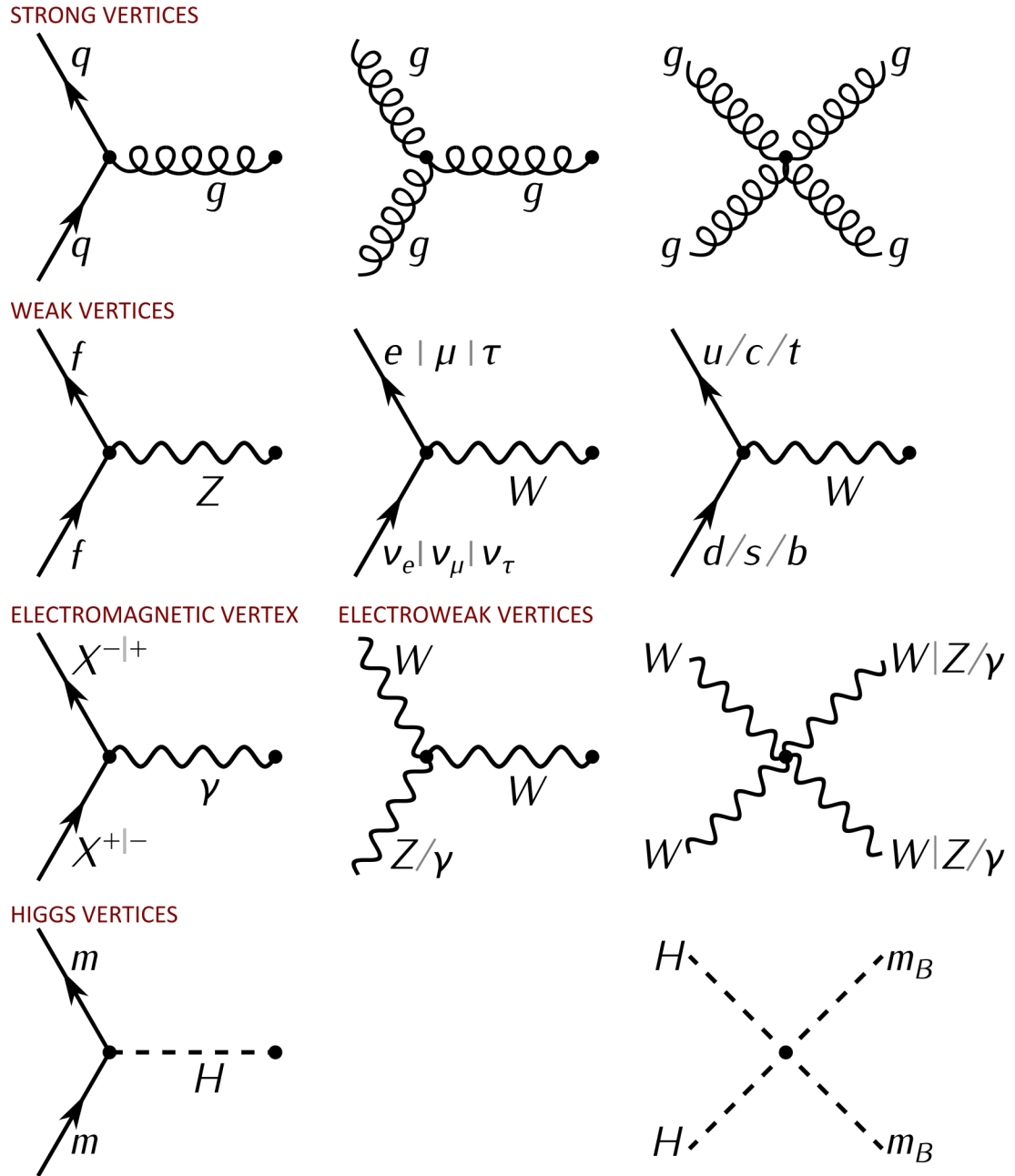
Figure 1.4: They Feynman diagrams of all interactions predicted by the Standard Model [11].

Chapter 2

Collisions at the LHC

One of the best ways to test the Standard Model is through high-energy particle collisions. The LHC is a 27 km hadron-hadron circular collider where protons or nucleons interact with each other in a multitude of ways, resulting in a myriad of possible final states involving particles which do not exist in our cold universe. By identifying these final states, and selecting those which correspond to processes of interest, it is possible to study these processes in detail. A number of detectors, including ATLAS and CMS, lie on the beam pipe for this purpose.

The LHC has undergone several phases. During Run 1 (2009-2013), the LHC ran at a center of mass energy $\sqrt{s}$ of 7-8 TeV. The energy was increased to $\sqrt{s} = 13$ TeV during Run 2 (2015-2018). Run 2 delivered a total integrated luminosity of 156 fb^{-1}. During Run 3, set to begin next year, the center of mass energy is set to increase up to 14 TeV, and it is expected that the integrated luminosity will double to 300 fb^{-1}.

In this chapter, we will shortly describe the physics of hadron colliders.

2.1 Factorization

The proton is a dynamic system. In a simplistic view, it is composed of three valence quarks, u, u, d bound together by gluons. The gluons interact both with the valence quarks and themselves, leading to a "sea" composed of gluons, quarks and anti-quarks of all flavors, which originate from gluon splitting. The sea is dominant at low energies, and suppressed at higher energies.

When two hadrons collide at high energies, the resulting interaction does not directly involve the hadrons as a whole but the *partons* which constitute the hadron. Thanks to asymptotic freedom, these partons are quasi-free. This means that, even when the interaction with the hadron as a whole is deeply inelastic, the parton-parton interaction which occurs is instead elastic.

Relativistic considerations also allow us to deduce that the time-scale is such that interactions involving more than one parton per hadron are suppressed. Indeed, in the rest frame of the proton, the time-scale of the interactions holding the proton together are of the order $1/m_p$. In the laboratory frame of the collision, this is boosted by a factor $\gamma = \sqrt{s}/2m_p$. Since the energies available at the LHC place us firmly in the ultrarelativistic limit, interactions with a virtual particle of energy $Q^2 \gg m_p^2$ occur on a time-scale much shorter than γ/m_p, and the parton probed has no time to communicate with the other partons.

In the case of deep inelastic scattering, after the constituent parton has been struck, the virtual particles emitted by the constituent as part of normal interactions within the hadron can no longer be reabsorbed. This effect is exacerbated at higher Q. The end result is a perturbative

evolution of the final state particles from the interaction, together with these liberated virtual particles, down to energies of the order of the Landau pole of QCD, Λ_{QCD}. At these energies, due to the running of the coupling constant α_S, it is no longer possible to describe QCD using perturbative physics. What follows is the hadronization process, where final state particles undergo non-perturbative interactions which transform them into relatively long-lived hadrons. Since the time-scale for hadronization is much longer when compared to that of the elementary process, the cross section for the collision is said to be *factorized* into a hard process described by an elementary cross section and functions describing the non-perturbative physics involved in the hadron before and after the interaction.
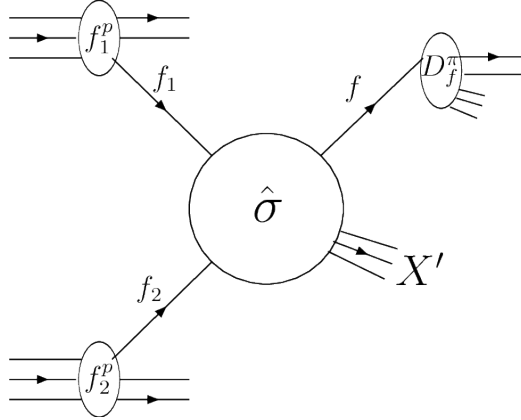

Figure 2.1: A schematic representation of factorization in a process which results in the production of pion in a proton-proton collision [17].

This discussion can be neatly summarized in a formula:

$$\frac{d\sigma}{dX} = \sum_{j,k} \int dx_1 dx_2 \, f_j(x_1, Q^2) f_k(x_2, Q^2) \, \frac{d\hat{\sigma}_{jk}(x_1 P_1, x_2 P_2, Q^2, \mu_F^2)}{d\hat{X}} \, F(\hat{X} \to X; Q^2, \mu_F^2). \quad (2.1)$$

Here, we are stating that the differential cross section with respect to a hadronic observable X can be written in terms of the parton-level cross section $d\hat{\sigma}/d\hat{X}$, differential in the parton-level observable $\hat{X}$. $f_j(x_1 P_1, \mu_F)$ and $f_k(x_2 P_2, Q^2)$ are known as Parton Distribution Functions (PDFs) and, at the lowest order in perturbation theory, describe the probability of extracting a parton of type j or k with momentum fraction x_1 or x_2, respectively from the colliding hadrons with momenta P_1 and P_2 when probed at energy Q^2. We must sum over all possible partons, and integrate over all momentum fractions. Finally, the function $F(\hat{X} \to X; Q^2, \mu_F^2)$ describes the (non-perturbative) transition from partonic states to hadronic states and indicates how to relate any partonic observable $\hat{X}$ with the measured hadronic observable X. This can be, for example, a Fragmentation Function, used when a single final-state hadron is observed, or a Jet Function, used instead when describing an aggregate transition to hadrons. We will describe all of the involved functions in more detail in subsequent sections.

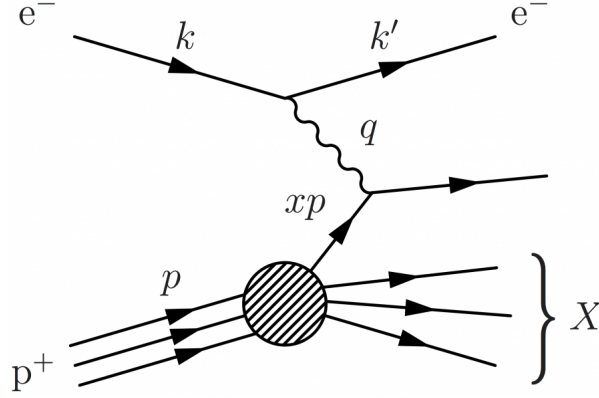Figure 2.1 is a schematic representation of factorization in a proton-proton collision described by Equation (2.1).

Figure 2.2: A schematic representation of the DIS process.

2.2 Parton Distribution Functions

As previously stated, Parton Distribution Functions contain information regarding the constituents of a hadron. To better understand them, we shall study in detail a benchmark process, known as Deep Inelastic Scattering (DIS).

2.2.1 Deep Inelastic Scattering

DIS is a process which investigates the insides of hadrons using leptonic probes. In this section we will focus on high-energy electron-proton inelastic scattering. In this process, the incoming electron exchanges a virtual electroweak boson with the proton. Specifically, we will look at the process

$$e^- \ p \to e^- \ X \tag{2.2}$$

where X represents an undetermined hadronic final state, mediated by a virtual photon.

Kinematics

A number of kinematic variables are needed to fully describe the process. Without loss of generality, we will work in the target rest frame (TRF), where the struck proton is at rest. With reference to Figure 2.2, we have

$$\begin{cases} k^\mu = (E, 0, 0, E) \\ p^\mu = (m_p, 0, 0, 0) \\ k^{\mu\prime} = (E', E' \sin\theta, 0, E' \cos\theta) \\ q^\mu = k^\mu - k^{\mu\prime} \end{cases} \tag{2.3}$$

where k^μ and $k^{\mu\prime}$ refer to the four-momenta of the initial and final state electron, respectively, p^μ the four-momentum of the initial state proton and q^μ refers to the four-momentum of the virtual photon. The quantities E and E' refer to the respective energies of the electron in the initial and final state, and θ is the angle at which the electron is scattered with respect to its initial momentum $\vec{k}$.

The first kinematic variable to consider is the energy of the virtual probe, known also as the scale of the process. Since DIS is a space-like process, we have $q^2 < 0$, where q^2 represents the

four-momentum of the virtual probe. Therefore we take

$$Q^2 = -q^2. \tag{2.4}$$

Next we introduce a series of Lorentz invariant variables. The first variable, y, is defined as

$$y = \frac{p \cdot q}{p \cdot k} \stackrel{\text{TRF}}{=} 1 - \frac{E'}{E}. \tag{2.5}$$

As is clear, in the TRF, y represents the fraction of energy lost during the inelastic process. For this reason, $0 \leq y \leq 1$. Next, we introduce a similar variable

$$\nu = \frac{p \cdot q}{m_p} \stackrel{\text{TRF}}{=} E - E' \tag{2.6}$$

which in the TRF represents the the energy lost by the electron in the scattering process. Finally, we introduce the variable Bjorken x, defined as

$$x = \frac{Q^2}{2p \cdot q} \stackrel{\text{TRF}}{=} \frac{Q^2}{2m_p\nu} \tag{2.7}$$

If we consider the definition of the invariant mass of the system,

$$W^2 = (p + q)^2 = p^2 + 2p \cdot q + q^2 = m_p^2 + Q^2 \left(\frac{1}{x} - 1\right) \geq m_p^2 \tag{2.8}$$

it becomes clear that x represents the "elasticity" of the process: x is limited to the range $0 \leq x \leq 1$, and $x = 1$ corresponds to a perfectly elastic collision, whereas $x = 0$ corresponds to a perfectly inelastic one. By definition, the deep inelastic limit of this process is the limit in which $Q^2 \to \infty$ while x is held constant.

Cross Section

It is possible to calculate the cross section for the inelastic scattering process described above. The most general Lorentz invariant cross section for the interaction considered is

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4} \left[\left(1 - y - \frac{m_p^2 y^2}{Q^2}\right) \frac{F^2(x, Q^2)}{x} + y^2 F_1(x, Q^2)\right]. \tag{2.9}$$

In the deep inelastic limit, this simplifies to

$$\frac{d^2\sigma}{dxdQ^2} \approx \frac{4\pi\alpha^2}{Q^4} \left[(1 - y) \frac{F^2(x, Q^2)}{x} + y^2 F_1(x, Q^2)\right]. \tag{2.10}$$

F_1 and F_2 are *structure functions*, which describe the internal structure of the proton.

Bjorken Scaling and the Callan-Gross Relation

An important observation involving the aforementioned structure functions is *Bjorken scaling*. This is the observation that, to first order, F_1 and F_2 are independent of Q^2, as shown in Figure 2.3. In addition to this scaling, experimental observations also found that, at sufficiently large Q^2

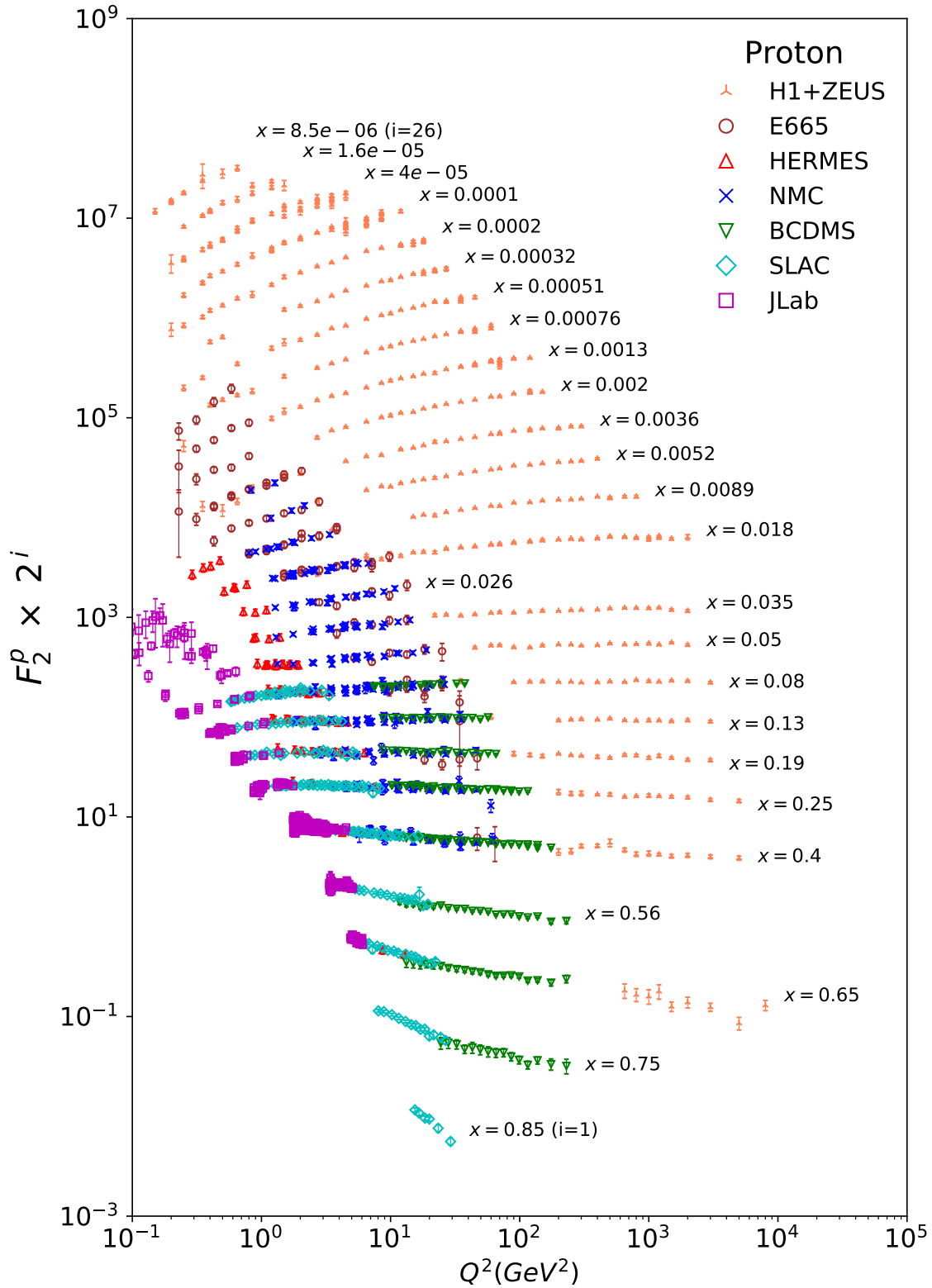$$F_2(x) = 2xF_1(x). \tag{2.11}$$

Figure 2.3: The observation of approximate scaling of the structure functions at various experiments. The observed violations are caused by higher-order contributions [18].
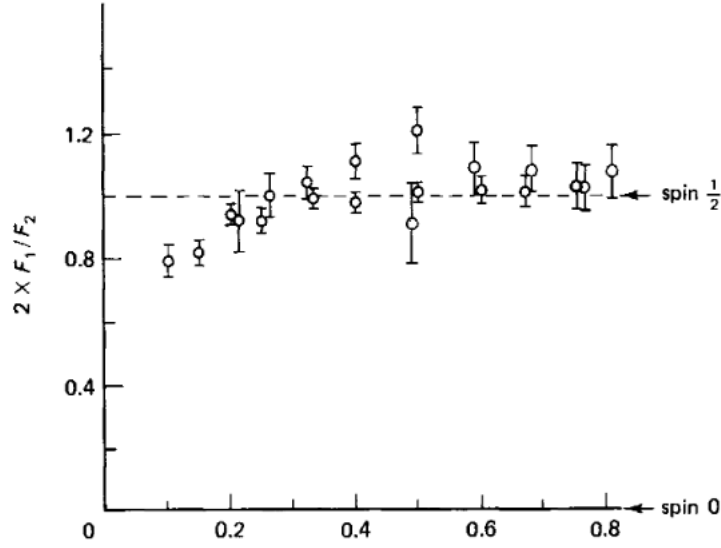
Figure 2.4: An experimental observation of the Callan-Gross relation [19].

This relation is known as the *Callan-Gross relation* and is shown in Figure 2.4. We would expect to see scaling if the scattering occurred against point-like particles, giving evidence to the composite nature of the proton. In addition to this, the Callan-Gross relation tells us that the constituent partons carry spin-1/2. We therefore have experimental evidence to support the statement that the proton is composed of point-like spin-1/2 particles, namely quarks[1]. The incoming electron elastically scatters against these constituents, explaining these observations and justifying the formula (2.1).

If we consider the DIS process in the infinite momentum frame, defined as the frame in which the energy of the proton in the initial state $E_p \gg m_p$, i.e. $p^\mu = (E_p, 0, 0, E_p)$, we can deduce that the four-momentum of the struck quark can be written as

$$p_q^\mu = (\xi E_p, 0, 0, \xi E_p) \tag{2.12}$$

where ξ represents the fraction of the proton's momentum carried by the quark. The quark in the final state of the electron-quark scattering will have four-momentum

$$(\xi p + q)^2 = \xi^2 p^2 + 2\xi p \cdot q + q^2 = m_q^2. \tag{2.13}$$

For this relation to hold, we must have $2\xi p \cdot q + q^2 = 0$, which implies that

$$\xi = \frac{-q^2}{2p \cdot q} = \frac{Q^2}{2p \cdot q} = x. \tag{2.14}$$

We can therefore identify Bjorken x with the momentum fraction of the struck constituent quark.

---

[1]The proton also has a significant gluon component, as will be discussed in a later section.

Determining PDFs

We can now proceed in actually determining the constituents of the proton. Experimentally, this is done using formula (2.1), which in the case of DIS becomes

$$\frac{d\sigma}{dE'd\Omega} = \sum_f \int_0^1 dx \; \frac{d\hat{\sigma}}{dE'd\Omega}(xP, q) \; \phi_f(x). \tag{2.15}$$

where $\phi_f(x)$ represents the PDF for the (anti)quark of flavor f within the proton. The partonic cross section is easily calculable in QED

$$\frac{d\hat{\sigma}}{dE'd\Omega} = \frac{4\alpha^2}{Q^4} E'^2 \cos^2 \frac{\theta}{2} e_f^2 \frac{2m_f x}{Q^2} \delta\left(x' - x\right) \left[1 + \frac{Q^2}{2m_f^2} \tan^2 \frac{\theta}{2}\right]. \tag{2.16}$$

By inserting (2.10) and (2.16) in (2.15), we find that

$$\begin{cases} F_1(x) = \frac{1}{2} \sum_f e_f^2 \phi_f(x) \\ F_2(x) = x \sum_f e_f^2 \phi_f(x). \end{cases} \tag{2.17}$$

We have managed to describe the structure functions in terms of the PDFs, and in turn found the Callan-Gross relation!

By experimentally measuring the cross section (2.15), we can deconvolve the PDF contribution. Figure 2.5 shows the PDFs of various partons within the proton at $Q^2 = 10$ GeV^2.

PDFs have the notable property of being *universal*. This means that, regardless of what is used to probe them, they will always be the same since they are an intrinsic property of a given hadron. They are intrinsically non-perturbative, but their evolution with energy is governed by the renormalization group equation obtained in perturbation theory, as described in the below.

Determining the proton PDFs is fundamental for making accurate predictions at the LHC, as they allow us to to obtain theoretical predictions for inclusive or exclusive variables, as in 2.1, and compare them to measured rates.

Gluon PDF

The discussion up to now has focused exclusively on quarks, since the original framework of the so-called *Quark Parton Model* did not include gluons. However, Figure 2.5 shows that, particularly at small x, i.e. at high energies, a significant fraction of the proton's constituents are gluons.

The gluon PDF can be determined in a manner analogous to the quark PDFs. With DIS, we must, however, consider higher-order QCD corrections to be able to study elementary processes such as $\gamma g \to q\bar{q}$. This is not ideal since the process of interest is suppressed. Alternatively, we can look at hadron-hadron collisions, where gluon interactions do appear at leading order for some processes.

PDF Evolution

In general, PDFs are functions both of the momentum fraction x and the scale of the hard process Q^2. At first order, this dependence is negligible, but the inclusion of the gluon leads to higher-order QCD corrections which lead to the scaling violations observed.

PDFs cannot be calculated from first principles; they must be measured from experiment. Thankfully, it is not necessary to perform measurements at different values of x and Q^2 to fully determine PDFs.
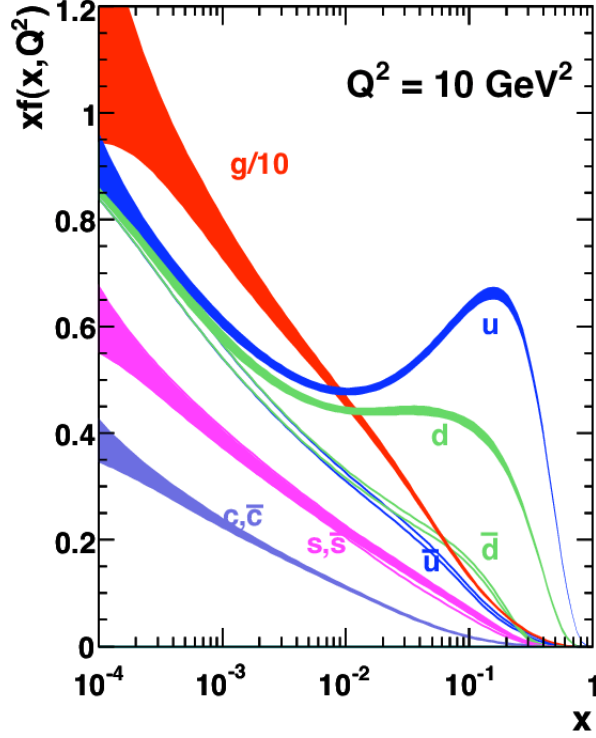
Figure 2.5: PDFs of the proton measured at various values of x at $Q^2 = 10$ GeV^2 [20].

The Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) Equations [21, 22] allow us to calculate the evolution of a PDF from the scale Q^2 to a different scale Q'^2. This is a notable achievement, as the PDFs can be measured at a certain energy, but re-utilized in a theoretical prediction for a process occurring at a different energy scale.

For the sake of brevity, we shall limit ourselves to citing this incredible result, rather than deriving the whole equation.

The evolution equation for the quark parton density is:

$$\frac{dq_f(x, Q^2)}{d \log Q^2} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y} \left[q_f(y, Q^2) P_{qq} \left(\frac{x}{y} \right) + g(y, Q^2) P_{qg} \left(\frac{x}{y} \right) \right]. \tag{2.18}$$

This equation states that, to calculate the PDF for a parton of flavor f at a given x and Q^2 we need only to integrate the parton and gluon PDFs at y and Q^2, along with the splitting functions P_{qq} and P_{qg}. P_{qq} gives the probability of finding a real quark q with a certain momentum fraction after the emission of a virtual gluon, and P_{qg} is the analogous function for the emission of a real quark after a gluon splitting.

$$\begin{cases} P_{qq}(z) = C_F \frac{1+z^2}{(1-z)_+} \\ P_{qg}(z) = T_R \left[z^2 + (1-z)^2 \right] \end{cases} \tag{2.19}$$

where $C_F = (N_c^2 - 1)/2N_c$ is the Casimir invariant of $SU(3)$, T_R is the trace of the Gell-Mann

matrices and the plus-prescription, defined as

$$[f(x)]_+ = f(x) - \delta(1-x) \int_0^1 f(z)\,dz, \tag{2.20}$$

has been used to regularize the divergent integral.

An analogous function exists for the evolution of the gluon PDF

$$\frac{dg(x,Q^2)}{d\log Q^2} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y} \left[\sum_f q_f(y,Q^2) P_{gq}\left(\frac{x}{y}\right) + g(y,Q^2) P_{gg}\left(\frac{x}{y}\right) \right] \tag{2.21}$$

where this time

$$\begin{cases} P_{gq}(z) = C_F \left[\frac{1+(1-z)^2}{z} \right] \\ P_{gg}(z) = 2C_A \left[\frac{1-z}{z} + \frac{z}{(1-z)_+} + z(1-z) \right] \end{cases} \tag{2.22}$$

where C_A is the Casimir operator for the adjoint representation of $SU(3)$.

2.3 Fragmentation Functions

Fragmentation functions play a role similar to that of PDFs, but rather than describing the possible initial states of the interaction, they describe the possible final states that may be observed.

To understand Fragmentation Functions, we refer to the *semi-inclusive* DIS process, where we observe one hadron in the final state. The kinematic variables are the same as in (2.3), though we must now include the four-momentum of the final-state hadron

$$P_h^\mu = (E_h, \vec{P_h}) \tag{2.23}$$

and one additional Lorentz invariant variable

$$z_h = \frac{P \cdot P_h}{P \cdot q} \stackrel{\text{TRF}}{=} \frac{E_h}{\nu} \tag{2.24}$$

which in the TRF represents the fraction of energy of the observed particle lost in the hadronization process.

The cross section can again be calculated without difficulty:

$$\frac{d\sigma}{dxdydz} = \frac{4\pi\alpha_S^2}{Q^4} \left(\frac{y}{2} + 1 - y \right) x \sum_f e_f^2 \phi_f(x) D_f(z) \tag{2.25}$$

where the sum over all flavors includes the anti-quarks. In this way, the Fragmentation Function $D_f(z)$ represents the probability of finding a hadron with fraction z of the available energy. The cross section remains factorized, allowing for information on the Fragmentation Functions by comparison with data, as they cannot be calculated in QCD. Finally, it should be said that the Fragmentation Functions depend on the quark from which they originate, hence the index.

2.4 Jets

Due to the non-perturbative nature of the hadronization process and the myriad of particles produced, it is not possible to calculate cross sections for all of the different possible hadronic final states. Instead, *jets* are considered in their place. Jets are, as the name suggests, groups of collimated particles which all originate from the same parent particle. They can be treated as singular objects, allowing for a notable simplification of the calculations.

2.4.1   Jet Definitions

In order to be able to precisely calculate jet cross sections and confront these with experimental data, it is necessarily to unambiguously *define* what a jet is. There are multiple possible definitions of jets, and we will briefly illustrate some examples of these.

Cone Algorithms

Historically, cone algorithms were the first class of jet algorithms introduced. The very first algorithm was used to classify jets in e^+e^- collisions and depended on two arbitrary parameters, δ and ϵ. If an event had a fraction of energy of at least $1 - \epsilon$ concentrated within two cones of half-angle δ, then that even was said to contain two jets [23]. The fact that the definition relied on the arbitrary parameters δ and ϵ meant that these had to be specified so that the predictions could be compared to data. This remains a general feature of jet algorithms to this day.

Cone algorithms have progressed in the years since their inception. Today, two of the most widely used code algorithms are iterative and fixed cone algorithms. In iterative cone algorithms, the direction of the jet is initially set by a particle i. All particles within distance R_{ij}

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 < R_{ij}^2 \tag{2.26}$$

in the rapidity-azimuthal angle plane are then taken as part of the jet, and their momenta are summed. The result of this summation is then used as the new seed, and the process is iterated until the jet cone is stable. To fully specify the algorithm, one must choose how to take the seed, and what to do when jets overlap.

Fixed cone algorithms function similarly. In this case, rather than iterating the cone direction, a cone is fixed around a seed, and that cone is called a jet. The particles within the radius R_{ij} are assigned to the jet, and removed from the event record. The algorithm proceeds until all possible jets have been identified.

k_t-Algorithms

k_t algorithms are part of a family of algorithms known as sequential recombination jet algorithms. The k_t algorithms use a momentum-weighted distance

$$\begin{cases} d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2} \\ d_{iB} = k_{ti}^{2p} \end{cases} \tag{2.27}$$

to establish which particles j lie closest to the particle i. If d_{ij} is less than the distance between i and the beam d_{iB}, i and j are combined into a jet. This procedure iterates over all particles.

These definitions also depend on two parameters: p and R. There exist three noteworthy cases: $p = 1$ is known as the k_t algorithm, and weighs the distance d_{ij} using the square of the transverse momentum of the softer particle; $p = 0$ is known as the Cambridge-Aachen algorithm, and features no weighting; $p = -1$ is known as the anti-k_t algorithm, and weights the distance using the inverse of the square of the transverse momentum of the harder particle.

Due to the distance used, the anti-k_t algorithm tends to cluster soft particles together with hard particles. This is a useful property as it tends to lead to jets centered about hard particles, and correctly recombines the soft radiation emitted from the hard seed together with that seed. If the hard particles are well-separated, this also leads to conical jets, as shown in Figure 2.6.
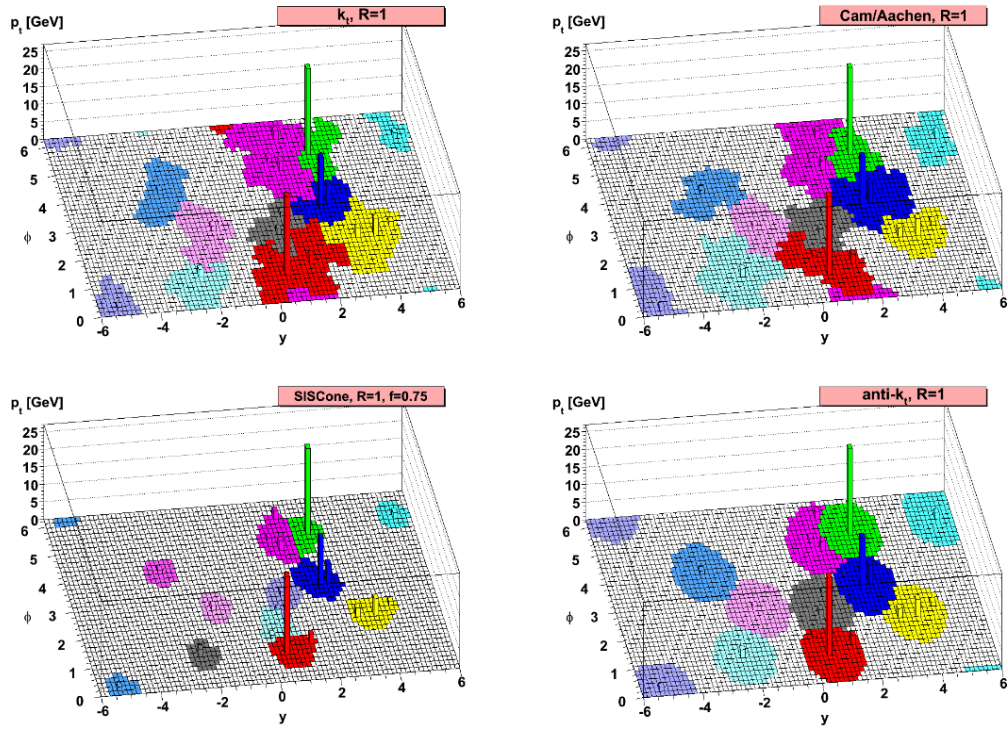
Figure 2.6: A representation of the jets formed from the same event using the three main k_t algorithms as well as SISCone, a commonly used cone algorithm. The jets clustered using anti-k_t are conical, and centered around the hard particles [24].

2.4.2   Jet Cross Sections

Once we have defined our jets, we can go on to calculate jet cross sections. Without going into too much detail, we will just illustrate a general feature which highlights the usefulness of jets.

If we calculate the two-hadron semi-inclusive cross section for e^+e^- into hadrons, we find that

$$\frac{d\sigma}{dy\,dz_1\,dz_2} = N_c \frac{\pi\alpha^2}{Q^2} \left(1 + \cos\theta\right) \sum_f e_f^2 D_f(z_1) D_f(z_2), \tag{2.28}$$

where again the sum over the flavors f runs over both quarks and antiquarks. N_c stands for the number of colors that can be produced, and θ is the angle between the momentum of the quarks produced and the momentum of the colliding leptons.

If rather than observing the two final-state hadrons we observe the jets surrounding them, the Fragmentation Functions are replaced by δ-functions, $\delta(1 - z_1)$ and $\delta(1 - z_2)$. After integrating, we find that

$$\frac{d\sigma^{jets}}{dy} = N_c \frac{\pi\alpha^2}{Q^2} \left(1 + \cos\theta\right) \sum_f e_f^2. \tag{2.29}$$

This is exactly the QED cross section for e^+e^- annihilation! The use of jets allows us to notably simplify calculations and make accurate predictions, since we neither have to consider fully inclusive cross sections, nor a fully exclusive one.

2.5   Higher Order Corrections

A number of higher order corrections to the elementary cross section are possible. These include real and virtual corrections in all possible combinations. In QCD, we currently know how to treat corrections up to next-to-next-to-next-to-leading order for a few select processes [25, 26]. In this section we will briefly discuss the importance of higher order corrections, as well as some challenges which arise during their calculation.

2.5.1   Infrared and Collinear Divergences

The matrix element for a radiative correction in QED and QCD can be calculated using the Feynman rules. We will consider the example of the emission of a photon for $e^+e^- \to e^+e^-$ in QED, but all considerations also hold in QCD. The matrix element for all possible real emissions and all possible photon polarizations is found to be

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_0 \frac{\alpha}{2\pi} \left[2\frac{p_- \cdot p_+}{(p_- \cdot k)(p_+ \cdot k)} - \frac{m_e^2}{(p_- \cdot k)^2} - \frac{m_e^2}{(p_+ \cdot k)^2}\right] \frac{d^3k}{\omega} \tag{2.30}$$

where $\left(\frac{d\sigma}{d\Omega}\right)_0$ represents the tree-level cross section, without any emission, p_- and p_+ represent the momenta of the outgoing electron and positron, and k^μ the momentum of the radiated photon. In the center-of-mass frame, where the kinematic variables can be expressed as

$$\begin{cases} p_-^\mu = (E, 0, 0, \beta E) \\ p_+^\mu = (E, 0, 0, -\beta E) \\ k^\mu = (\omega, \omega \sin\theta \cos\phi, \omega \sin\theta \sin\phi, \omega \cos\theta) \end{cases} \tag{2.31}$$
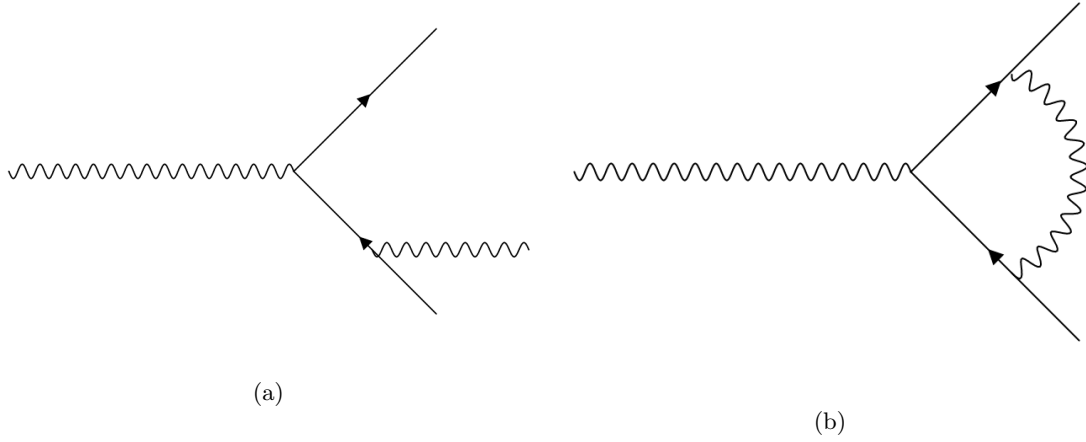
(a)

(b)

Figure 2.7: Feynman diagrams representing real (a) and virtual (b) corrections to the final state of e^+e^- annihilation in QED.

where $\beta = \sqrt{1 - \frac{m_e^2}{E^2}}$ and ω is the energy of the photon, equation (2.30) becomes

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_0 \frac{\alpha}{2\pi} \frac{1}{\omega} \frac{\beta^2 \sin^2\theta}{(1 - \beta^2 \cos^2\theta)^2} d\omega \, d\cos\theta \tag{2.32}$$

where we have already integrated in $d\phi$.

This expression is divergent in two cases: when the radiated particle is either very soft ($E \ll 1$) or collinear to the lepton from which it is emitted ($\theta \ll 1$). These cases are known as *infrared* and *collinear* divergences, respectively, and they are problematic as they prohibit the calculation of a finite cross section at next-to-leading order.

Integrating in $d\cos\theta$ resolves the collinear divergence

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_0 \frac{1}{\omega} \frac{2\alpha}{\pi} \left[\frac{\beta^2 + 1}{2\beta} \ln\left(\frac{1 + \beta}{1 - \beta}\right) + 1\right] d\omega \tag{2.33}$$

which, when $s \gg m^2$, simplifies to

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_0 \frac{1}{\omega} \frac{2\alpha}{\pi} \left[\ln\frac{s}{m_e^2} - 1\right] d\omega. \tag{2.34}$$

Lastly, we must integrate over all energies $d\omega$. However, because of finite detector resolution, the experimental cross section can only be sensitive to those photons radiated over a threshold ΔE. The experimental cross section contains two parts

$$\left(\frac{d\sigma}{d\Omega}\right)_{exp} = \left(\frac{d\sigma}{d\Omega}\right)_{elastic} + \left(\frac{d\sigma}{d\Omega}\right)_{\omega < \Delta E}, \tag{2.35}$$

one corresponding to the tree-level cross section, and the other with a radiative correction with $\omega < \Delta E$. When we integrate over $d\omega$, the second term on the right-hand size diverges as can be seen from (2.34).

Thankfully, a brilliant solution to this problem exists. If we assign a mass λ to the radiated photon, the (2.34) becomes

$$\left(\frac{d\sigma}{d\Omega}\right)_{exp} = \left(\frac{d\sigma}{d\Omega}\right)_{elastic} + \left(\frac{d\sigma}{d\Omega}\right)_{\omega<\Delta E} \ln \frac{\Delta E}{\lambda}. \tag{2.36}$$

If we also consider the virtual corrections and calculate the relative cross section, which also contributes to (2.35), we find

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_{0} \frac{1}{\omega} \frac{2\alpha}{\pi} \left[\ln \frac{s}{m_e^2} - 1\right] \ln \frac{\lambda}{E} d\omega \tag{2.37}$$

We can sum the real and virtual corrections, and find a total cross section independent from λ! The Kinoshita–Lee–Nauenberg theorem guarantees that this cancellation occurs at all orders in perturbation theory [27]. We can now safely take the limit of $\lambda \to 0$ and integrate over ω.

2.5.2   Infrared and Collinear Safety for Jets

As mentioned in the previous section, infrared and collinear (IRC) divergences will, in theory, exactly cancel. In practice, however, this is not always the case. Depending on the type of jet algorithm used, it may happen that the cancellation breaks and an infinite cross section is calculated. Obviously, this is a problem since the measured cross section is by definition finite. It is therefore important from a theoretical standpoint to define jets in a way that is IRC safe.

An observable $\mathcal{O}$ is said to be IRC safe when the following properties are satisfied:

$$\mathcal{O}(X; p_1, \ldots, p_n, p_{n+1} \to 0) \to \mathcal{O}(X; p_1, \ldots, p_n) \tag{2.38}$$

$$\mathcal{O}(X; p_1, \ldots, p_n \parallel p_{n+1}) \to \mathcal{O}(X; p_1, \ldots, p_n), \tag{2.39}$$

i.e. when the observable reduces to that of n particle case, in absence of the soft or collinear emission. An example of infrared safe and unsafe jets is shown in Figure 2.8, while Figure 2.9 illustrates an example of collinear safe and unsafe jets.

The anti-k_t algorithm is an example of a jet algorithm which is IRC safe. Due to its weighting, when a soft particle is emitted, this will tend to cluster together with the hard center of the jet. On the other hand, when a particle is emitted collinear to another, the distance Δ_{ij} will be very small, and it will again be clustered together with its parent particle.

2.6   Conclusions

Over the past few decades, colliders have allowed us to study, understand, and ultimately build the Standard Model. Notable achievements include the discovery of the Z boson through the observation of weak neutral currents at Gargamelle in 1973 [29, 30], the discovery of the W boson by the UA1 and UA2 collaborations in 1983 [31, 32], and the confirmation of the existence of three generations of matter at LEP in 1998 [33]. From here began an extensive campaign to measure the properties of these particles with increasing precision, which is still in course to this day [34, 35]. At the end of the '90s, only one key ingredient of the model had yet to be found: the Higgs boson.
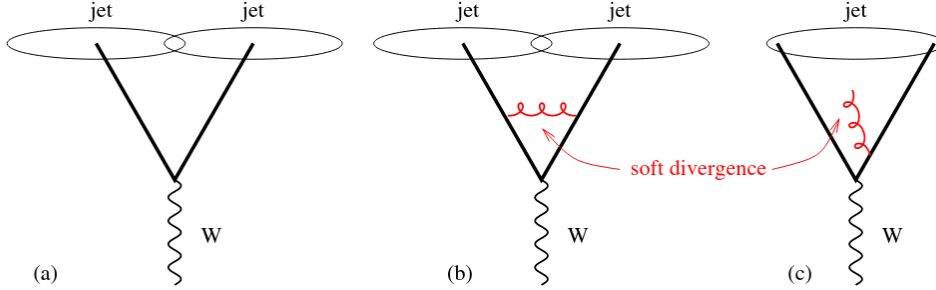
Figure 2.8: A schematic representation of an IR unsafe jet. The emission of a soft gluon changes the event from a two-jet event to a one-jet event due to its effect on the clustering algorithm [28].
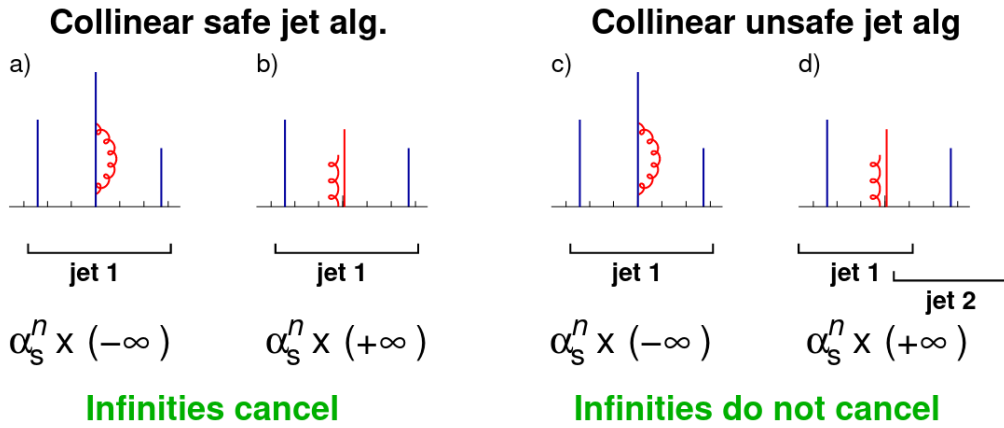

Figure 2.9: A schematic representation of a collinear safe and unsafe jet. The blue lines represent particles, and the length corresponds to the particle's transverse momentum. The horizontal axis represents rapidity. If the emission of a collinear gluon leads to the formation of a second jet, the cancellation is broken and the cross section diverges, as shown on the right. If, on the other hand, it and its parents are treated as a single particle by, for example, summing their momenta, the emission satisfies the definition given by Equation (2.38) [28].

Chapter 3

Higgs Physics

Over the past several decades, colliders have allowed us to study and understand the Standard Model. At the turn of the century, only one piece of the puzzle was left: the Higgs boson, expected as the remaining degree of freedom from the Higgs mechanism described in Section 1.4. The LHC was tasked with finding this elusive particle, and succeeded in 2012 [37, 38].

The discovery of the Higgs boson ushered in a new era of physics. Since then, one of the main goals of the ATLAS and CMS programs has been to robustly test the Standard Model by studying the properties of the Higgs. This requires knowledge on the production and decay mechanisms of the Higgs.

These mechanisms could be studied as a function of the Higgs mass m_H even prior to the discovery of the Higgs. Now that m_H is known, all the other parameters are fixed and the relevant total cross sections depend exclusively on $\sqrt{s}$.

3.1   Higgs Production Mechanisms

At the LHC, there are four different processes which can produce the Higgs. Their cross sections at $\sqrt{s} = 13$ TeV as a function of mass are shown in Figure 3.1, and a short summary is provided below, in the order of decreasing cross section.

Gluon-Gluon Fusion

Gluon-gluon fusion is the process which dominates Higgs production at the LHC. In the Standard Model, there is no direct coupling between the Higgs and the gluon, though an indirect coupling is possible through a virtual top quark loop, as shown in Figure 3.2a. The top quark loop is favored due the high mass of the top, though there is a small contribution from the bottom.

Vector Boson Fusion

Vector boson fusion is the second most production mechanism, responsible for about 10% of the total cross section. This process is interesting because it proceeds via a pure electroweak exchange.

Figure 3.2b shows the Feynman diagram for the process. The initial-state quarks deviate only slightly from their initial direction, leading to two jets near the beam axis in opposite regions of the detector. This provides a typical experimental signature of the process.
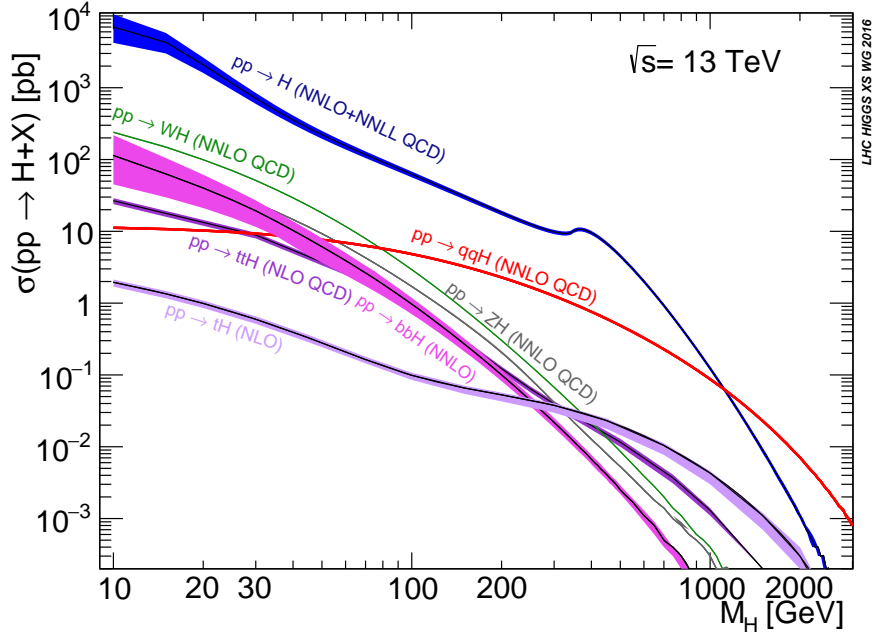
Figure 3.1: The cross sections of various Higgs production mechanisms as a function of m_H at current LHC energies. The thickness of the lines represent various theoretical uncertainties [39].

Higgstrahlung

Higgstrahlung (named after bremsstrahlung) is another process which relies on the Higgs coupling to two vector bosons. In this case, the Higgs radiates off a virtual Z or W. Experimentally, this channel can be identified by the decay of vector boson into leptons. Figure 3.2c shows the Feynman diagram for this process.

Heavy Quarks Associated Production

The last production channel of the Higgs is $q\bar{q} \rightarrow q\bar{q}H$, as shown in Figure 3.2d. Once again, since the Higgs coupling favors heavy masses, the top quark dominates this channel. However, the coupling is also possible for other heavy quarks, such as the bottom. This channel provides a direct way to measure the Yukawa coupling.

3.2   Higgs Decay Mechanisms

As opposed to the Higgs production mechanisms, which depend in part on the structure of the hadrons used in collisions, Higgs decay mechanisms depend exclusively on the properties of the Higgs. The branching ratios of the various decay channels, again as a function of m_H, are shown in Figure 3.3, while the Feynman diagrams representing the main decay channels are shown in Figure 3.4. Since they are many, we will only focus on a few key processes.
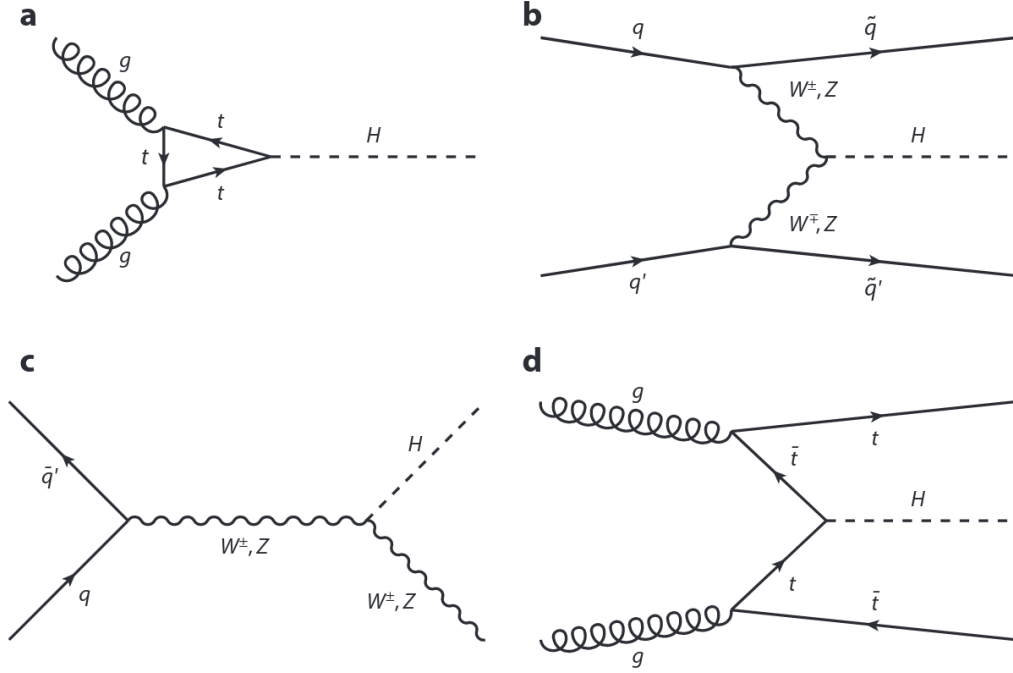
Figure 3.2: Higgs production in association with heavy quarks.

$H \to b\bar{b}$

The decay of the Higgs to $b\bar{b}$ is by far the most important decay channel. Despite the fact that the Higgs couples more strongly to the top, the decay would only be possible if m_H were above ~ 300 GeV because of conservation of energy.

The extremely large branching ratio means that, experimentally, it is of utmost importance to be able to observe this channel if we want to study the properties of the Higgs in detail. To this aim, flavor tagging algorithms are a fundamental tool in determining whether a jet originates from a b-quark, a c-quark, τ, or a light-quark or gluon.

A jet originating from a heavy quark can be identified based on some unique properties which stem from the high mass of the quark. For example, the relatively long lifetime of b-hadrons, of the order of 1.5 ps, combined with the high energies involved result in the formation of a secondary vertex, which can distance anywhere from a few hundred μm up to ~ 1 cm from the primary vertex. This secondary vertex gives rise to displaced tracks with respect to the primary vertex, from which the secondary vertex can be recognized. It is also possible to measure the "mass" of the secondary vertex, which will be related to the mass of the b-hadron. The decay products of a b-hadron are characterized by a large transverse momentum with respect to jet axis when compared to other jet constituents. In addition to this, in about 20% of cases, the decay of the b-hadron at the secondary vertex results in an soft electron or muon, whose properties allow for the selection of a pure sample of b-jets. The combination of this information is used as input to algorithms, which can either be classical or make use of machine learning techniques, in order to determine with what likelihood a given jet is a b-jet. Figure 3.5 illustrates the geometry of these secondary vertices.

Because these tagging techniques are general, similar considerations hold for $H \to \tau^+\tau^-$ and
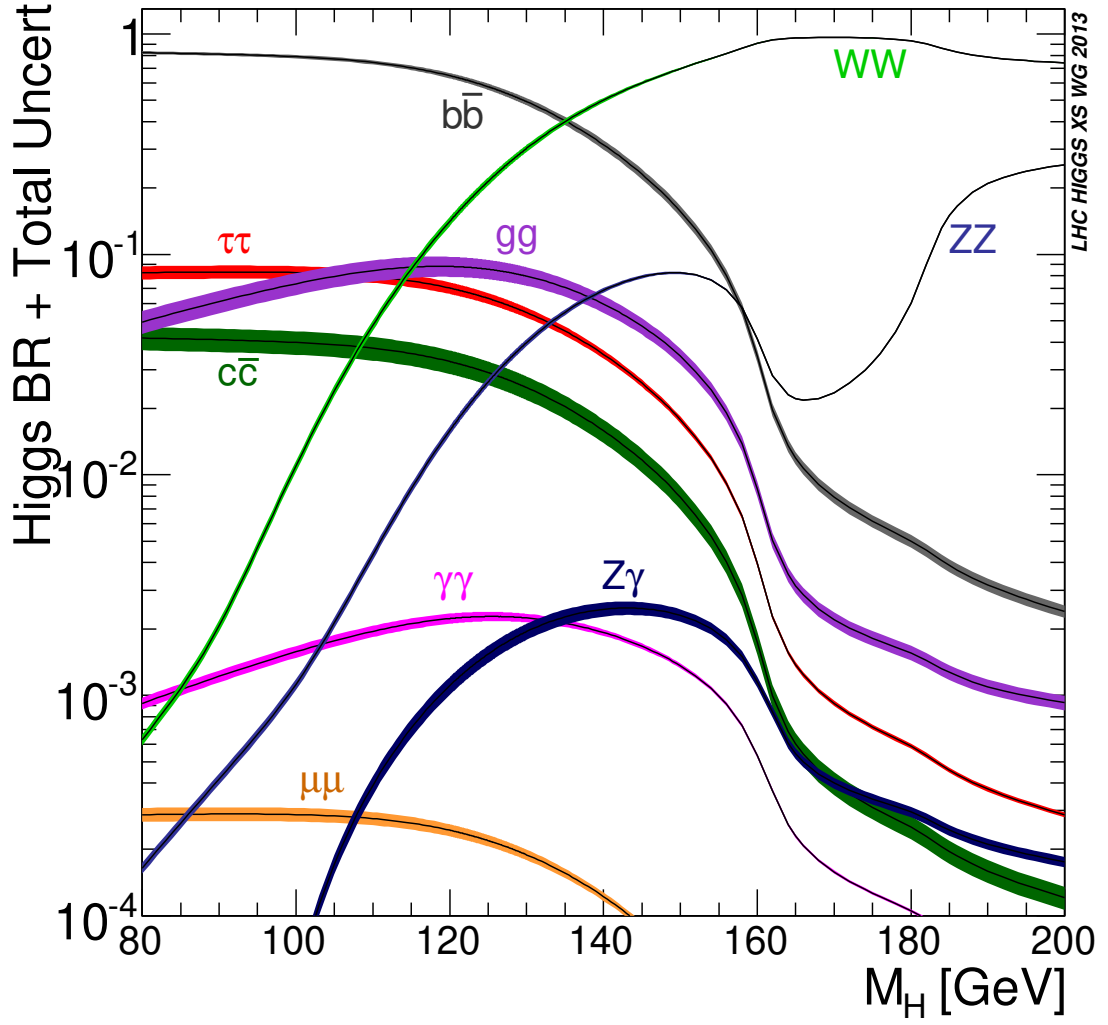
Figure 3.3: The branching ratios of the various decay channels of the Higgs boson as a function of m_H. The width of the lines represents theoretical uncertainties [40].
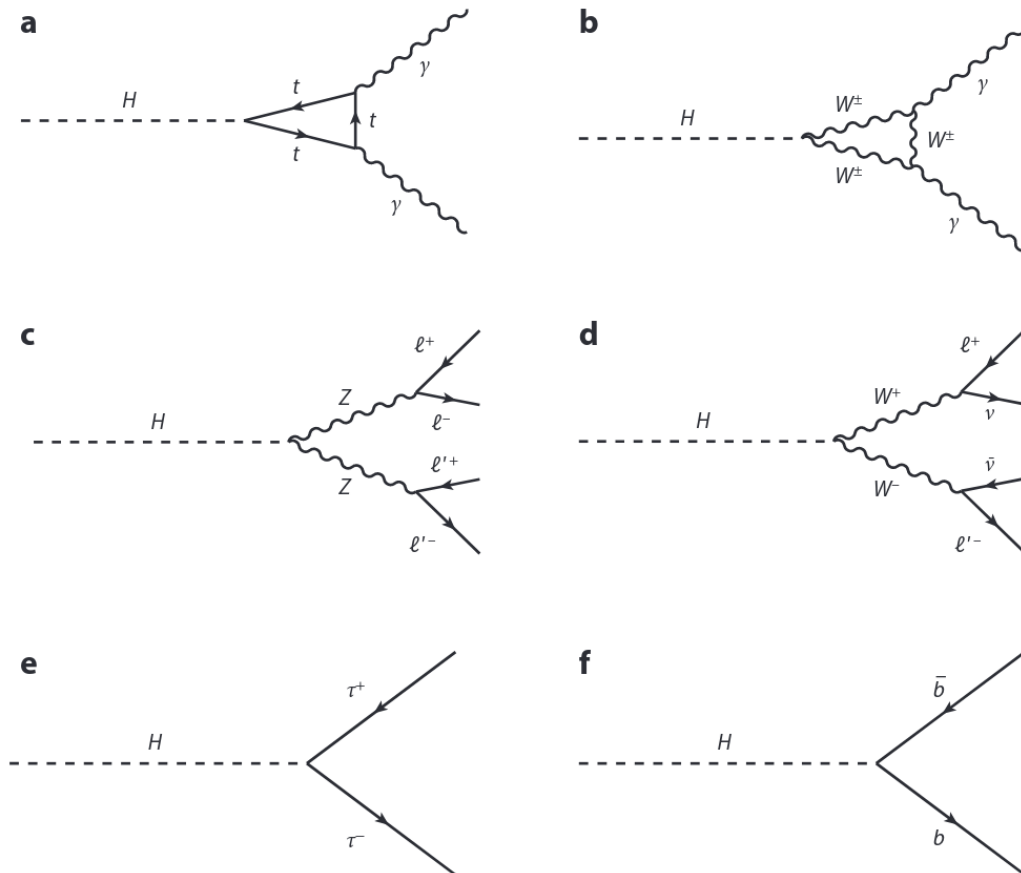
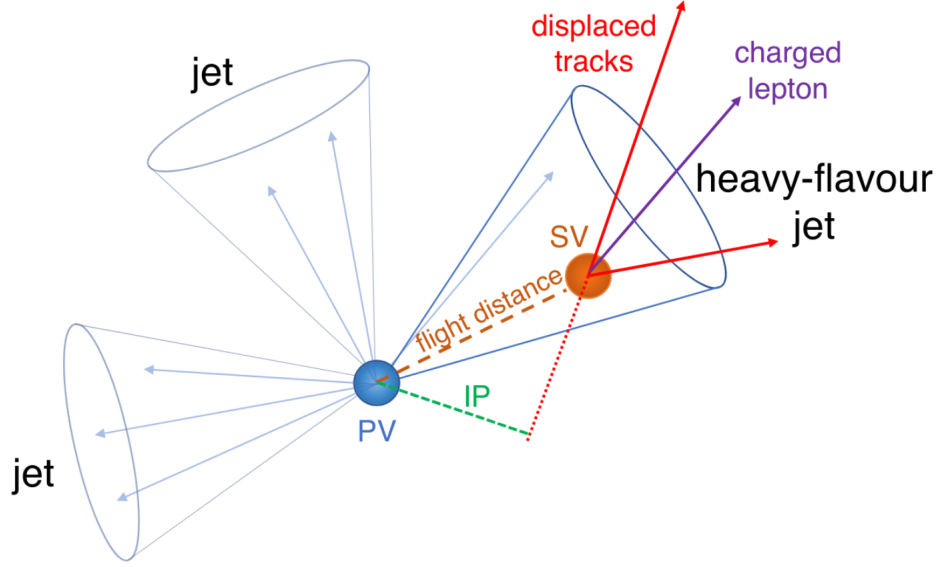Figure 3.4: The Feynman diagrams of the principal decay channels of the Higgs boson.

Figure 3.5: A schematic diagram of the secondary vertex present in decays of b-hadrons and the variables used for b-tagging [41].

$H \to c\bar{c}$. Although $H \to gg$ also leads to two jets, these cannot be effectively discriminated from the QCD background.

$H \to WW, H \to ZZ, H \to \gamma\gamma$

The large QCD background is the foremost obstacle that physicists must contend with when making discoveries at the LHC. For this reason, the decay channels $H \to WW, H \to ZZ$ and $H \to \gamma\gamma$ were crucial in the discovery of the Higgs boson. These channels are characterized by a clean experimental signal, thanks to the possibility of leptonic decays and the ease with each photons can be identified. Specifically, the channels $H \to ZZ^{(*)} \to 4\ell$, $H \to \gamma\gamma$, $H \to WW^{(*)} \to e\nu\mu\nu$, when combined with data from the channels $H \to b\bar{b}$ and $H \to \tau^+\tau^-$ allowed for the discovery of the Higgs in 2012 by the ATLAS and CMS collaborations. The infamous plot showing the discovery of the Higgs in the $\gamma\gamma$ channel is shown in Figure 3.6.

3.3   Properties of the Higgs

New particles do not come with a label. How can we be certain that the particle discovered in 2012 is the Standard Model Higgs? Weinberg's model predicted several properties of the Higgs, however others such as the particle's mass and width were merely constrained by previous experiments. In this section we shall review the Higgs' properties.

Mass

Several channels offer direct access to the mass of the Higgs. In particular, the ZZ and $\gamma\gamma$ decay channels allow physicists to measure the mass with an error of just 0.5 GeV. Using Run 1 data, ATLAS measured the Higgs mass to be $m_H = 125.37 \pm 0.37(\text{stat.}) \pm 0.18(\text{syst.})$ [42], while CMS
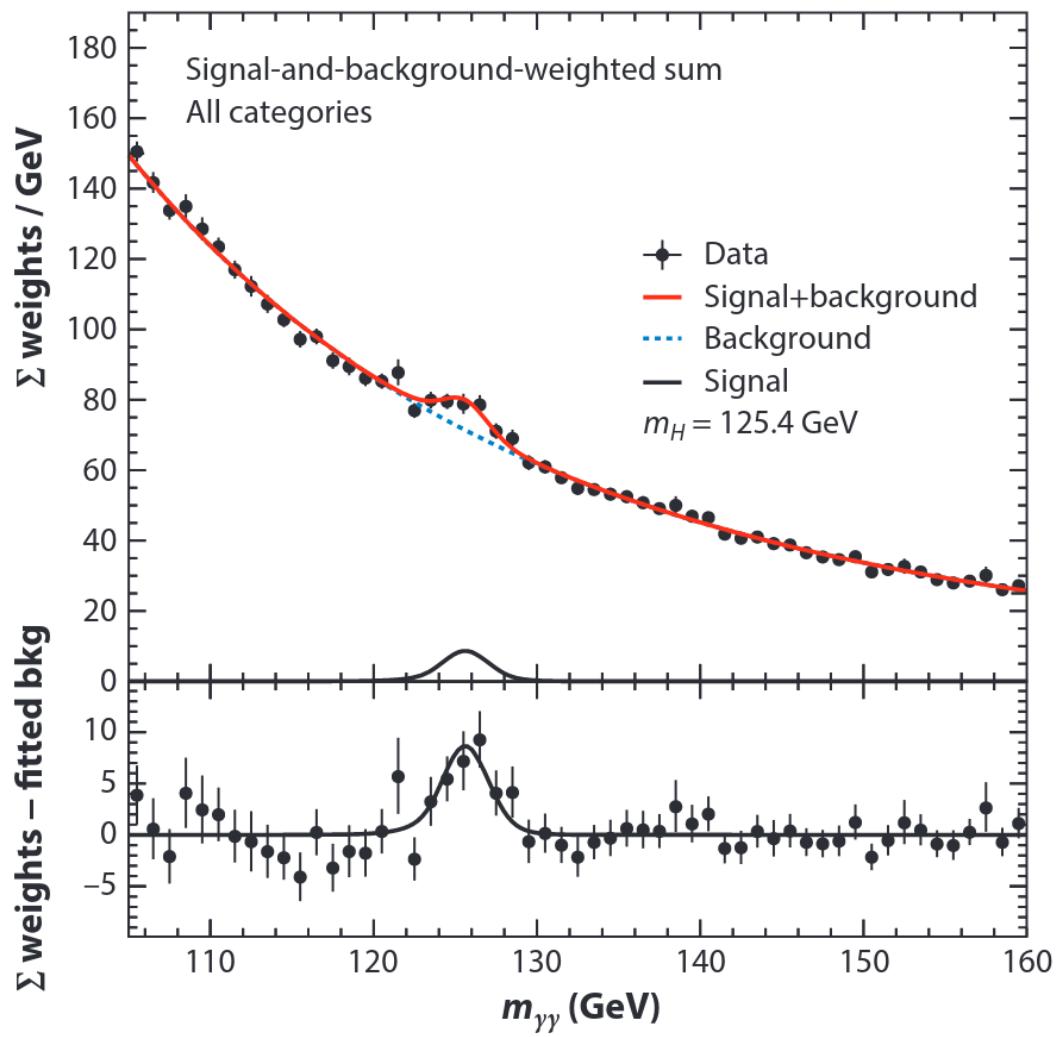
Figure 3.6: The observation of a signal peak above the background distribution at 125 GeV in the $\gamma\gamma$ channel at ATLAS [37].

found $m_H = 125.02^{+0.26}_{-0.27}$(stat.)$^{+0.14}_{-0.15}$(syst.) GeV [43]. The combination of these measurements is $m_H = 125.09 \pm 0.21$(stat.) ± 0.11(syst.) GeV [44]. There was no Standard Model prediction for the mass of the Higgs; this measurement represents a constraint for the model.

Width

The Higgs boson width is accessible from a number of measurements. One of the most recent measurements has been taken by the CMS collaboration. They looked at on-shell and off-shell H production and studied in particular the 4ℓ final state to find a width of $3.2^{+2.8}_{2.2}$ MeV, assuming Standard Model couplings [45]. This measurement is in agreement with the Monte Carlo prediction cited by the collaboration, $4.1^{+5.0}_{-4.0}$ MeV.

Couplings

It is also of interest to measure the couplings of the Higgs bosons to the fermions and vector bosons. The Standard Model predicts the these to be proportion to m_f and m_V^2, respectively. Loop decays play a special role in determining the fermionic couplings since the decay rate depends on the strength of the coupling. Otherwise, it is also possible to access the couplings to the vector bosons, specifically the W, through the branching ratios of $H \rightarrow WW$ decays and interference with $H \rightarrow \gamma\gamma$.

Deviation from the predicted value of the couplings is measured through κ_f and κ_V. These quantities indicate the ratio of the measured coupling to the predicted value

$$\kappa^2 = \frac{\Gamma}{\Gamma_{SM}}. \tag{3.1}$$

Figure 3.7a shows the measured limits of the scale factors by CMS through several different channels using data from Run 1. As can be seen, the measured values are compatible with the Standard Model prediction. Figure 3.7b instead shows the mass dependence of the couplings as measured by CMS. This too follows the Standard Model prediction.

Spin-Parity

The Standard Model predicts that the Higgs boson has a 0^+ spin-parity state. Since all observed final states are characterized by integral spin, it is clear that the particle from which they originate must also exhibit this property. However, several models had to be tested to verify the Standard Model prediction, as well as ensure that the observed particle was the same throughout all channels.

It is possible to access spin-parity information through the $\gamma\gamma$, WW, and ZZ channels [47–49]. In particular, either the direction of emission of the photons or the angular information from the decays of the vector bosons are considered. Experimentally, the information available from the WW channel is diluted with respect to that available from the ZZ channel due to the momentum lost to the neutrinos. Figure 3.8 shows the angular variables considered in the $pp \rightarrow X \rightarrow ZZ \rightarrow 4\ell$ channel.

All observations are consistent with a particle with spin-parity consistent with the Standard Model prediction. When combined with information coming from the production rate and couplings, physicists and confidently assert that the particle observed in 2012 was indeed the Higgs boson.
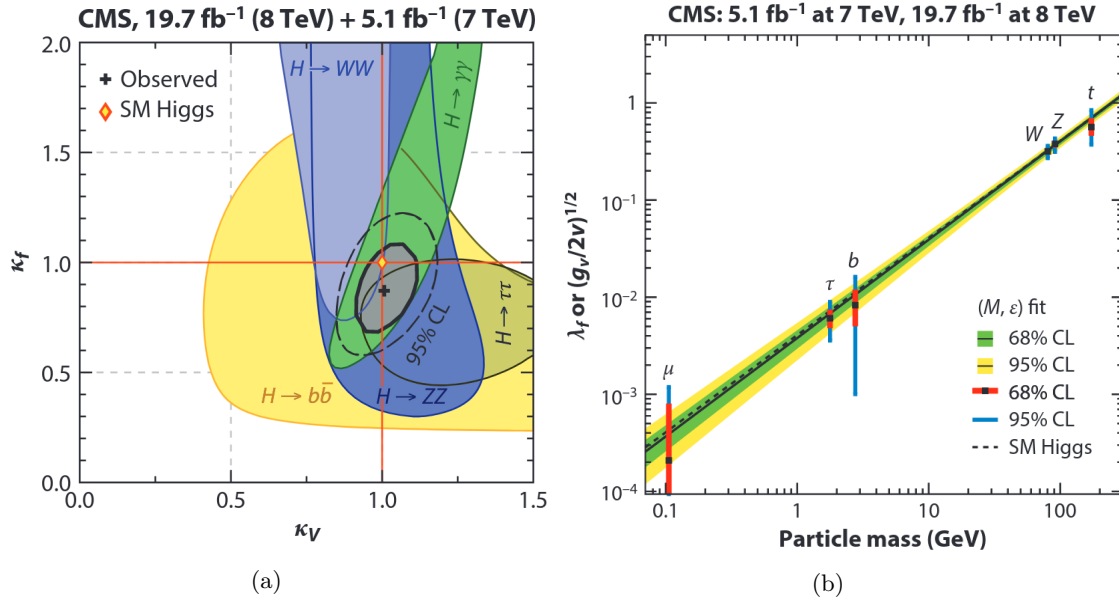
(a)   (b)

Figure 3.7: The measured limits (a) of the scale factors κ_f and κ_V and the mass dependence of the Higgs couplings (b) as measured by CMS [46].
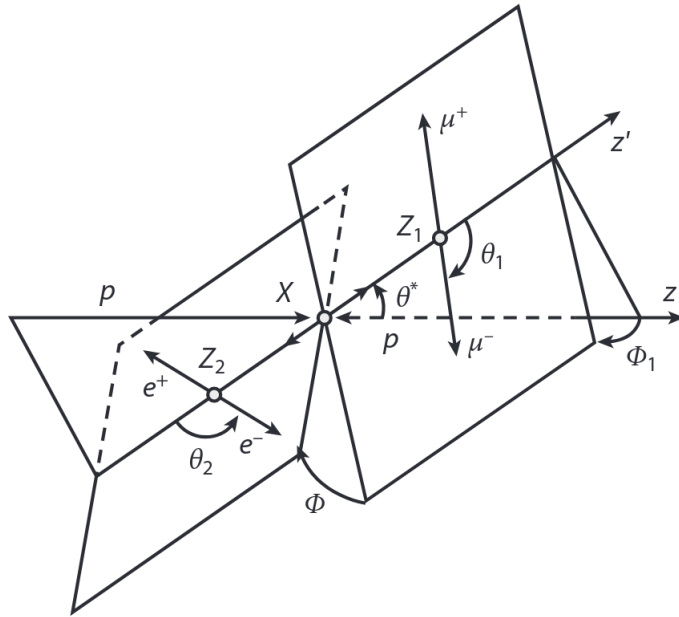

Figure 3.8: The angular variables considered in the ZZ channel to measure the spin-parity state of the parent particle [46].
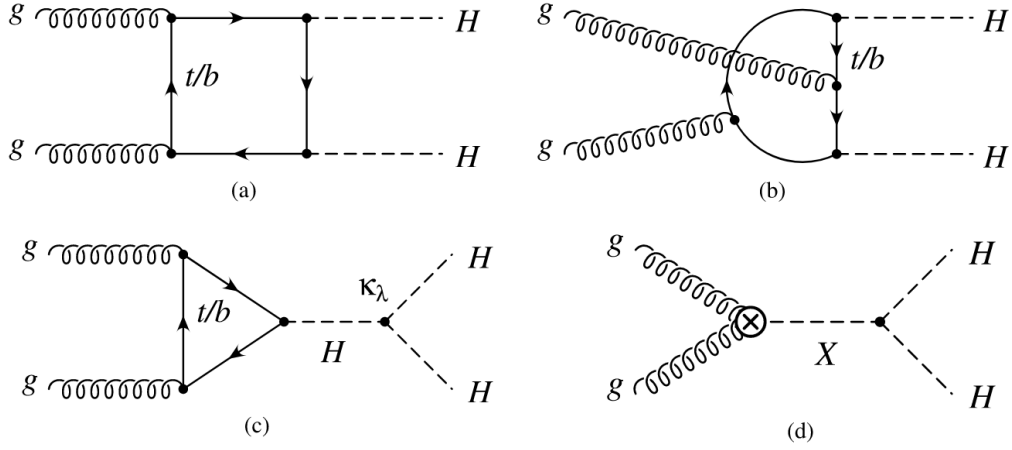
Figure 3.9: The leading-order Feynman diagrams for double Higgs production [51].

3.4   New Discoveries

Having confirmed the existence of the Higgs boson, physicists have begun to rigorously test all processes involving this particle and search for new, unknown ones.

Double Higgs Production

Within the field of precision physics, some Higgs couplings have yet to be tested. In particular, the Higgs self-coupling remains to be precisely measured at ATLAS and CMS.

Due to the small Higgs production cross section, double Higgs production is an incredibly rare event. However, it is expected that by the end of Run 3, this process should be observed [50].

The process is of interest due to possible cosmological implications of a deviation from the Standard Model prediction, and because it is a probe of physics beyond the Standard Model. Specifically, there are models which predict decays of spin-2 Kaluza-Klein gravitons or other, heavier Higgs bosons into a Higgs boson pair. Figure 3.9 shows the leading order Feynman diagrams for two Higgs bosons produced via gluon-gluon fusion.

Currently, ATLAS has placed limits on the scale factor κ_H which measures the deviation from the expected Standard Model couplings. These limits come from constraints placed by single Higgs production. At a 95% confidence level, κ_H is limited to the range $-5.0 < \kappa_H < 12.0$ [51].

Dark Matter Searches

Over the past few decades, cosmological observations have led to the conclusion that visible matter makes up just a small part of our universe. The rest is made up of what is known as dark matter and dark energy.

We are sure of the existence of dark matter because we can see its gravitational influence. The gravitational interactions of dark matter are indicative of the fact that it has mass. There are several theories as to the nature of dark matter, but one of the leading hypotheses is the particle nature of dark matter.

If dark matter is a massive particle, it is natural to assume that it gains mass via the Higgs mechanism. This idea has led to the search for production of unknown, invisible particles at colliders via their interaction with the Higgs boson.

Without diving into too much detail, dark matter searches at colliders in general rely on the same principle as precision physics: precise measurements of known cross sections to search for any possible deviations from Standard Model predictions. Given how little we know about dark matter, these Higgs portals are one of the most promising avenues to find out more about the invisible universe [52].

3.5 Conclusions

The discovery of the Higgs boson finally completed our understanding of the Standard Model. However, there are still many pieces of the puzzle that are missing, such as dark matter. To gain a full understanding of the model, we must be able to produce, reconstruct and detect the Higgs boson at current and future colliders.

Detection of $H \to b\bar{b}$ decays is an important part of this mission, as it is the main decay channel of the Higgs boson. Due to the large QCD background of the process, it is hard to discern events containing the Higgs boson within the QCD haystack. In the next chapter, we will describe a framework which we have developed to help overcome this hurdle.

Chapter 4

A Study of Color Sensitive Variables

Due to its large branching ratio, the $H \to b\bar{b}$ decay is fundamental to unveiling the properties of the Higgs coupling to fermions. In this chapter, we will describe a method that we have originally developed to discriminate b-jets originating from a color singlet, such as the Higgs boson, from those originating from a color octect.

We have used simulated data to train two different machine learning algorithms on high-level, color sensitive variables which have been introduced in literature. We will first introduce these variables, and then go on to discuss the details of the simulation, the analysis of the simulated events, and our findings.

4.1 The Observables

In total, this method makes use of eight variables existing in literature and presented in [53–55]. These are $t_{\parallel a}, t_{\perp a}, \theta_{pa}, t_{\parallel b}, t_{\perp b}, \theta_{pb}, D_2$ and the color ring $\mathcal{O}$.

4.1.1 The Pull Variables

The first variables we will consider are related to the *pull vector*. With reference to Figure 4.1, consider an event in which two jets, J_a and J_b, with centers (y_a, ϕ_a) and (y_b, ϕ_b), respectively, are emitted. The pull vector $\vec{t}$ relative to jet J_a is defined as

$$\vec{t} = \frac{1}{p_{ta}} \sum_{i \in J_a} p_{ti} |\vec{r}_i|^2 \hat{r}_i \tag{4.1}$$

where p_{t_a} is the transverse momentum of the jet, and the sum runs over all the the jet constituents. y and ϕ represent rapidity and azimuthal angle, and $\vec{r}_i$ is the distance vector between the jet and its i-th constituent in the y-ϕ plane

$$\vec{r}_i = (y_i - y_a, \phi_i - \phi_a). \tag{4.2}$$

In particular, we would like to consider the projections of $\vec{t}$ along two lines: one, generated by the unit vector which points from the center of J_a to the center of J_b

$$\hat{n}_{\parallel} = \frac{1}{\sqrt{\Delta y^2 + \Delta \phi^2}} (\Delta y, \Delta \phi), \tag{4.3}$$
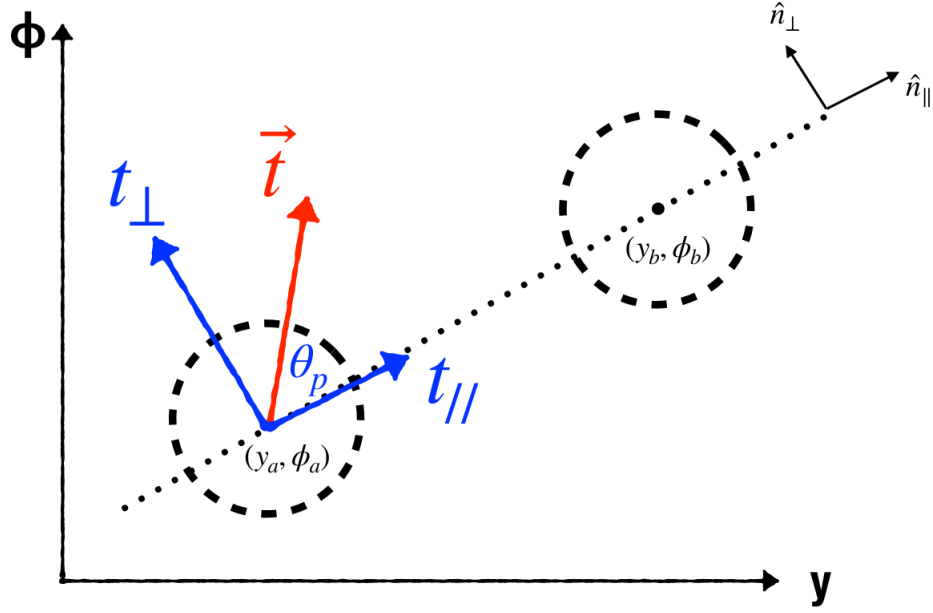
Figure 4.1: The construction of the pull vector in the (y, ϕ) plane [53].

and the other generated by the unit vector perpendicular to $\hat{n}_{\parallel}$

$$\hat{n}_{\perp} = \frac{1}{\sqrt{\Delta y^2 + \Delta \phi^2}} \left(-\Delta \phi, \Delta y \right), \tag{4.4}$$

i.e.

$$t_{\parallel} = \vec{t} \cdot \hat{n}_{\parallel} \tag{4.5}$$

$$t_{\perp} = \vec{t} \cdot \hat{n}_{\perp} \tag{4.6}$$

We would also like to consider θ_p, known as the pull angle, defined as

$$\theta_p = \arccos \frac{t_{\parallel}}{|\vec{t}|}. \tag{4.7}$$

The pull vector is sensitive to the different color connections which are present in the decay of a color singlet and the decay of a color octet.

An illustration of this feature is shown in Figure 4.2. Because of the different color connections, the radiation from jets originating from a singlet decay will tend to be emitted between the two jets, causing the pull vector of J_a to point in the direction of J_b and vice-versa. In the octect case, the pull vectors will instead tend to point in different directions.

Of all the variables considered, the pull angle has been shown to be one of the most effective discriminators of the two different color configurations [56]. However, there was a noticeable mismatch between theoretical predictions for θ_p and experimental measurements [57].

The theoretical difficulties in calculating θ_p stem from the fact that it is not an IRC safe variable. However, $t_{\parallel}$ and $t_{\perp}$ *are* IRC safe observables.
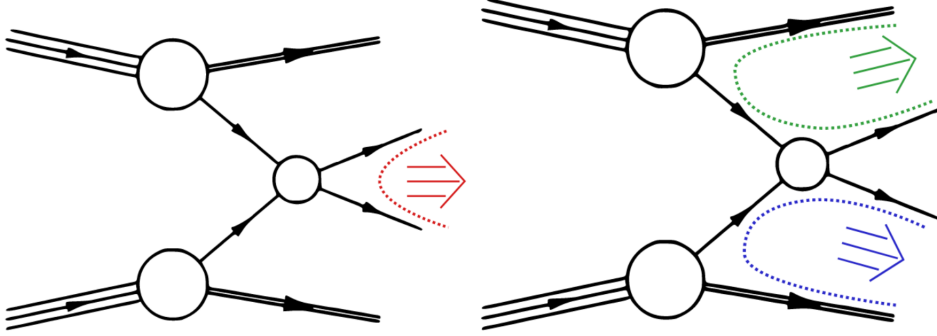
Figure 4.2: A representation of the color flow for the decay of a color singlet (left) vs. a color octect (right) at a collider experiment. In the case of a color singlet, the two colored particles stemming from the decay are color-connected only to each other, while for the case of an octect, the particles are color-connected to the rest of the proton, as it is from there that their color originates [56].

Our method includes these two variables for this purpose. Although it is not IRC safe, θ_p is also included. We are aware of a potential overlap in including both $t_\parallel$, $t_\perp$ and θ_p. However, this is not an issue since the machine learning algorithms are robust against interdependence between input variables.

We consider these three variables for both jets J_a and J_b.

4.1.2 D_2

To introduce the observable D_2, we must first familiarize the reader with energy correlator functions (ECFs) [58]. These are a class of observables sensitive to jet substructure. Specifically, they are sensitive to a the number of prongs of a jet, or how many distinct subjets compose a given jet. This is particularly useful in the boosted regime, defined by $p_T \gg 1$, where kinematics forces jets, which would otherwise be separated, to come together, as shown in Figure 4.3. The $(N+1)$-point correlator is used to determine if a jet has N prongs.

The definition of the N-point ECF is based on the the p_T and angular distance between the components of a jet:

$$ECF(N,\beta) = \sum_{i_1 < i_2 < \cdots < i_N \in J} \left(\prod_{a=1}^{N} p_{Ti_a} \right) \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^{N} R_{i_b i_c} \right)^\beta , \tag{4.8}$$

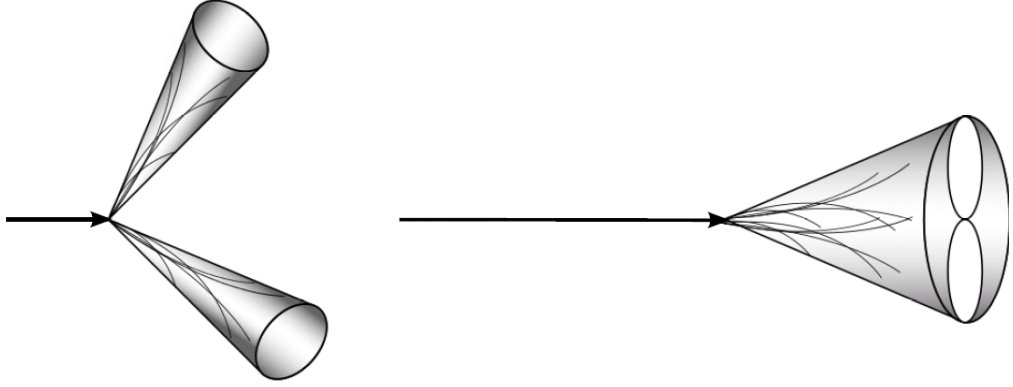where β is an arbitrary parameter. If $\beta > 0$, the observable is IRC safe. For the sake of clarity,

Figure 4.3: An illustration of a 2-prong jet caused by boosted kinematics [59].

let us write out the first few ECFs:

$$ECF(0, \beta) = 1$$

$$ECF(1, \beta) = \sum_{i \in J} p_{Ti}$$

$$ECF(2, \beta) = \sum_{i < j \in J} p_{Ti} p_{Tj} R_{ij}^{\beta}$$

$$ECF(3, \beta) = \sum_{i < j < k \in J} p_{Ti} p_{Tj} p_{Tk} \left(R_{ij} R_{ik} R_{jk} \right)^{\beta}.$$

There exists also a normalized N-point ECF, $e_N^{(\beta)}$, which differs from the definition (4.8) by a factor proportional to the p_T of the jet

$$e_N^{(\beta)} = \left(\frac{1}{p_{TJ}} \right)^N ECF(N, \beta) \tag{4.9}$$

Clearly, $ECF(N, \beta) = 0$ when it is computed on a jet with fewer than N constituents. In addition to this, $ECF(N + 1, \beta)$ will be significantly smaller than $ECF(N, \beta)$ if calculated on a jet containing only N subjets. This naturally leads to the consideration of the following ratio:

$$r_N^{\beta} = \frac{ECF(N + 1, \beta)}{ECF(N, \beta)}. \tag{4.10}$$

The ratio r_N^{β} is useful to discriminate N-prong jets from $N + 1$-prong jets. The D_2 variable [54] is a variation on this ratio, specified to the case of 1-prong and 2-prong jets, and calculated using the normalized ECFs

$$D_2(\beta) = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}. \tag{4.11}$$

Due to the different color configurations, a large jet originating from the decay of a color singlet in the boosted regime will tend to exhibit a 2-prong substructure, and thus will tend to have a larger value of D_2. On the other hand, in the case of a color octect, the same decay will tend to have a 1-prong structure, leading to smaller values for the observable.

For the purpose of our study, we have chosen the value $\beta = 2$.

4.1.3 Color Ring

The final variable we have considered is known as the *color ring*. To derive the observable, we must consider the boosted decay of a color singlet into two jets, which, at the parton level, correspond to (anti)quarks or gluons. We will consider the decay into two hard partons and the subsequent emission of a low-energy gluon.

The matrix element describing this transition is as follows [55]:

$$|\mathcal{M}_S|^2 = -\mathbf{T}_\alpha \cdot \mathbf{T}_\beta \frac{n_a \cdot n_b}{(n_a \cdot k)(n_b \cdot k)}.$$ (4.12)

In this expression, n_a and n_b are the light-like vectors parallel to the hard jets, k is the impulse of the soft emission, and $\mathbf{T}_i$ is the color operator. The dependence on α_S has been neglected as it is not relevant at this level. In what follows, Latin indices are used to refer to kinematics while Greek indices refer to color.

We know that color must be conserved, and since we are dealing with a singlet decay, we must have

$$\mathbf{T}_\alpha + \mathbf{T}_\beta = 0.$$ (4.13)

This necessarily implies that $\beta = \overline{\alpha}$, and so

$$- \mathbf{T}_\alpha \cdot \mathbf{T}_\beta = \mathbf{T}_\alpha^2 = C_S \mathbf{1}$$ (4.14)

where C_S is Casimir operator for either the fundamental or adjoint representation of $SU(3)$, depending on the decay, and we have explicitly written the identity matrix.

Let us now consider the same matrix element for a background process, in which a color octect decays into two hard partons, with a subsequent soft emission in the boosted regime. Thanks to the boosted kinematics of the event, we can safely assume that the two hard partons are closer in angle to each other than to any other colored object, allowing for the factorization of the matrix element into two parts. The first describes the dipole formed by the hard partons and the emission of the soft radiation. The second contains additional contributions from the initial state of the event and from extra jets, if present. This term will be neglected since it will lead to a constant. We can thus write

$$|\mathcal{M}_B|^2 = \sum_{\alpha,\beta} \left[-\mathbf{T}_\alpha \cdot \mathbf{T}_\beta \frac{n_a \cdot n_b}{(n_a \cdot k)(n_b \cdot k)} - \mathbf{T}_\alpha \cdot \mathbf{T}_\gamma \frac{n_a \cdot \overline{n}}{(n_a \cdot k)(\overline{n} \cdot k)} - \mathbf{T}_\beta \cdot \mathbf{T}_\gamma \frac{n_b \cdot \overline{n}}{(n_b \cdot k)(\overline{n} \cdot k)} \right]$$ (4.15)

where $\overline{n}$ is a light-like vector antiparallel to the system. $\overline{n}$ has overall color γ which is fixed by the relation

$$\mathbf{T}_\alpha + \mathbf{T}_\beta + \mathbf{T}_\gamma = 0.$$ (4.16)

We can use this to simplify the expression (4.15), which becomes

$$|\mathcal{M}_B|^2 = \sum_{\alpha,\beta} \left[-\mathbf{T}_\alpha \cdot \mathbf{T}_\beta \frac{n_a \cdot n_b}{(n_a \cdot k)(n_b \cdot k)} + (\mathbf{T}_\alpha^2 + \mathbf{T}_\alpha \cdot \mathbf{T}_\beta) \left(\frac{n_a \cdot \overline{n}}{(n_a \cdot k)(\overline{n} \cdot k)} + \frac{n_b \cdot \overline{n}}{(n_b \cdot k)(\overline{n} \cdot k)} \right) \right].$$ (4.17)

The background matrix element has been written in terms of two color factors:

$$\begin{cases} C_B = -\mathbf{T}_\alpha \cdot \mathbf{T}_\beta \\ \tilde{C}_B = \mathbf{T}_\alpha^2 + \mathbf{T}_\alpha \cdot \mathbf{T}_\beta. \end{cases}$$ (4.18)
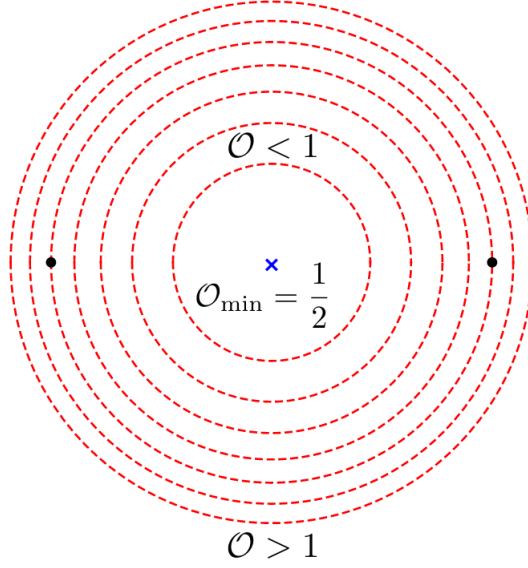
Figure 4.4: A geometric interpretation of the color ring. The two black dots represent the two hard jets, and the contour passing through them has value $\mathcal{O} = 1$. Depending on the direction of k, $\mathcal{O}$ will assume different values. If k is collinear to either a or b, $\mathcal{O}$ assumes its minimum value [55].

Again, the exact color factor depends on the process considered. We are interested in $H \to b\bar{b}$ as our signal process, which fixes $C_S = C_F$ and $C_B = C_F - C_A/2$ and $\widetilde{C}_B = C_A/2$. A decay into gluons, e.g. $H \to gg$ would fix these constants differently.

It is a well-established fact that the optimal variable to discriminate these two configurations is monotonic in their ratio [60]. The ratio turns out to be

$$\frac{|\mathcal{M}_S|^2}{|\mathcal{M}_B|^2} \simeq \frac{(n_a \cdot \overline{n})(n_b \cdot k)}{(n_a \cdot n_b)(\overline{n} \cdot k)} + \frac{(n_b \cdot \overline{n})(n_a \cdot k)}{(n_a \cdot n_b)(\overline{n} \cdot k)} \tag{4.19}$$

where we have left out the color factors as they are just constants. If we assume the collinear limit and utilize the small-angle approximation, this expression simplifies to

$$\frac{|\mathcal{M}_S|^2}{|\mathcal{M}_B|^2} \simeq \frac{\theta_{ak}^2 + \theta_{bk}^2}{\theta_{ab}^2} \equiv \mathcal{O}. \tag{4.20}$$

$\mathcal{O}$ is the definition of the jet color ring. This variable gets its name from its geometric interpretation, shown in Figure 4.4. Radiation from color singlets will tend to fall between the two jets, leading to values of $\mathcal{O} < 1$, while in the case of color octects, we will tend to have $\mathcal{O} > 1$.

Unfortunately, without any sort of momentum-weighting, (4.20) cannot be IRC safe. However, if we assume that the emission of the soft gluon leads to a distinct subjet within a larger jet, then IRC safety is recovered.
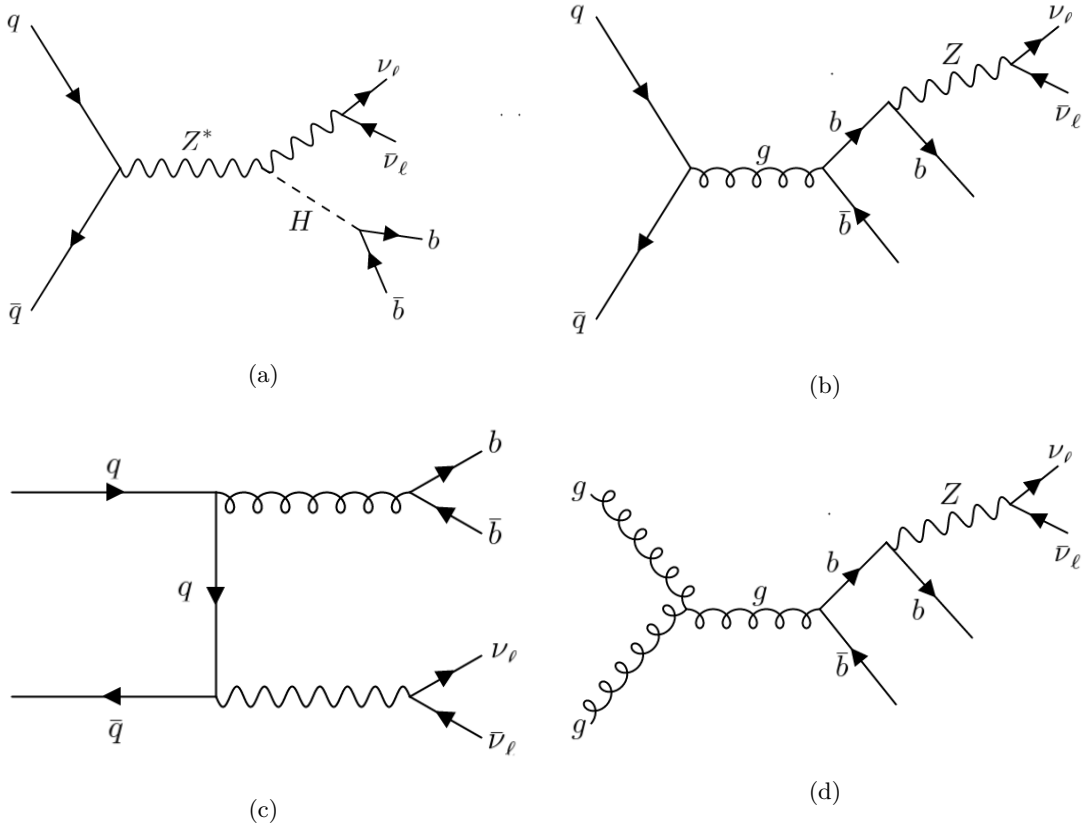
Figure 4.5: The Feynman diagrams constituting the signal (a) and background (b)-(d).

4.2 Simulation

The aforementioned variables are measured on events produced via Monte Carlo simulation. Using `MadGraph5_aMC` v2.8.3.2 [61], we generated two hard processes:

$$pp \to Z(\nu_\ell \overline{\nu}_\ell) H(b\overline{b}) \tag{4.21}$$

$$pp \to b\overline{b}\nu_\ell \overline{\nu}_\ell. \tag{4.22}$$

The first process, hereby referred to as the *signal*, represents the most likely decay of the Higgs boson, a color singlet. The second process, or the *background*, contains all QCD diagrams which lead to the same final state as the signal. Figure 4.5 shows the Feynman diagrams constituting the signal and the background. Both processes were generated with a 200 GeV cut on the p_T of the neutrino pair in the final state. This was done to ensure that the events generated were firmly in the boosted regime. The outputs of the parton-level simulations were saved as Les Houches Event Files [62].

These hard processes were subsequently showered in `PYTHIA8` v8.305 [63]. This program simulates the perturbative evolution of the event (e.g. radiation) as well as hadronization to produce particle-level events. The simulation included both Multi-Parton Interactions and underlying events. The output of the `PYTHIA8` shower was saved as a `HepMC3` file [64, 65]. `HepMC3` v3.2.2 was used.

Finally, rather than simulating an entire detector, `DELPHES` v3.5.0 was used to perform a fast detector simulation [66, 67]. This allowed us to understand how our method would perform in reality without having to run a computationally expensive full simulation, which in addition is strongly detector dependent. From `DELPHES`, we extracted both the Monte Carlo truth of the event, containing the particle-level information, as well as the reconstructed events including the `DELPHES`-smeared detector effects.

The `DELPHES` simulation was run using a version of the ATLAS card modified to fit our analysis standards. The output was stored as a `ROOT` file, and `ROOT` v6.22/06 was used [68].

In total, we simulated 1M signal and background events.

4.3   Analysis

We analyzed the simulated data in both the truth case and the reconstructed case. These cases differ in just one aspect: for the truth case, all stable particles with $p_T > 0.5$ GeV were clustered into jets. In the reconstructed case, the jets were clustered using the simulated calorimeter towers and tracks. All electromagnetic calorimeter towers with energy $E > 0.5$ GeV and significance $S > 2.0$ and all hadronic calorimeter towers with energy $E > 1.0$ GeV and significance $S > 2.0$ are considered. Tracks are required to have $p_T > 0.5$ GeV. The jets are clustered using `FastJet` v3.3.4 [69].

We can now discuss the details of the analysis. First, the constituents are clustered into Large Jets with radius $R = 1.0$. For each event, we choose the jet with the highest p_T as the Large Jet. We only accept the event if the jet has $p_T > 250$ GeV and $|y| < 1.5$. The restriction on y comes from experimental considerations: due to limitations related to the detector acceptance, b-tagging outside of this region comes with large errors.

We also cluster the constituents into smaller jets with radius $R = 0.2$. We identify those jets which distance $\Delta R < 0.8$ from the selected Large Jet, and call these subjets. We then proceed to identify the b-subjets which originate from the b-partons, through a process known as *b-labelling*. We do this by first identifying the b-partons originating from the hard scattering in the event record. For each b-parton, we find the closest subjet. If the b-parton has $p_T > 5.0$ GeV and the subjet is within $\Delta R = 0.2$ from the b-parton, we label the subjet as a b-subjet.

For the event to be accepted, we require that there be two unique b-labelled subjets with $p_T > 10$ GeV. The pull variables are calculated on these two b-labelled subjets, and D_2 is calculated on the Large Jet. For the color ring to be defined, there must also be a third non-b subjet within $\Delta R = 0.8$ from the Large Jet. In a majority of cases, this third jet is not present. To avoid discarding too many events, in these cases we assign a default value of $\mathcal{O} = -1$ to the color ring. This allows for higher statistics, but also provides useful information to the machine learning algorithms.

Table 4.1 shows the percentage of events which passed the selections in all cases considered. The cuts were more severe on the background events than the signal events in both the truth and reconstructed cases. Two cuts are significantly more important than the rest: the p_T cut on the Large Jet, responsible for about 60% of all cuts, and the rapidity cut on the Large Jet, responsible for another 10%.

Figure 4.6 shows the distributions of the eight variables for the truth case. In Figure 4.7, the same distributions are shown in the reconstructed case. In both cases, clear differences between signal and background allow for discrimination of the two. These differences are especially noticeable for the $\mathcal{O}$ and D_2 variables. On the other hand, we can see how θ_{pa} and θ_{pb} are effective discriminators in the truth case, but less so in the reconstructed case.
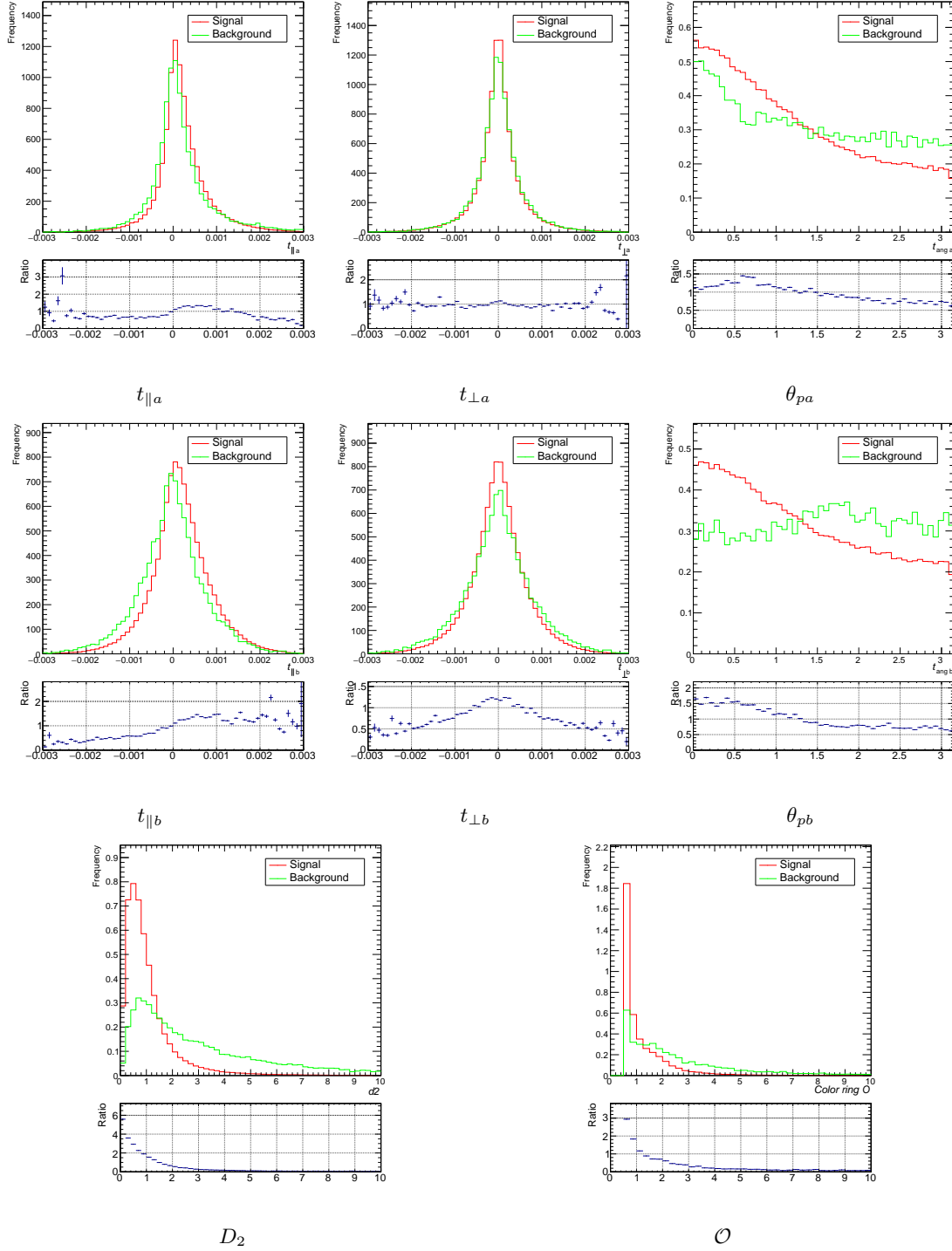
Figure 4.6: The distributions of the eight variables after all selection cuts in the truth case together with the ratio between signal and background events.
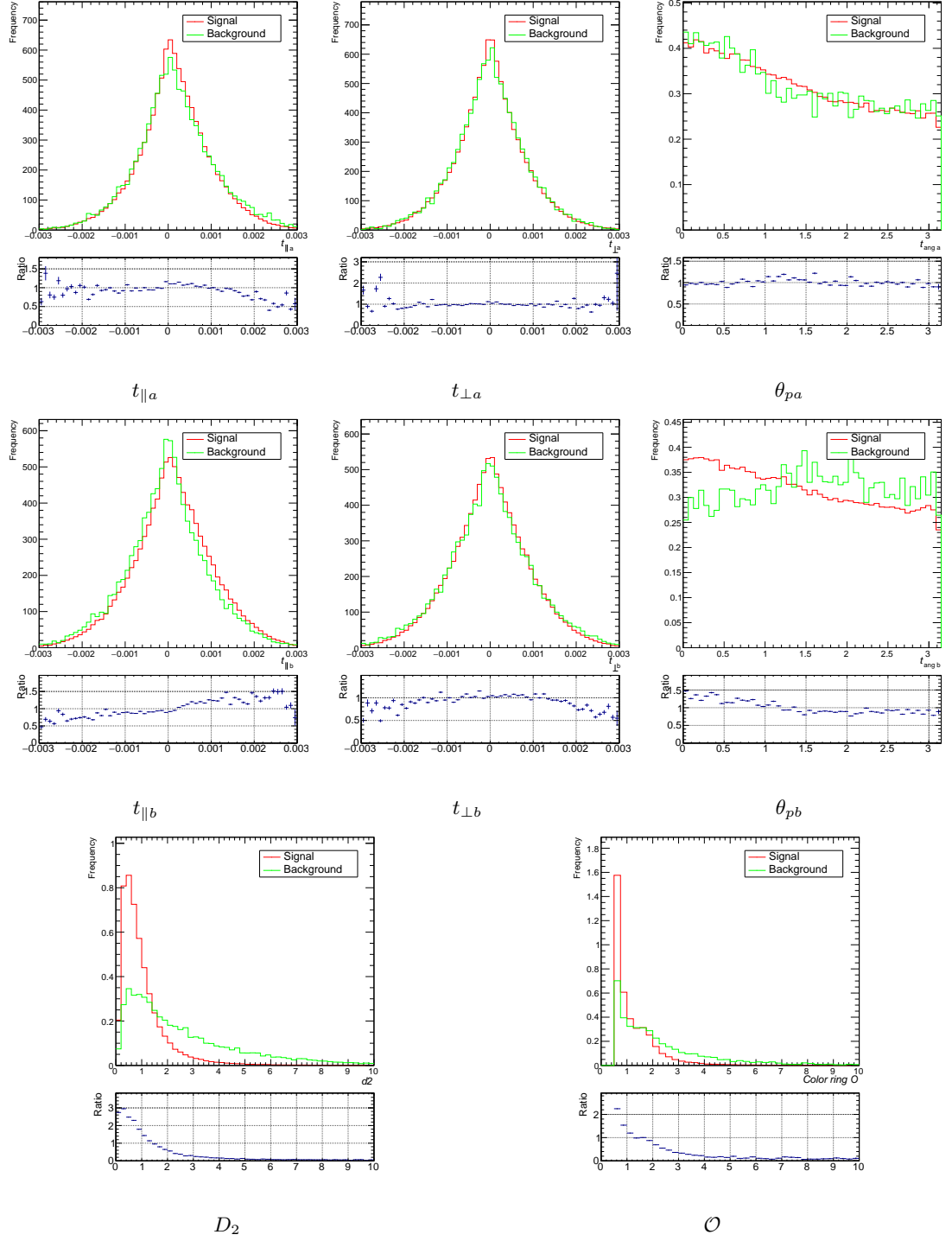
Figure 4.7: The distributions of the eight variables after all selection cuts in the reconstructed case together with the ratio between signal and background events.

Truth	
	Events Passed
Signal	20%
Background	1.6%

Reco	
	Events Passed
Signal	17%
Background	1.3%

Table 4.1: The approximate efficiencies of the analysis after all cuts were applied.

4.4 Machine Learning Architecture

The output variables were used to train two different machine learning algorithms: a Multilayer Perceptron (MLP) and Boosted Decision Tree (BDT) from the `ROOT` Toolkit for Multivariate Analysis (TMVA) library [70]. In this section, we will limit ourselves to describing the architectures of these algorithms. For a more detailed discussion on how they work, see Appendix A.

The MLP is a neural network made of one hidden layer containing nine neurons. All signal and background events are passed to the network in one batch. 50% of the data is used as a training sample and the other 50% as a test sample. The number of epochs is set to 600, and a Rectified Linear Unit activation function is used. The learning rate is set to 0.02. No downsampling is present.

The BDT is made up of 50 trees, with a maximum depth of 5 and minimum node size set to 2.5% of the total number of events. The Gini index is used as the optimization criterion. AdaBoost has been chosen as the boosting model and the number of cuts is set to 80. Also in this case, we use a 50/50 train/test sample, and there is no downsampling.

4.5 Results

We are now ready to present our findings. In Figure 4.8, the ROC curves from both the MLP and BDT are shown for the truth and reconstructed cases. The areas under the ROC curves (AUCs) are reported in Table 4.2.

	AUC - Test Sample	
	Truth	Reco
BDT	0.818	0.777
MLP	0.817	0.773

Table 4.2: The areas under the ROC curves found.

The AUCs show that in the truth case there both machine learning algorithms exhibit good signal/background discrimination. The results are unsurprisingly slightly worse for the reconstructed case due to the introduction of detector effects, but still quite good. This method is thus promising for application to real data.

We can also rank the variables based on their importance. The ranking for the BDT is reported in Table 4.3. D_2 is the most discriminant variable, and the others are all of similar importance. For the reconstructed case, $\mathcal{O}$ gains additional importance with respect to the pull variables.

Figure 4.9 shows the correlation between the variables for the signal and for the background in the reconstructed case. Unsurprisingly, $t_{\|a}$ is correlated with θ_{pa} and $t_{\|b}$ is correlated with θ_{pb},

Variable Ranking - Truth		
Rank	Var.	Importance
1	D_2	$3.743{\times}10^{-1}$
2	θ_{pb}	$1.083{\times}10^{-1}$
3	$t_{\parallel a}$	$9.508{\times}10^{-2}$
4	θ_{pa}	$9.506{\times}10^{-2}$
5	$t_{\parallel b}$	$9.729{\times}10^{-2}$
6	$\mathcal{O}$	$8.637{\times}10^{-2}$
7	$t_{\perp b}$	$8.151{\times}10^{-2}$
8	$t_{\perp a}$	$7.206{\times}10^{-2}$

Variable Ranking - Reco		
Rank	Var.	Importance
1	D_2	$3.536{\times}10^{-1}$
2	$\mathcal{O}$	$1.116{\times}10^{-1}$
3	θ_{pa}	$1.020{\times}10^{-1}$
4	$t_{\perp b}$	$9.347{\times}10^{-2}$
5	$t_{\perp a}$	$9.137{\times}10^{-2}$
6	$t_{\parallel a}$	$8.795{\times}10^{-2}$
7	θ_{pb}	$8.136{\times}10^{-2}$
8	$t_{\parallel b}$	$7.862{\times}10^{-2}$

Table 4.3: The importance of the variables found by the BDT using the data from the truth case and the reconstructed case.

as they are related to each other by definition. There is also some correlation present between D_2 and $\mathcal{O}$ in the background.

Finally, Figures 4.10 and 4.11 show the result of a Kolmogorov-Smirnov test [71] performed on the MLP and the BDT to check for overfitting. Since the test probability $\rho_{KS} > 0.01$ in all cases, no overtraining has occurred.

4.6   Mass Bias

Since our goal is to develop a tagger purely sensitive to the color configuration of the decaying particle, ideally our machine learning algorithms would be insensitive to the invariant mass of system. However, this is not the case.

In Figure 4.12a we show how different cuts on the output of the BDT depend on the invariant mass of the system in the truth case. We found that the sensitivity to the invariant mass was introduced mainly through the D_2 observable. Figure 4.12b shows how the dependence of the output of the invariant mass is eliminated when the BDT is trained without this observable. The same effect was found in the reconstructed case.

Given the fact that the D_2 observable is ranked as most important, the removal of this variable comes at the price of the efficiency of the algorithm. Focusing again on the truth case, Table 4.4 shows how AUCs of the MLP and BDT change when the algorithms are trained without the D_2 observable.

	AUC w/o D_2
BDT	0.710
MLP	0.709

Table 4.4: The values of the AUC when D_2 is not included.

In an attempt to reduce the mass bias while still training the algorithms with D_2, we generated 100k new events for the signal and artificially varied the mass of the Higgs boson. Specifically, we generated 20k events with mass $m_H = 25$ GeV, $m_H = 50$ GeV, $m_H = 75$, $m_H = 100$ GeV and $m_H = 125$ GeV and combined them into one file. After showering these events in PYTHIA8 and running the analysis, we used the output to train the machine learning algorithms. For the training, 100k background events were also generated in the same manner as described in 4.2.

In Figure 4.13 we again show how the output of the BDT depends on the invariant mass of the system. The dependence on the mass is noticeably reduced. However, with the reduction of the mass bias came a reduction of efficiency. The AUC of the BDT reduced to the values shown in Table 4.5, despite the fact that D_2 was still included.

	AUC
BDT	0.651
MLP	0.663

Table 4.5: The values of the AUC when the invariant mass of the Higgs boson is varied.

We have been able to find the reason for the reduction of the AUC. Figure 4.14 shows the distribution of the D_2 variable in the truth case when generating events with and without varying the mass of the Higgs boson. When the mass of the Higgs boson is varied, the D_2 distribution of the signal noticeably shifts towards that of the background. Effectively, the discrimination potential of the D_2 observable is reduced, explaining the reduction of the AUC.

4.7 Conclusions

Overall, we can conclude that this method for distinguishing the different color configurations of the Higgs boson and the gluon is effective both before and after the inclusion of detector effects. Unfortunately, the machine learning algorithms trained for the purpose of this classification are in part sensitive to the mass of the decaying boson. We found that the elimination of this bias comes at the cost of classification efficiency. In the future, we hope to find other ways to eliminate the bias, either by another machine learning approach or by debiasing a posteriori, in order to obtain a pure color tagger [72, 73].

A paper summarizing the results obtained is in preparation. We are also presenting the method described above to the ATLAS Hbb Working Group, as the experimental community could largely profit from it.
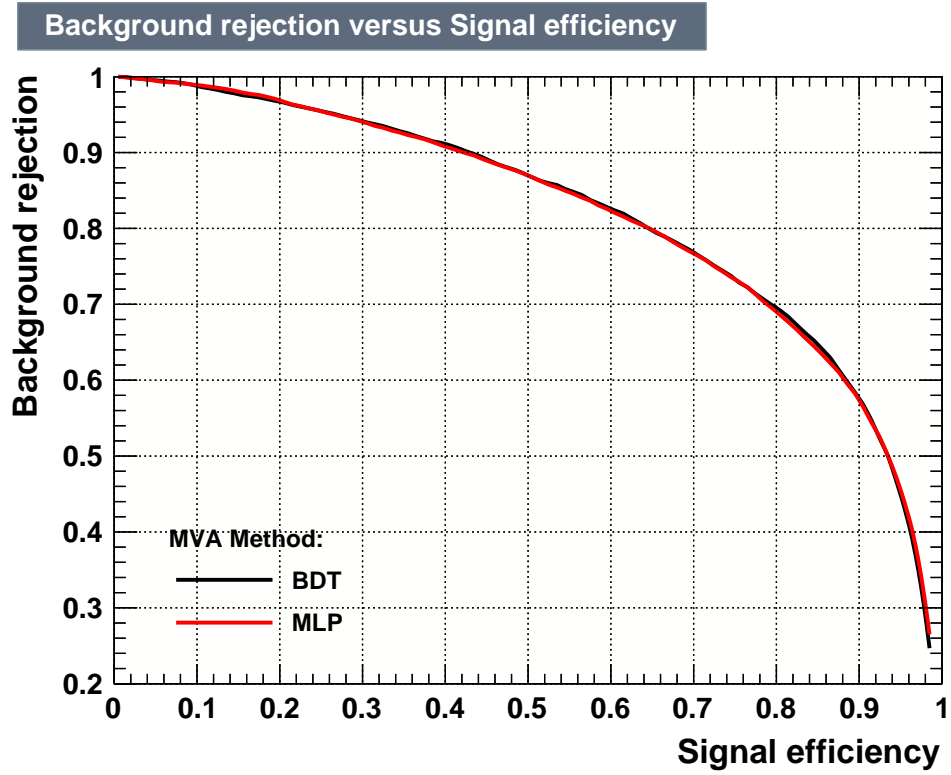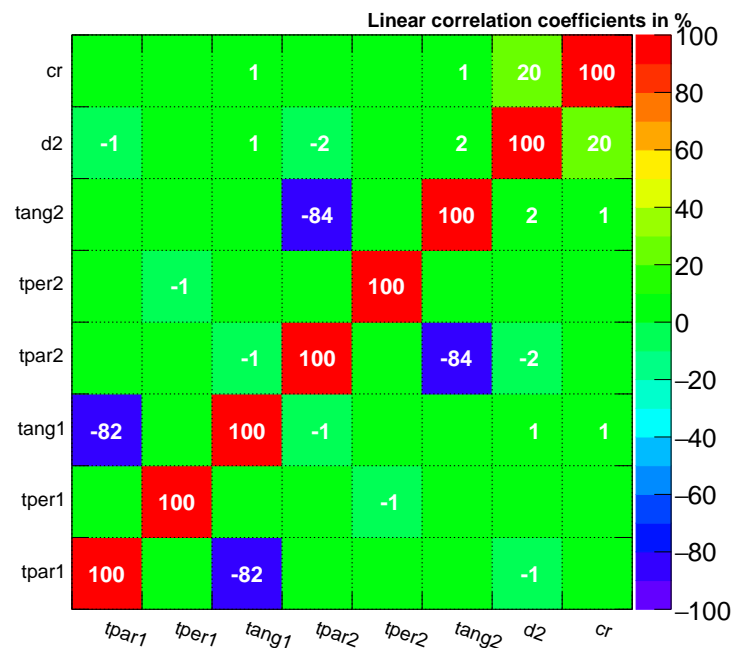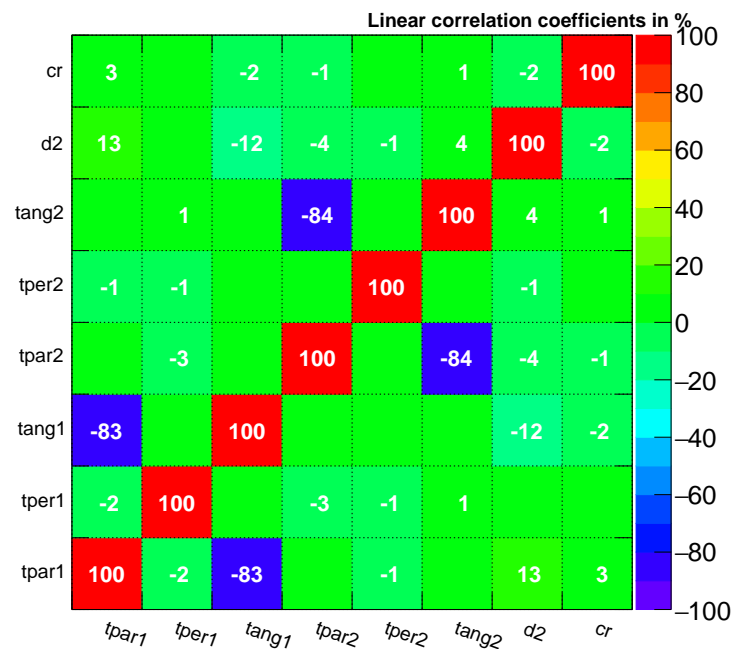
(a)
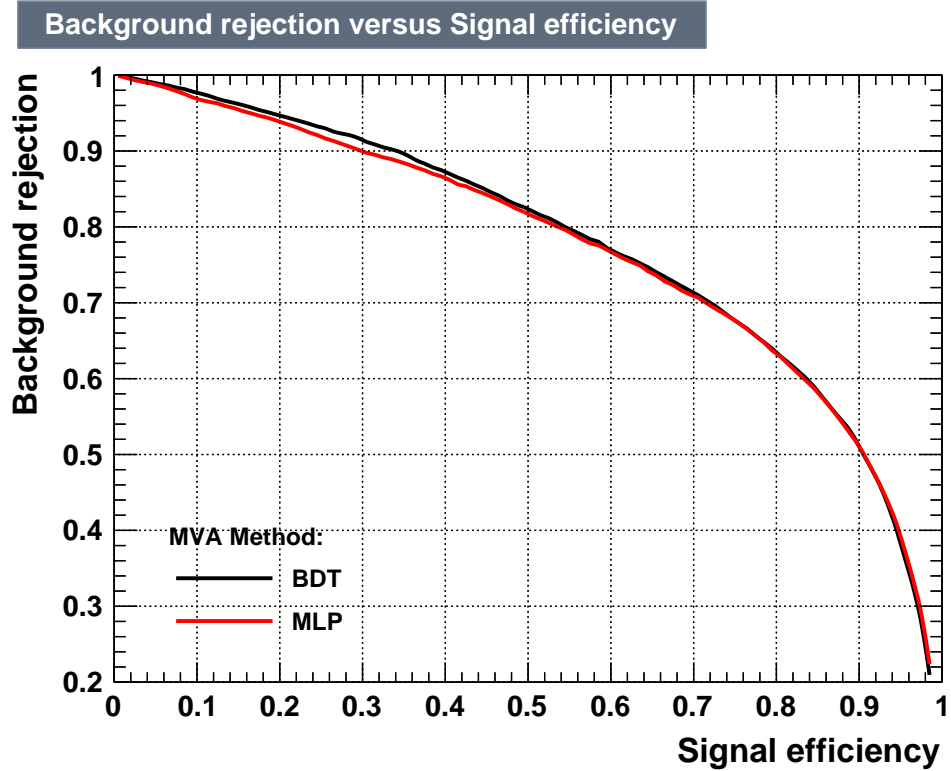
(b)

Figure 4.8: The ROC curves showing background rejection as a function of signal efficiency for the truth case (a) and for the reconstructed case (b) for both the BDT and the MLP.

Correlation Matrix (signal)

Linear correlation coefficients in %

	tpar1	tper1	tang1	tpar2	tper2	tang2	d2	cr
cr			1			1	20	100
d2	-1		1	-2		2	100	20
tang2				-84		100	2	1
tper2		-1			100			
tpar2			-1	100		-84	-2	
tang1	-82		100	-1			1	1
tper1		100			-1			
tpar1	100		-82				-1	

(a)

Correlation Matrix (background)

Linear correlation coefficients in %

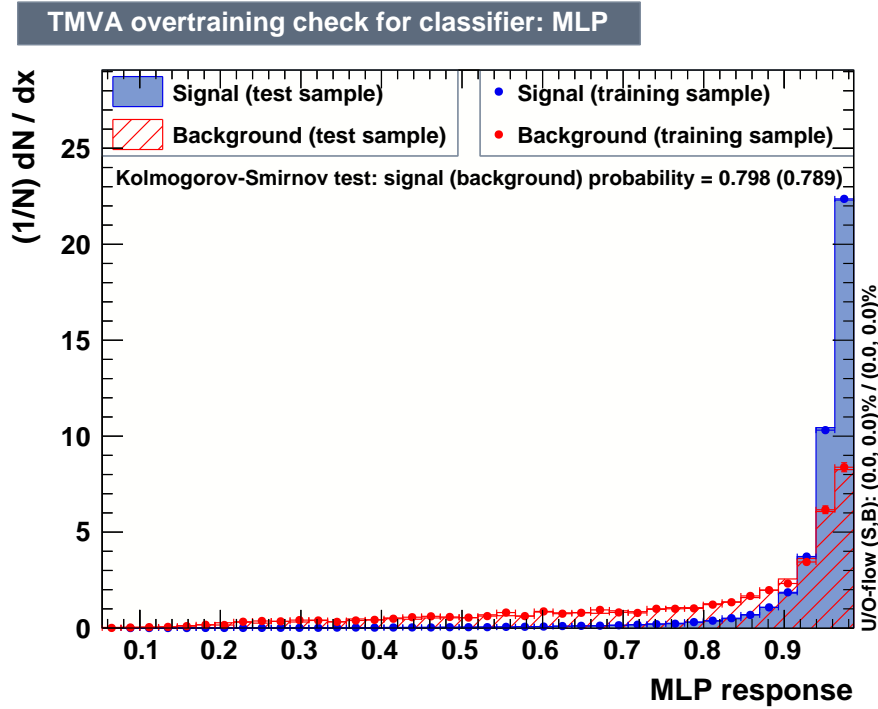|	tpar1	tper1	tang1	tpar2	tper2	tang2	d2	cr
cr	3		-2	-1		1	-2	100
d2	13		-12	-4	-1	4	100	-2
tang2		1		-84		100	4	1
tper2	-1	-1			100		-1	
tpar2		-3		100		-84	-4	-1
tang1	-83		100				-12	-2
tper1	-2	100		-3	-1	1		
tpar1	100	-2	-83		-1		13	3

(b)

Figure 4.9: The correlation matrix found by the MLP in the reconstructed case for the signal (a) and for the background (b).
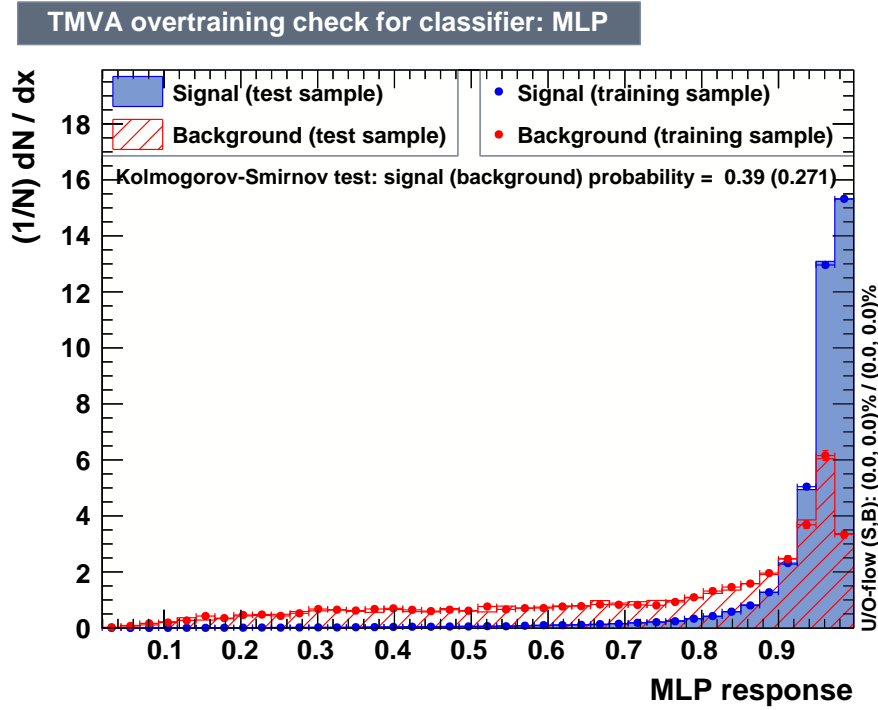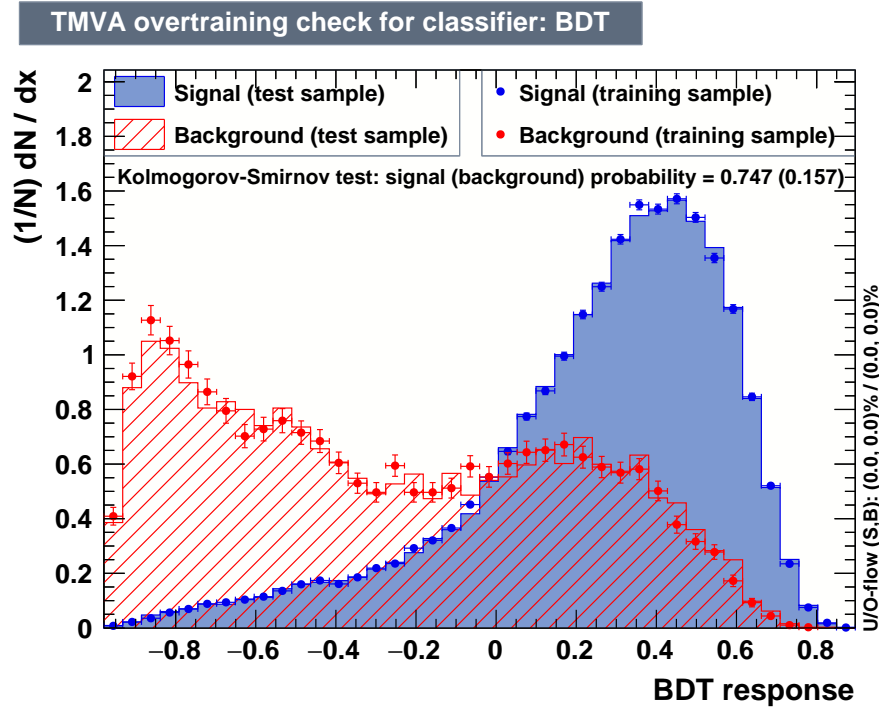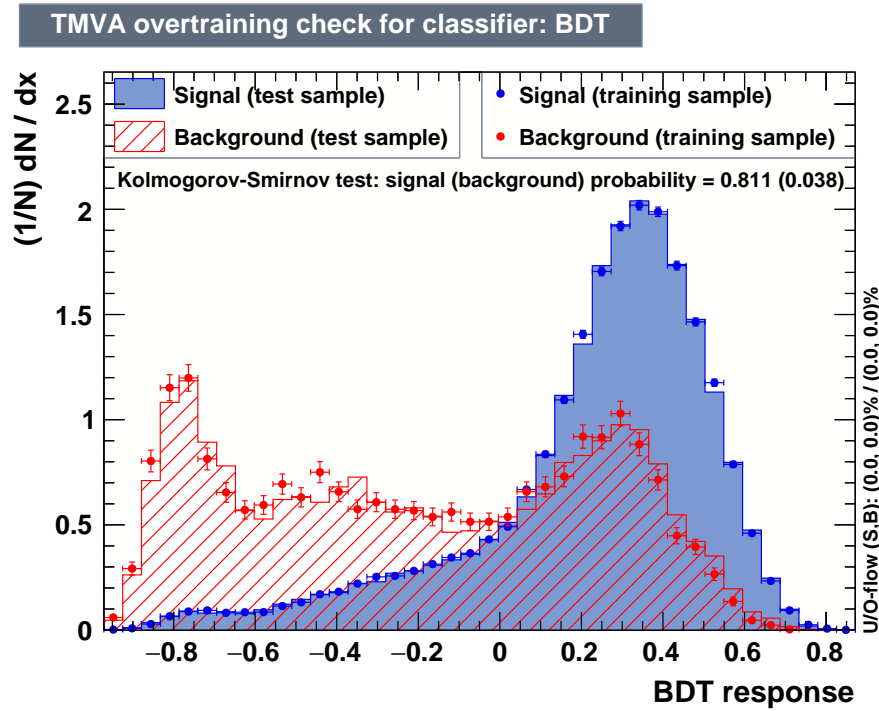
(a)

(b)

Figure 4.10: The result of a Kolmogorov-Smirnov test performed on the MLP in the truth case (a) and the reconstructed case (b).
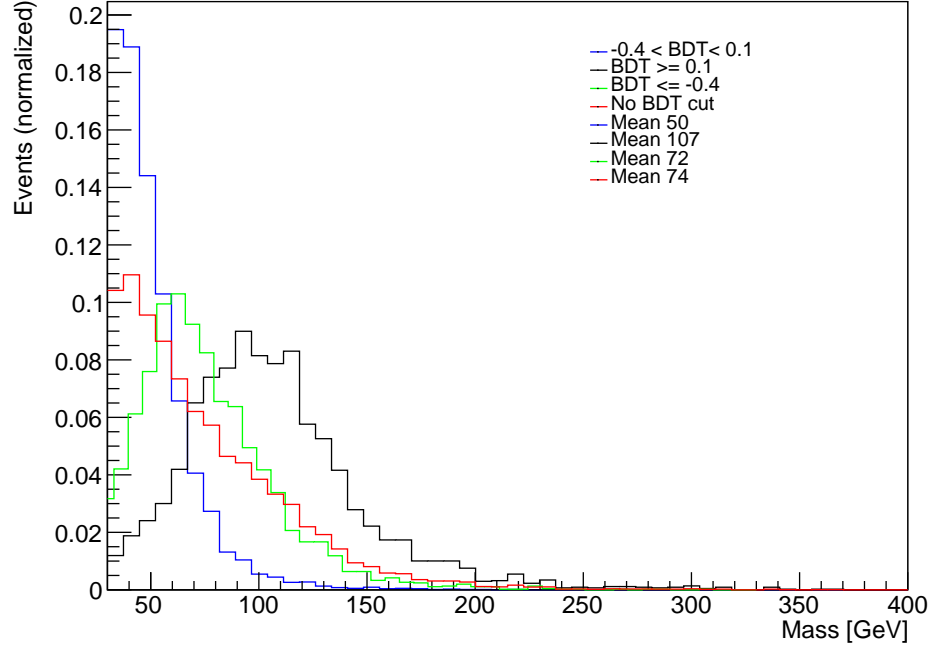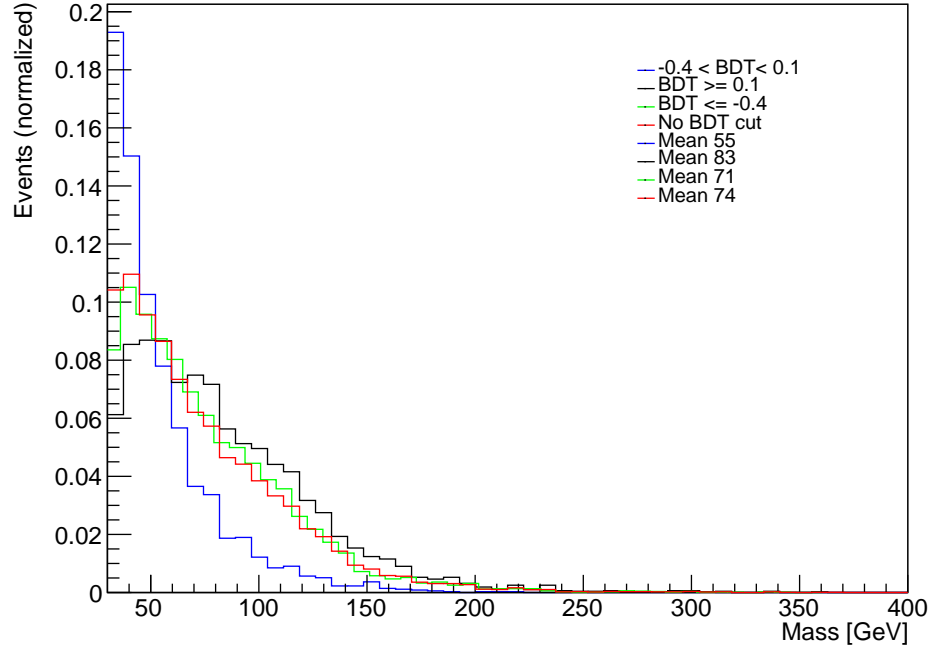
(a)

(b)

Figure 4.11: The result of a Kolmogorov-Smirnov test performed on the BDT in the truth case (a) and the reconstructed case (b).

(a)

(b)

Figure 4.12: The distribution of different cuts of the output of the BDT as a function of the invariant mass of the system with the inclusion of D_2 (a) and without the inclusion of D_2 (b)
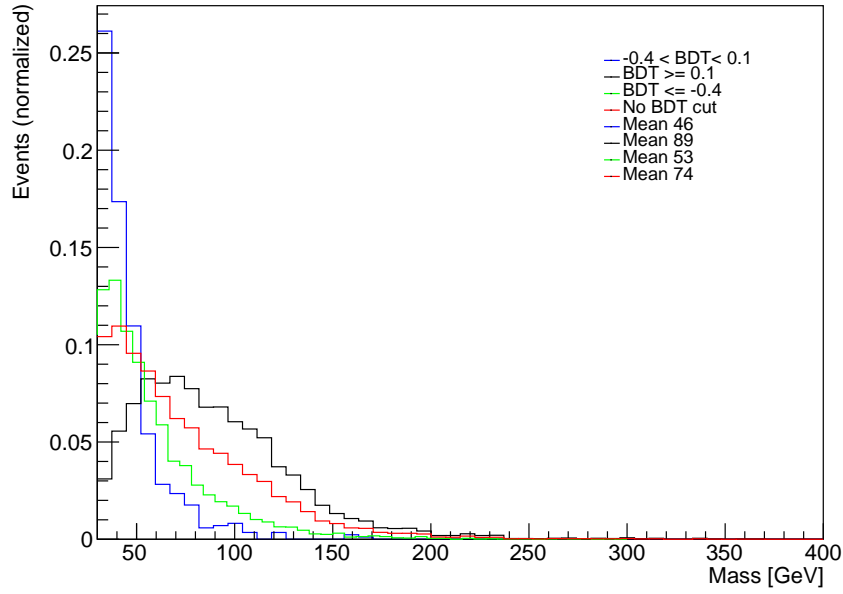
Figure 4.13: The distribution of different cuts of the output of the BDT as a function of the invariant mass of the system when this has been varied when generating hard scattering events.
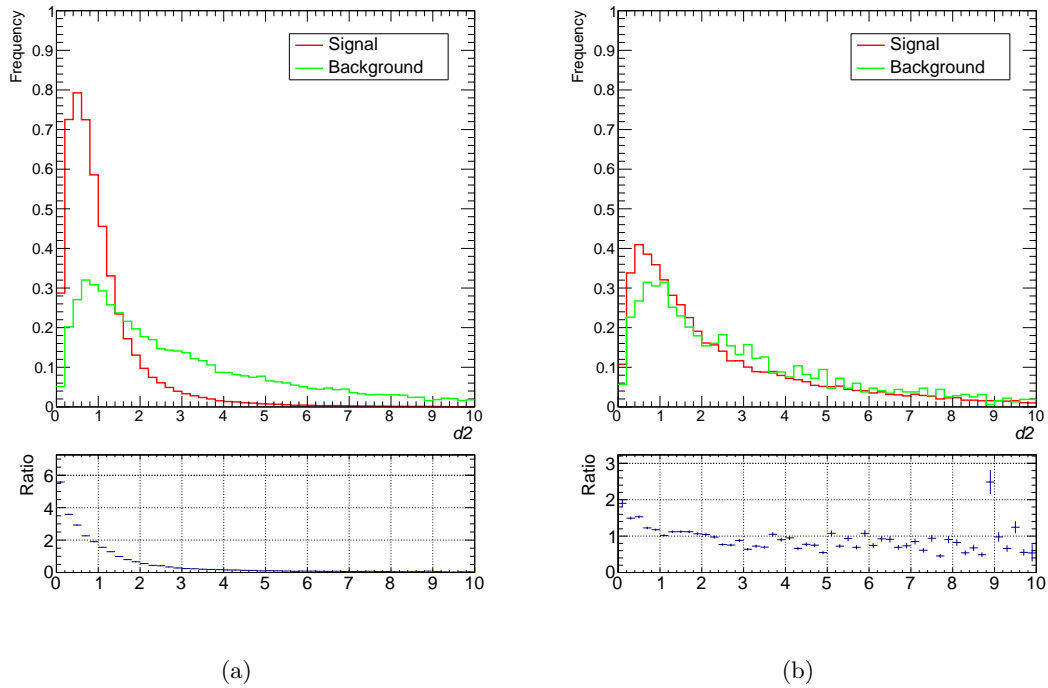

(a)   (b)

Figure 4.14: The distribution of D_2 before (a) and after (b) varying the mass of the Higgs boson.

Conclusions and Outlook

In this thesis, we have developed a machine learning based method capable of distinguishing decays originating from color singlets from those originating from color octects. Specifically, we have trained a Multilayer Perceptron and a Boosted Decision Tree on eight high-level, color sensitive variables in order to develop a color tagger.

Given the high value of the area under the ROC found when using simulated particle-level data and simulated data including detector effects (about 0.82 in the first case and 0.77 in the second), we can conclude that the method is not only effective in theory, but also shows promising prospects for application to experiment. A paper summarizing the main results is under preparation.

The ATLAS Xbb Tagger group has already expressed interest in this work. In the future, we plan to run a full simulation of the ATLAS detector, to then test the performance of the method using real data. At that point, the method can be integrated into the existing b-tagging algorithms used at ATLAS and can provide additional information regarding the partonic origin of jets.

We also hope to improve this method by introducing other variables. In particular, we would like to include the Lund Jet Plane [74], a phase-space representation of jet substructure which has already been successfully measured at ATLAS [75]. Preliminary results are extremely encouraging.

In the coming years, the increased statistics as a result of the LHC luminosity upgrade promise to provide us with more insight on the properties of the Higgs boson. With luck, precision studies led by both theoretical and experimental efforts will lead to new discoveries or to some deviation from the current understanding of particle physics, which may be the footprint of a new, previously unforeseen sector of physics.

Appendix A

Machine Learning

A.1 Machine Learning Basics

Machine learning is the generic term for a class of algorithms which are able to learn from a series of examples.

Specifically, in *supervised* machine learning, the type dealt with in this thesis, a set of examples $\{(\vec{x}_i, y_i)\}_{i=1}^{n}$ are provided. $\vec{x}_i$ represents the *feature vector* of the i-th event, and y_i is its label. The machine learning algorithm is *trained* on these examples. This training involves, in general, the minimization of some quantity. We can then test how well the algorithm performs using a subset of the provided dataset set aside before the training for this specific purpose. This subset is known as the *test set*.

The training is a delicate process. If, for example, the model is not trained on meaningful variables or when it is too simple, there is risk of underfitting. This means that, when presented with an event from the dataset on which it was trained, the model is not able to reliably assign the correct label to that event. On the other hand, if the model is too complex or there are too few training examples, we run the risk of overfitting. The difference between underfitting, a good fit, an overfitting is shown in Figure A.1 .

Once we have developed a model, we must also be able to quantify its performance. Once way of doing so is by looking at the area under the ROC curve (AUC). The ROC curve, or Receiver Operating Characteristic curve, can be used to assess the performance of classifiers, specifically
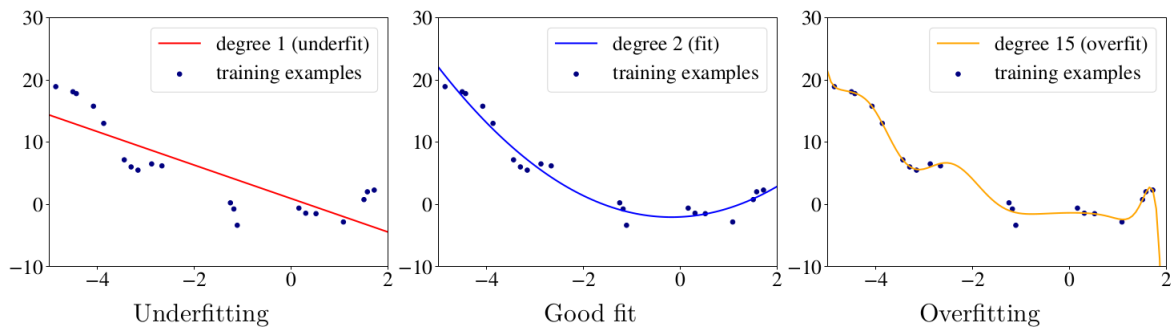

Figure A.1: An example of underfitting, a good fit, and overfitting of a dataset [76].
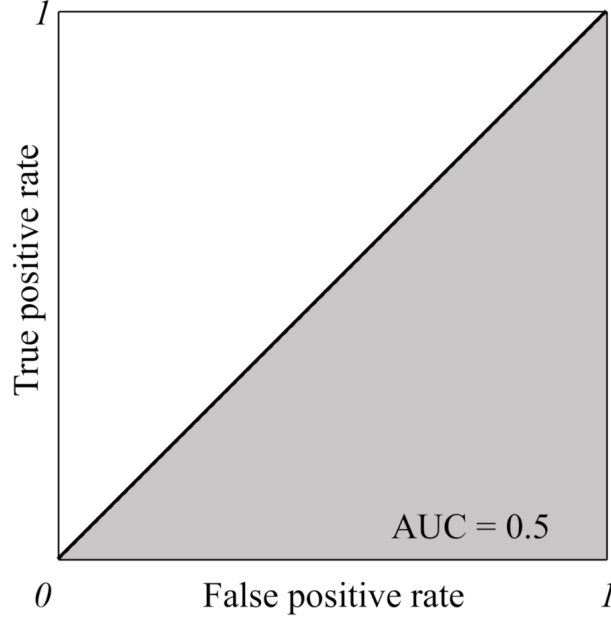
Figure A.2: An example of a ROC curve corresponding to a random guesser [76].

those classifiers which assign a label with a certain probability. These curves are built by varying the threshold above which an event is classified as signal. We then plot the True Positive Rate (TPR) against the False Positive Rate (FPR), where these are defined as

$$TPR = \frac{TP}{TP + FN} \tag{A.1}$$

$$FPR = \frac{FP}{FP + FN} \tag{A.2}$$

where TP (true positive) stands for the number of events correctly classified as signal, FP (false positive) the number of events correctly classified as background, and FN (false negative) the number of events incorrectly classified as background. The area under the curve built in this manner measures how well the model performs. An example of a ROC curve is shown in Figure A.2. A perfect model would have an $AUC = 1$ since, for all possible thresholds, $TPR = 1$ and $FPR = 0$. On the other hand, a model which guesses randomly would be placed on the diagonal, and the resulting ROC curve would have $AUC = 1/2$.

A.2   Neural Networks

A neural network is a type of machine learning algorithm inspired by biology. The network is a web in which each node, called a neuron, receives an input from the neurons in the previous layer and transmits an output to those which succeed it. In the final layer, the neuron returns a number between $[0, 1]$. In this way, a decision is made regarding the nature of the event which was given in input. A schematic representation of a neural network is shown in Figure A.3.

Formally, a neural network is a composite function. If the network contains p layers, the
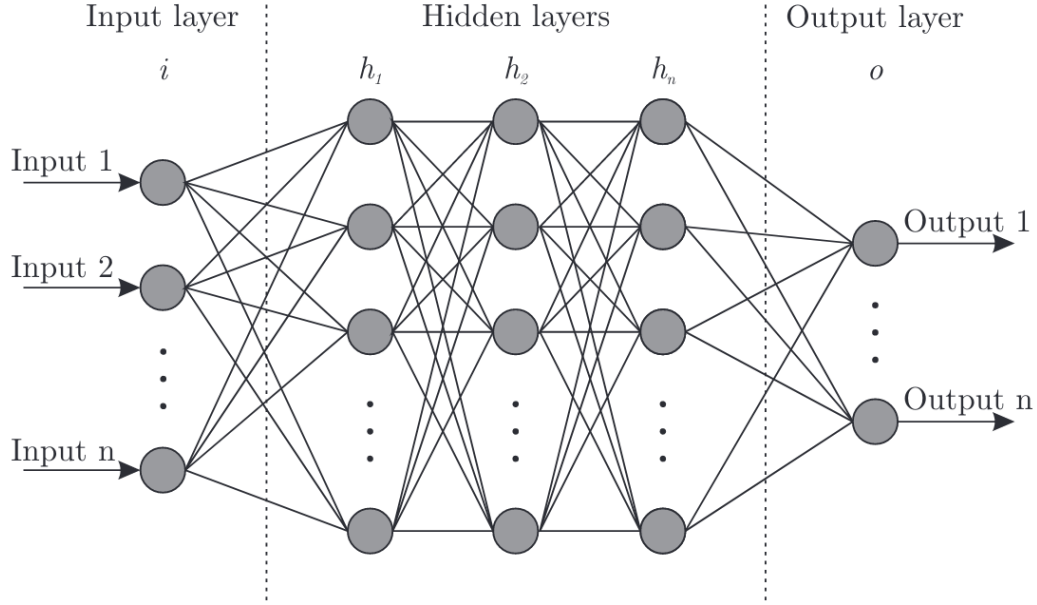
Figure A.3: The neurons composing an artificial neural network and the connections between them [77].

function is of the form

$$y = f_{NN}(\vec{x}) = (\mathbf{f}_p \circ \cdots \circ \mathbf{f}_1 \circ \mathbf{f}_0)(\vec{x}). \tag{A.3}$$

Each function $\mathbf{f}_m$ is of the form

$$\mathbf{f}_l(\vec{z}) = \mathbf{g}_l\left(\mathbf{W}_l \vec{z} + \vec{b}_l\right) \tag{A.4}$$

where $\mathbf{g}_l$ is an activation function, and $\mathbf{W}_l$ and $\vec{b}_l$ are a matrix and vector parameter, respectively, to be minimized as part of the backpropagation algorithm. We will describe both below.

A.2.1 Backpropagation

The neural network is trained to minimize the cost function. This is a function which measures how accurately the network predicts the label of an event belonging to the training set. This function can be minimized by optimizing the parameters $\mathbf{W}_l$ and $\vec{b}_l$.

We can view $\mathbf{W}_l$ as a matrix composed of column vectors $\vec{W}_{l,u}$, where u is the number of neurons in the $(l-1)$-th layer. The function (A.4) gives an output $\vec{y}$, still of dimension u, which we can use to define a cost function

$$C = \frac{1}{2N} \sum_{l,i} \|\mathbf{y}_l(\vec{x}_i) - \mathbf{g}_l(\vec{x})_i\|^2 \tag{A.5}$$

where the sum runs over all layers l and events i and $\mathbf{y}$ is a function which returns the true label of $\vec{x}$.

To avoid the complication which comes when considering multiple indices, we will consider a

one-dimensional problem. We wish to minimize (A.5), which now takes the form

$$C = \frac{1}{2N} \sum_{i=1}^{N} (y_i - (Wx_i + b))^2. \tag{A.6}$$

To do so, we must calculate the derivatives

$$\begin{cases} \frac{\partial C}{\partial W} = \frac{1}{N} \sum_{i=1}^{N} x_i(y_i - (Wx_i + b)) \\ \frac{\partial C}{\partial b} = \frac{1}{N} \sum_{i=1}^{N} (y_i - (Wx_i + b)) \end{cases} \tag{A.7}$$

These are the components of the gradient of (A.4). The gradient is a vector which points in the direction of maximum rate of change of a function. We can follow it "downhill" to find a minimum. This process is known as gradient descent.

The gradient descent algorithm proceeds in multiple iterations known as *epochs*. At the first epoch, the parameters W and b must be initialized. We then calculate the partial derivatives (A.7) to find the gradient, and update the parameters accordingly:

$$\begin{cases} W \rightarrow W - \alpha \frac{\partial C}{\partial W} \\ b \rightarrow b - \alpha \frac{\partial C}{\partial b} \end{cases} \tag{A.8}$$

where α is a parameter known as the *learning rate* which determines how far downhill we want to go during each epoch.

When dealing with gradient descent in multiple dimensions, we must simply apply the gradient descent algorithm on the singular components of the vectors. Since, in general, we are dealing with a nested function of the form (A.3), we must apply the chain rule successively, leading to a product of partial derivatives. If these are exceedingly small, their product will be even smaller, leading to a low descent rate, a problem known as the *vanishing gradient problem*. Exceedingly large gradients likewise lead to the *exploding gradient problem*.

A.2.2 Activation Functions

The activation function is a nonlinear function which defines the output of the neuron. There are several functions to choose from, but common choices are:

1. *Sigmoid*

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

2. *tanh*

$$tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

3. *Rectified linear unit (ReLU)*

$$ReLU(z) = \begin{cases} 0, \ \forall z \leq 0 \\ z, \ \forall z > 0 \end{cases}.$$

These functions each have their own advantages and disadvantages. For example, the sigmoid and tanh functions are continuously differentiable, allowing for the use of methods such as gradient descent during backpropagation. On the other hand, *ReLU* is both non-differentiable in $z = 0$ and has a step-function as a derivative, making the gradient descent unfeasible without
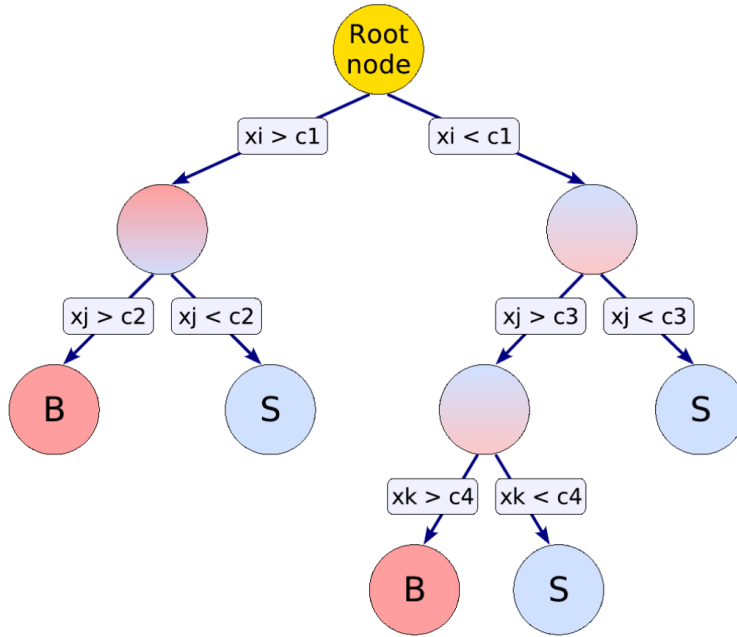
Figure A.4: The successive splitting of a decision tree based on the threshold t of a feature $x^{(j)}$ [78].

regularizing modifications. On the other hand, if the neural network is particularly deep, i.e. it contains multiple hidden layers between the input layer and the output layer, the sigmoid and tanh functions are subject to the *vanishing gradient problem*. This occurs when the product of the repeated partial derivatives taken as part of the backpropagation algorithm become too small. Since $\frac{d}{dz}ReLU(z) = 1 \; \forall z > 0$, this problem is drastically reduced.

A.3 Boosted Decision Trees

A boosted decision tree (BDT) is a type of decision tree which makes use of a boosting algorithm to create a strong classifier by combining the results of several trees (weak learners).

A.3.1 Decision Trees

A decision tree relies on successive ramifications to classify an event as signal or background, as shown in Figure A.4. At each node, the data is split based on a specific criterion. At a terminal node, known as a leaf, a decision is taken regarding the nature of the event.

Formally, consider a set $\mathcal{S} \doteq \{(\vec{x}_i, y_i)\}_{i=1}^{N}$ of labelled examples, where, as before, $\vec{x}_i \in \mathbb{R}^N$ is a feature vector and $y_i \in \{-1, 1\}$ is its label. We begin at a start node containing all examples and start with a constant model

$$f = \frac{1}{|\mathcal{S}|} \sum_{(\vec{x}, y) \in \mathcal{S}} y \tag{A.9}$$

which gives the same prediction for all inputs $\vec{x}$.

A search is run over all features $j = 1, \ldots, D$ and all thresholds t, and the dataset $\mathcal{S}$ is split into

$$\begin{cases} \mathcal{S}_- \doteq \{(\vec{x}, y) | (\vec{x}, y) \in \mathcal{S}, x^{(j)} < t\} \\ \mathcal{S}_+ \doteq \{(\vec{x}, y) | (\vec{x}, y) \in \mathcal{S}, x^{(j)} \geq t\} \end{cases}, \tag{A.10}$$

and each subset goes into a new new leaf. To find the optimal splitting, the algorithm must evaluate an *optimization criterion*. There are various possible criteria, but we will limit ourselves to illustrating two examples.

1. **Log-likelihood**

   The entropy of a set $\mathcal{S}$ is defined as

   $$H(\mathcal{S}) \doteq -f \ln f - (1 - f) \ln(1 - f). \tag{A.11}$$

   When a node is split, the entropy of the split is simply given by the sum of the entropies of the two successive leaves

   $$H(\mathcal{S}_-, S_+) \doteq \frac{|\mathcal{S}_-|}{|\mathcal{S}|} H(\mathcal{S}_-) + \frac{|\mathcal{S}_+|}{|\mathcal{S}|} H(\mathcal{S}_+) \tag{A.12}$$

   Since entropy is maximized when all variables are equiprobable and minimized when the variable assumes just one value, by minimizing the entropy of the node after the splitting, we are guaranteed to find the optimal splitting.

2. **Gini Index**

   Another possible optimization criterion is the Gini index. We start by defining the purity P of a sample in a given node,

   $$P \doteq \frac{\sum_s w_s}{\sum_s w_s + \sum_b w_b} \tag{A.13}$$

   where the sums run over the signal (s) and background (b) events and w_i indicates a weight given to the event.

   The Gini index G for a given node is then given by

   $$G = \left(\sum_{i=1}^{N} w_i \right) P(1 - P) \tag{A.14}$$

   where the sum runs over all the events which enter the node.

   Since the product $P(1 - P) = 0$ when a sample is composed entirely of signal or entirely of background events, the aim is again to minimize the entropy, this time given by the sum of the Gini index in the two leaves formed after a splitting, $G_1 + G_2$.

Once the splitting has been optimized, the algorithm iterates until one of the following conditions are met at a leaf:

- All examples have been correctly classified

- There is no attribute that can split the node

- The maximum depth of the tree has been reached

- The information gain produced by the splitting is less than some parameter ε.

The decision to split a node is taken *locally*, i.e. at the node itself. This makes a decision tree a weak classifier, since there is no guarantee to find an optimal solution. Several techniques, such as boosting, can be used to improve the model.

A.3.2 Boosting

The boosting algorithm combines several weak learners to build a stronger predictive model. The basic idea is to successively create models which reduce the errors of the previous model.

AdaBoost, short for Adaptive Boosting [79], is one of the most widely used boosting techniques. In this treatment, we will assume that the Gini index is being used as the optimization criterion.

Starting from the set of all labelled events $(\vec{x}_i, y_i)$, we assign a weight to each member of the training set. Initially, this weight is the same for all events

$$w = \frac{1}{N}. \tag{A.15}$$

For each tree m, we can calculate the error made by the classifier

$$err_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq T_m(\vec{x}_i))}{\sum_{i=1}^{N} w_i} \tag{A.16}$$

where $T(\vec{x}_i) = \pm 1$ according to whether the feature vector $\vec{x}_i$ is classified as a signal or background event and the function I depends on the accuracy of the prediction. $I(y_i \neq T_m(\vec{x}_i)) = 1$ if $y_i \neq T_m(\vec{x}_i)$ and $I(y_i \neq T_m(\vec{x}_i)) = 0$ if $y_i = T_m(\vec{x}_i)$. From here, we proceed to calculate the quantity

$$\alpha_m = \beta \ln \left(\frac{1 - err_m}{err_m} \right) \tag{A.17}$$

where β is a positive parameter usually set to unity, though $\beta = 1/2$ also appears often in literature. We use α_m to modify the weight assigned to each incorrectly classified event

$$w_i \rightarrow w_i \exp \left[\alpha_m I(y_i \neq T_m(\vec{x}_i)) \right]. \tag{A.18}$$

We normalize the weights to ensure that they still sum to unity

$$w_i \rightarrow \frac{w_i}{\sum_{i=1}^{N} w_i}. \tag{A.19}$$

In doing so, we have increased the weight of the incorrectly classified events.

We then iterate the algorithm, developing a new classifier with the updated weights, until either all events have been correctly labelled or the maximum number of iterations has been reached.

Bibliography

[1] Wikimedia Commons. File:standard model of elementary particles.svg — wikimedia commons, the free media repository, 2021. [Online; accessed 11-September-2021].

[2] Wikimedia Commons. File:running coupling constants.svg — wikimedia commons, the free media repository, 2020. [Online; accessed 19-August-2021].

[3] R. Keith Ellis, W. James Stirling, and B. R. Webber. *QCD and collider physics*, volume 8. Cambridge University Press, 2 2011.

[4] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957.

[5] M. Goldhaber, L. Grodzins, and A. W. Sunyar. Helicity of Neutrinos. *Phys. Rev.*, 109:1015–1017, Feb 1958.

[6] Eite Tiesinga, Peter J. Mohr, David B. Newell, and Barry N. Taylor. Codata recommended values of the fundamental physical constants: 2018. *Rev. Mod. Phys.*, 93:025010, Jun 2021.

[7] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[8] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[9] Wikimedia Commons. File:mexican hat potential polar with details.svg — wikimedia commons, the free media repository, 2020. [Online; accessed 13-September-2021].

[10] K. Zuber. *Neutrino Physics, Second Edition*. Series in High Energy Physics, Cosmology and Gravitation. Taylor & Francis, 2011.

[11] Wikimedia Commons. File:standard model – all feynman diagram vertices.svg — wikimedia commons, the free media repository, 2021. [Online; accessed 11-September-2021].

[12] Gerard 't Hooft. Renormalizable Lagrangians for Massive Yang-Mills Fields. *Nucl. Phys. B*, 35:167–188, 1971.

[13] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, New York, 2013.

[14] L. H. Ryder. *Quantum Field Theory*. Cambridge University Press, 6 1996.

[15] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Addison-Wesley, Reading, USA, 1995.

[16] Antonio Pich. The Standard Model of Electroweak Interactions. In *2010 European School of High Energy Physics*, 1 2012.

[17] Gerry Bunce, Naohito Saito, Jacques Soffer, and Werner Vogelsang. Prospects for spin physics at rhic. *Annual Review of Nuclear and Particle Science*, 50, 08 2000.

[18] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[19] David Griffiths. *Introduction to elementary particles*. 2008.

[20] Jessica Metcalfe, Simone Seidel, Rouzbeh Allahverdi, Alice Bean, and Igor Gorelov. Development of planar and 3d silicon sensor technologies for the atlas experiment upgrades and measurements of heavy quark production fractions with fully reconstructed d-star mesons with atlas. 08 2021.

[21] V.N. Gribov and L.N. Lipatov. Deep inelastic electron scattering in perturbation theory. *Physics Letters B*, 37(1):78–80, 1971.

[22] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.

[23] George F. Sterman and Steven Weinberg. Jets from Quantum Chromodynamics. *Phys. Rev. Lett.*, 39:1436, 1977.

[24] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[25] Claude Duhr. Qcd at nnlo and beyond. *Nuclear and Particle Physics Proceedings*, 273-275:2128–2135, 2016. 37th International Conference on High Energy Physics (ICHEP).

[26] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Franz Herzog, and Bernhard Mistlberger. Higgs Boson Gluon-Fusion Production in QCD at Three Loops. *Phys. Rev. Lett.*, 114:212001, 2015.

[27] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650–677, 1962.

[28] Gavin P. Salam. Towards Jetography. *Eur. Phys. J. C*, 67:637–686, 2010.

[29] F. J. Hasert et al. Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment. *Phys. Lett. B*, 46:138–140, 1973.

[30] F.J. Hasert et al. Search for elastic muon-neutrino electron scattering. *Physics Letters B*, 46(1):121–124, 1973.

[31] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV. *Phys. Lett. B*, 122:103–116, 1983.

[32] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett. B*, 122:476–485, 1983.

[33] M. Acciarri et al. Determination of the number of light neutrino species from single photon production at LEP. *Phys. Lett. B*, 431:199–208, 1998.

[34] Morad Aaboud et al. Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 78(2):110, 2018. [Erratum: Eur.Phys.J.C 78, 898 (2018)].

[35] Michele De Gruttola. Measurement of W, Z and Top properties with CMS. *Nuovo Cim. C*, 035N1:307–318, 2012.

[36] M. Mangano. QCD and the physics of Hadronic Collisions. *CERN Yellow Rep. School Proc.*, 4:27–62, 2018.

[37] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

[38] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[39] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2/2017, 10 2016.

[40] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 7 2013.

[41] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.

[42] Georges Aad et al. Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data. *Phys. Rev. D*, 90(5):052004, 2014.

[43] Vardan Khachatryan et al. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *Eur. Phys. J. C*, 75(5):212, 2015.

[44] Georges Aad et al. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[45] Albert M Sirunyan et al. Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state. *Phys. Rev. D*, 99(11):112003, 2019.

[46] William Murray and Vivek Sharma. Properties of the higgs boson discovered at the large hadron collider. *Annual Review of Nuclear and Particle Science*, 65(1):515–554, 2015.

[47] Georges Aad et al. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Phys. Lett. B*, 726:120–144, 2013.

[48] Georges Aad et al. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. *Eur. Phys. J. C*, 75(10):476, 2015. [Erratum: Eur.Phys.J.C 76, 152 (2016)].

[49] Vardan Khachatryan et al. Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. *Phys. Rev. D*, 92(1):012004, 2015.

[50] Xin Chen. Prospects of LHC Higgs Physics at the end of Run III. In *International Workshop on Future Linear Colliders*, 3 2017.

[51] Georges Aad et al. Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} =$13 TeV with the ATLAS detector. *Phys. Lett. B*, 800:135103, 2020.

[52] Giorgio Arcadi, Abdelhak Djouadi, and Martti Raidal. Dark Matter through the Higgs portal. *Phys. Rept.*, 842:1–180, 2020.

[53] Andrew Larkoski, Simone Marzani, and Chang Wu. Safe Use of Jet Pull. *JHEP*, 01:104, 2020.

[54] Andrew J. Larkoski, Ian Moult, and Duff Neill. Power Counting to Better Jet Observables. *JHEP*, 12:009, 2014.

[55] Andy Buckley, Giuseppe Callea, Andrew J. Larkoski, and Simone Marzani. An Optimal Observable for Color Singlet Identification. *SciPost Phys.*, 9:026, 2020.

[56] Jason Gallicchio and Matthew D. Schwartz. Seeing in Color: Jet Superstructure. *Phys. Rev. Lett.*, 105:022001, 2010.

[57] Andrew J. Larkoski, Simone Marzani, and Chang Wu. Theory Predictions for the Pull Angle. *Phys. Rev. D*, 99(9):091502, 2019.

[58] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. Energy Correlation Functions for Jet Substructure. *JHEP*, 06:108, 2013.

[59] Mozer, Matthias. Jet Substructure Tools in CMS. In *ATLAS/CMS MC workshop*, 2016.

[60] J. Neyman and E. S. Pearson. *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, pages 73–108. Springer New York, New York, NY, 1992.

[61] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[62] Johan Alwall et al. A Standard format for Les Houches event files. *Comput. Phys. Commun.*, 176:300–304, 2007.

[63] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[64] Andy Buckley, Philip Ilten, Dmitri Konstantinov, Leif Lönnblad, James Monk, Witold Pokorski, Tomasz Przedzinski, and Andrii Verbytskyi. The hepmc3 event record library for monte carlo event generators. *Computer Physics Communications*, 260:107310, 2021.

[65] Matt Dobbs and Jorgen Beck Hansen. The HepMC C++ Monte Carlo event record for High Energy Physics. *Comput. Phys. Commun.*, 134:41–46, 2001.

[66] S. Ovyn, X. Rouby, and V. Lemaitre. DELPHES, a framework for fast simulation of a generic collider experiment. 3 2009.

[67] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014.

[68] Fons Rademakers, Philippe Canal, Axel Naumann, Olivier Couet, Lorenzo Moneta, Vassil Vassilev, Sergey Linev, Danilo Piparo, Gerardo GANIS, Bertrand Bellenot, Enrico Guiraud, Guilherme Amadio, wverkerke, Pere Mato, TimurP, Matevž Tadel, wlav, Enric Tejedor, Jakob Blomer, Andrei Gheata, Stephan Hageboeck, Stefan Roiser, marsupial, Stefan Wunsch, Oksana Shadura, Anirudha Bose, CristinaCristescu, Xavier Valls, Raphael Isemann, and Kim Albertsson. root-project/root: v6.20/06, June 2020.

[69] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J. C*, 72:1896, 2012.

[70] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. 3 2007.

[71] Frank J. Massey. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[72] James Dolen, Philip Harris, Simone Marzani, Salvatore Rappoccio, and Nhan Tran. Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure. *JHEP*, 05:156, 2016.

[73] Ian Moult, Benjamin Nachman, and Duff Neill. Convolved Substructure: Analytically Decorrelating Jet Substructure Observables. *JHEP*, 05:002, 2018.

[74] Frédéric A. Dreyer, Gavin P. Salam, and Grégory Soyez. The Lund Jet Plane. *JHEP*, 12:064, 2018.

[75] Georges Aad et al. Measurement of the Lund Jet Plane Using Charged Particles in 13 TeV Proton-Proton Collisions with the ATLAS Detector. *Phys. Rev. Lett.*, 124(22):222002, 2020.

[76] A. Burkov. *The Hundred-Page Machine Learning Book*. Andriy Burkov, 2019.

[77] Facundo Bre, Juan Gimenez, and Víctor Fachinotti. Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158, 11 2017.

[78] Colomobo, F., Friese, R., Kambeitz, M. Classification using Boosted Decision Trees. In *KSETA Tutorial*, 2013.

[79] Y. Freund and R. Schapire. Experiments with a new boosting algorithm. In *ICML*, 1996.

[80] Hai-Jun Yang, B. Roe, and Ji Zhu. Studies of boosted decision trees for miniboone particle identification. *Nuclear Instruments & Methods in Physics Research Section A-accelerators Spectrometers Detectors and Associated Equipment*, 555:370–385, 2005.