STUDY OF JET-SUBSTRUCTURE TO IDENTIFY TWO-BODY DECAYS OF BOOSTED PARTICLES IN ATLAS

Relatore Interno: Prof. Attilio Andreazza
Relatore Esterno: Prof. Mario Campanelli
Correlatrice: Dr. Pauline Bernat

Tesi di laurea di:
LAURA MANENTI
Matr. 773314
Codice P.A.C.S.: 13.38.-b

Anno Accademico 2010-2011
Ad Alberto, che ha tradito la fisica, ma ne sposerà una.
Abstract

The activity hereby described has been performed within the ATLAS experiment at the Large Hadron Collider at CERN.

The work has been conducted within the ATLAS group at University College London, in the areas of jet and vector bosons physics.

The structure of jets produced by hadronically-decaying weak vector bosons with high transverse momentum was investigated. The idea being to eventually apply the same reconstruction techniques to the search of the Higgs boson.

The analysis has been carried out on both simulated events using the Monte Carlo (MC) method and data from the ATLAS detector in 2011.

Up to now, Standard Model precision measurements favor a low-mass Higgs boson. Previous experiments at LHC, combined with the recent results from ATLAS and CMS, have not excluded the existence of a low-mass Higgs between 115-140 GeV. The Higgs boson would have similar kinematics to W and Z bosons, if its mass would be within this range.

Furthermore, Z and W bosons have not been observed in the hadronic channel yet, because of the high background due to QCD events with jet production in the final state.

In this work, the possibility to observe the $W, Z \rightarrow qq$ decay channel has been investigated using jet reconstruction techniques and multi-variate analysis.

A success in employing these techniques on hadronic jet $W, Z$ decays would open the prospect to implement them in the Higgs boson search in the $b\bar{b}$ decay channel.

Since the transverse momentum distribution ($p_T$) for $W, Z$ jets decreases less rapidly than the one for QCD jets, in order to increase the purity of the sample, only vector bosons with high transverse momentum have been selected for this study.

The vector boson will then decay into two neighboring jets, which would be reconstructed as a single fat jet by standard jet-reconstruction algorithms.

The algorithm used in this analysis was Cambridge/Aachen with a 1.2 jet radius. A comparison on the typical kinematic variables for MC and real data for these reconstructed jets has been carried out.

The jet substructure has been revealed using a technique called filtering, recently developed at UCL and LPTHE.
MC level studies verify that the filtering method improves the signal resolution in the mass spectrum, increases the signal over background ratio and flattens the background shape around the peak region, facilitating the observation of the signal in the mass spectrum.

Signal-background discrimination was further improved using multivariate methods. Eight variables have been used. Four of these have been defined in the center of mass of the vector boson, where the event topology allows a better signal-background discrimination.

Three variables have been excluded, because they were weakly discriminating.

The remaining variables have been evaluated on the filtered jets and combined with different classifiers. Different cuts have then been applied to the classifiers, depending on the signal efficiency required.

The Fisher method gives the best results, with a background rejection of 27.8%, for a signal efficiency of 90%.

The signal over background ratio (S/B) increases from 0.3% to 2.6%, after combining the filtering method with MVA, in the W,Z mass region (70-100 GeV).

For a luminosity of 1.56 $fb^{-1}$ - which corresponds to the one of the data we have used - the $S/\sqrt{B}$ ratio is increased from 2.3 to 6.4.

The non-optimal agreement between MC and real data led us to apply the filtering and the MVA only at MC level.

This analysis gave the first indications on how to observe Z and W bosons in the hadronic decay channel, employing the techniques described above.

Data and MC disagreement will have to be further investigated, in order to quantitatively employ the techniques developed on data.
Riassunto

L'attività svolta in questa tesi magistrale si inserisce nell’esperimento ATLAS presso il Large Hadron Collider al CERN. Il lavoro, condotto presso il gruppo ATLAS della University College of London, rientra nell’ambito della fisica dei jet e dei bosoni vettori. Oggetto dell’analisi è stato lo studio della struttura di jet prodotti in decadimenti adronici di bosoni vettori W e Z di alto impulso, con la finalità di applicare le tecniche di ricostruzione utilizzate alla ricerca del bosone di Higgs. L’analisi è stata effettuata sia su eventi simulati col metodo Monte Carlo (MC), sia su dati raccolti da ATLAS nel 2011.

Allo stato attuale, le misure di precisione del Modello Standard della fisica delle particelle favoriscono un bosone di Higgs di bassa massa. Esperimenti precedenti a LHC, combinati con i recenti risultati di ATLAS e CMS, non hanno ancora escluso l’Higgs a bassa massa fra 115-140 GeV. Se la massa del bosone di Higgs fosse su questa scala, allora il bosone di Higgs avrebbe una cinematica simile a quella di W e Z. Ad ora, i bosoni Z e W non sono ancora stati osservati nel canale adronico per via dell’elevato background dovuto ad eventi QCD con produzione di jet nello stato finale.

Il presente lavoro ha indagato la possibilità di osservare il decadimento $W, Z \rightarrow qq$ utilizzando tecniche di ricostruzione della struttura del jet e tecniche di analisi multivariata. Appurato il buon funzionamento delle tecniche sopra citate per W,Z in canale adronico, le stesse potrebbero venire applicate alla ricerca del bosone di Higgs che decade in una coppia di quark b.

Poiché la distribuzione in momento trasverso (pt) di jet provenienti da W e Z decreased meno rapidamente di quella di jet prodotti da interazioni di QCD, si è scelto di studiare bosoni vettori di alto impulso trasverso, così da aumentare la purezza del campione. Il bosone vettore, decadendo, produrrà due jet molto vicini fra loro, che, con procedure standard di ricostruzione di jet, vengono ricostruiti come un unico jet di grande dimensione. L'algoritmo usato per la ricostruzione del jet è stato Cambridge/Aachen con raggio 1.2. Su tali jet è stato effettuato un confronto fra MC e dati reali per tipiche variabili cinematiche.

Per poter risalire alla sottostruttura del jet è stata applicata una tecnica chiamata “filtering”, sviluppata di recente a UCL e LPTHE. Lo studio a livello MC
rivela che il filtering migliora la risoluzione del segnale nello spettro di massa, aumenta il rapporto segnale-fondo e “appiattisce” il background di modo da facilitare l’osservazione di un picco nella massa.

Si è poi proceduto a studiare tecniche di separazione del segnale dal fondo utilizzando metodi multivariati. In totale si è fatto uso di otto variabili. Quattro di queste sono state definite nel centro di massa del bosone, dove la topologia dell’evento permette, tramite l’uso delle stesse, una maggiore di discriminazione segnale-fondo. Tre variabili sono infine state scartate perché poco discriminanti.

Le variabili rimaste sono state valutate sui jet ottenuti dopo la procedura di filtering e combinate in vari tipi di classificatori. Diversi tagli sono poi stati applicati ai classificatori a seconda dell’efficienza richiesta sul segnale. Il metodo che dà i risultati migliori è il discriminante di Fisher che, per un’efficienza del segnale del 90%, dà una riduzione del fondo di 27,8%. Il rapporto segnale-fondo (S/B) combinando filtering e MVA passa dallo 0,3% al 2,6% nella regione di massa che interessa Z e W (70-100 GeV). Per una luminosità di 1.56 fb$^{-1}$ (pari a quella dei dati a nostra diposizione) il rapporto passa da 2,3 a 6,4. Poiché l’accordo fra MC e dati non è ottimale per i jet cui è stato applicato il filtering, l’analisi multivariata è stata svolta solo a livello MC.

Questo lavoro di tesi ha fornito le prime indicazioni per una possibile osservazione di Z e W in canale adronico tramite l’uso delle tecniche sopra esposte. Il disaccordo tra dati e MC dovrà essere ulteriormente investigato, allo scopo di poter raggiungere una migliore descrizione dei dati, e poter applicare in maniera quantitativa le tecniche descritte.
Introduction

This study is part of a much larger ongoing effort to study hadronic decays of heavy particles: top quarks, W and Z bosons, and the Higgs boson. For all these decays the QCD background is much larger than the signal. One way to partly overcome this difficulty is to go at large transverse momenta ($p_T$), where the signal over background ratio increases.

High $p_T$ particles will then decay into two neighboring jets, which are reconstructed as a single fat jets by standard algorithms. Several techniques are currently being investigated to study these massive, boosted, hadronically-decaying particles. The most common being the use of discriminating jet shape variables evaluated on jets reconstructed through the well-understood and -calibrated anti-$k_t$ algorithm.

In the present work, the possibility to improve signal-background discrimination has been explored using a recently developed jet sub-structure reconstruction procedure, called filtering, on Cambridge-Aachen jets. This technique has been combined with a multivariate analysis (MVA). The analysis was performed on $W, Z \rightarrow qq$. This channel is one of the most difficult, since no $b$-tagging can be used to considerably reduce the background.

If, on one side, the filtering procedure is expected to reduce the uncertainties due to standard jet-reclustering techniques, on the other, it suffers of the fact that such reconstructed jets are less well understood, and in fact exhibit a worse agreement between data and simulation.

The use of the multivariate techniques here explored can and will be used in future measurements, especially the search for inclusive production of Higgs bosons decaying in the dominant $bb$ mode, where, even after the $b$-tagging requirements, the signal over background ratio is similar to that of $W, Z \rightarrow qq$.

For any observation of new particles (Higgs, or unknown heavy states) to be credible, it must first be demonstrated that the combination of the filtering combined and the MVA is well understood on a well-known process. This is why testing the techniques described above on “standard candles” weak vector bosons is so crucial.
Chapter 1

The Electroweak Theory

"Behind the complicated details of the world stand the simplicities."
— Graham Greene

The electromagnetic interaction and the weak nuclear force are unified in the context of the electro-weak theory, which constitutes a fundamental pillar of the so-called Standard Model of particle physics. In quantum theory, all forces are mediated by the exchange of a particle. The electroweak force is mediated by a neutral vector boson, the photon, and two massive ones, charged and neutral, the $W_{\pm}$ and the $Z$. The massive bosons have been discovered in 1983 at the CERN SPS collider. This Chapter gives an overview of the theoretical underpinnings of the electroweak bosons within the Standard Model. It shows how the vector bosons arise from imposing local gauge invariance on the Lagrangian; how the electroweak symmetry is broken; and finally how the $W$ and $Z$ bosons become massive.

1.1 The Standard Model

The Standard Model (SM) of particle physics is a quantum field theory (QFT) that describes three, out of the four fundamental forces, which make the subatomic particles interact: electromagnetic, weak and strong force (the gravitational force not being included). The fundamental particles of the SM are summarised in Tables 1.1 and 1.2.

In QFT the Lagrangian formalism is used and particles are described in terms of quantized fields.

In the SM, invariance under local gauge transformations on the fields is a guiding principle in writing the proper Lagrangian for the specific interaction. Demanding gauge symmetries leads directly to the inclusion of gauge bosons that mediate the interactions within the theory.

The SM theory is a combination of local gauge symmetry groups:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y.$$
14 Chapter 1. The Electroweak Theory

The bosons – integer spin particles obeying Bose-Einstein statistics. The bosons with spin-1 are the gauge bosons, which mediate the interactions: photon, $W$ and $Z$ are the electroweak gauge bosons; gluons are the QCD gauge bosons. The Higgs boson, responsible for generating the fundamental particle masses, has not been observed but searches have excluded the existence of the Higgs with $m_H < 114.4$ GeV [1] and $160$ GeV $< m_H < 170$ GeV [12].

\begin{table}
\centering
\begin{tabular}{lll}
\hline
Spin & Boson & Mass (GeV) \\
\hline
0 & Higgs (H) & > 114.4 (95\% C.L.) \\
\multicolumn{3}{l}{\text{photon} (\gamma) \hspace{1cm} 0} \\
\multicolumn{3}{l}{\text{W} (W) \hspace{1cm} 80.403 \pm 0.029} \\
\multicolumn{3}{l}{\text{Z} (Z) \hspace{1cm} 91.1876 \pm 0.0021} \\
\multicolumn{3}{l}{\text{gluon} (g) \hspace{1cm} 0} \\
\hline
\end{tabular}
\caption{The bosons – integer spin particles obeying Bose-Einstein statistics. The bosons with spin-1 are the gauge bosons, which mediate the interactions: photon, W and Z are the electroweak gauge bosons; gluons are the QCD gauge bosons. The Higgs boson, responsible for generating the fundamental particle masses, has not been observed but searches have excluded the existence of the Higgs with $m_H < 114.4$ GeV [1] and $160$ GeV $< m_H < 170$ GeV [12].}
\end{table}
1.2. U(1) Invariance, QED Lagrangian

The associated gauge bosons are the spin-1 bosons listed in Table 1.2.

The $SU(3)_C$ symmetry term results in quantum chromodynamics (QCD) – a theory which describes the strong interaction of particles which carry color charge, i.e. as quarks and gluons. As the present analysis has been done on hadronically decaying vector bosons, a meaningful discussion of QCD is not within the scope of this thesis. The electroweak sector of the Standard Model derived from the $SU(2)_L \otimes U(1)_Y$ symmetries will only be described. As it will be shown, these symmetries result in the electroweak gauge bosons: the massive W and Z; and the massless photon.

In more concrete terms, the electroweak Lagrangian is invariant under transformations belonging to both the SU(2) and U(1) groups. These transformations are unitary (hence U) and in the case of SU(2) have determinant +1 (special). Furthermore, these transformations are “local” – they are space-time dependent.

1.2 U(1) Invariance, QED Lagrangian

Notice that the free Dirac Lagrangian:

$$L = \bar{\psi} (i \gamma^\mu \partial_\mu + m) \psi$$

is invariant under the global gauge transformation:

$$\psi \rightarrow \psi' = e^{-iq\alpha} \psi$$

(1.2)

(where $\alpha$ is any real number and $q$ a dimensionless constant\(^1\), for then $\bar{\psi} \rightarrow e^{-iq\alpha} \bar{\psi}$, and in the combination $\bar{\psi} \psi$ the exponential factors cancel out.

The family of gauge transformations $U(\alpha) \equiv e^{i\alpha}$ forms a unitary Abelian group known as the unitary group U(1) \cite{24}. $\alpha$ is said to be the parameter of the transformation, whose generator, in the matrix representation, is simply $I$ (remember that a generic element of any group is given by exponentiating the generators together with the parameters of the transformation).

But what if the phase factor, $\alpha$, is different at different space-time points; that is, what if $\alpha$ is a function of $x^\mu$, so that $U(\alpha(x)) = e^{-iq\alpha(x)}$?

$$\psi \rightarrow \psi' = e^{-iq\alpha(x)} \psi$$

(1.3)

Is the Lagrangian invariant under such a local transformation?

The second term in (1.1) is clearly invariant:

$$m \bar{\psi} \psi = m \bar{\psi} e^{i\alpha(x)} e^{-i\alpha(x)} = m \bar{\psi} \psi$$

(1.4)

However, as $\alpha(x)$ is space-time dependent, $\partial_\mu \alpha(x) \neq 0$ and therefore:

$$\partial_\mu \psi \rightarrow \partial_\mu \psi' = e^{-iq\alpha(x)} \partial_\mu \psi - iq(\partial_\mu \alpha(x)) e^{-iq\alpha(x)} \psi$$

(1.5)

\(^1q\) will be eventually regarded as the electric charge, the coupling constant between the electromagnetic and the Dirac field.
Chapter 1. The Electroweak Theory

The second term shows that the Lagrangian of (1.1) is not locally invariant under (1.3) due to the second term in (1.5). However, it can be made invariant if a gauge field, $A_\mu = A_\mu(x)$, is included in the theory. This field must transform as:

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha(x) \quad (1.6)$$

$A_\mu$ is used to modify the derivative $\partial_\mu$ into the covariant derivative:

$$D_\mu = \partial_\mu + iqA_\mu \quad (1.7)$$

which has the desired behavior that:

$$D_\mu \psi \rightarrow D'_\mu \psi = e^{-iq\alpha(x)} D_\mu \psi \quad (1.8)$$

thus leading to the following Lagrangian, which is invariant under the local U(1) transformation:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu + m)\psi + \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \quad (1.9)$$

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the energy-momentum tensor for the field $A_\mu$ and the last term describes the motion of free vector bosons.

So far we have seen that demanding local phase invariance under U(1), we are forced to introduce a vector field $A_\mu$, which, through the covariant derivative, $D_\mu$, couples to the Dirac particle with strength $q$ [24]. The full Lagrangian must then include a “free” term for the gauge field, $\frac{1}{4} F^{\mu\nu} F_{\mu\nu}$ [23]. A mass term for this field would have the form $M^2 A_\mu A^\mu$, which is not invariant under U(1). If the symmetry wants to be preserved, it follows that mass terms are not permitted, that is the particle associated to the field $F^{\mu\nu}$ has to be massless.

The vector field $A_\mu$, so added to the initial Dirac free Lagrangian, reveals to be the electromagnetic potential, with the photon being correctly massless. Remarkably, the requirement of local gauge invariance, applied to the free Dirac Lagrangian, generates all of electrodynamics. Explicitly the interaction term between the Dirac field, $\psi$, and the photon field, $A_\mu$, the Lagrangian of Eq. (B) becomes:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu + m)\psi + e\bar{\psi}\gamma^\mu A_\mu \psi + \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \quad (1.10)$$

In the Standard Model, the $U(1)_Y$ symmetry is associated not with the electric charge $q$, but with weak hypercharge, $Y$, (hence $U(1)_Y$). However, the $SU(2)_L \otimes U(1)_Y$ is spontaneously broken, leaving a $U(1)_{EM}$ with electric charge as its generator, as observed in nature. This will be discussed more in Section (1.9).
1.3 SU(3) Invariance, QCD Lagrangian

In an analogous way, the structure of QCD can be inferred through the above idea, replacing the $U(1)$ gauge group by the non-Abelian$^2$ $SU(3)_C$ group of phase transformations on the quark color fields:

$$L_{QCD} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - gG^A_\mu(\bar{\psi}\gamma^\mu T^a_\mu \psi) - \frac{1}{4}G^A_{\mu\nu}G^A_{\mu\nu}$$  \hspace{1cm} \text{(1.11)}$$

where for a gluon field:

$$G^A_\mu = \partial_\mu G^A - \partial_\mu G^A - g_{fABC}G^B_\mu G^C_\nu$$  \hspace{1cm} \text{(1.12)}$$

the quarks carry color charge, $r, g, b$, anti-quarks carry anti-charge, $\bar{r}, \bar{g}, \bar{b}$ and the force is mediated by massless gluons. Just as for the photon, local gauge invariance requires the gluons to be massless.

The strong interaction is invariant under rotations in color space: this is an exact $SU(3)$ symmetry. In QED, photons are electrically neutral and hence do not carry the charge of the EM interaction. In contrast, gluons do carry color charge (one unit of color and one of anti-color) and so QCD, through the non-abelian term, $G^A_{\mu\nu}$, allows gluon self-interactions: both triple and quartic gluon vertices are possible, as can be seen in equation (2.3). Color charge is conserved at all vertices.

$$L_{QCD} = \bar{q}q + G^2 + g\bar{q}qG + g^2G^3 + g^2G^4$$  \hspace{1cm} \text{(1.13)}$$

1.4 SU(2) Invariance

Imposing local gauge invariance under SU(2) transformations follows a similar logic to the U(1) case. However, there are additional features which result from the richer group structure. The principal difference is that SU(2) has $2^2-1=3$ generators, $T_i$, of transformations rather than the one of U(1). $T_i$ obey the group algebra:

$$[T_i, T_j] = i\epsilon_{ijk}T_k$$  \hspace{1cm} \text{(1.14)}$$

The special unitary local transformations generated by these $T_i$ take the form:

$$U(\theta) = e^{-igW^T \cdot \theta(x)}$$  \hspace{1cm} \text{(1.15)}$$

where $g_W$ is the weak coupling strength and $\theta = (\theta_1, \theta_2, \theta_3)$ are the space-time dependent parameters of the transformation.

In an analogous manner to the U(1) example above, SU(2) local gauge invariance is maintained by use of the covariant derivative:

$$D_\mu = \partial_\mu + igT \cdot W_\mu$$  \hspace{1cm} \text{(1.16)}$$

$^2$The group is non-Abelian since not all the generator of the group commute with each other
which again has the desired transformation properties (compare to (1.8)):

$$D_\mu \chi \rightarrow D_\mu' \chi' = U(\theta) D_\mu \chi$$

(1.17)

where $\chi$ is now an SU(2) doublet (analogous to the U(1) singlet $\psi$ in the previous section).

Three gauge fields, $W_\mu^i (i \in \{1, 2, 3\})$, had to be introduced, as three is the number of generators for SU(2).

This covariant derivative gives rise to the interactions between the SU(2) gauge fields and the particles represented by $\chi$. The coupling $g_W$ is the same for all SU(2) multiplets (a consequence of the non-Abelian nature of SU(2)). This is in contrast to the U(1) case, where $q$ could take on different values, e.g. $-1$ for the electron, but $-1/3$ for the down quark.

The field strength tensor for the gauge fields is defined as follows:

$$F^{\mu\nu} = \partial^{\mu} W^{\nu} - \partial^{\nu} W^{\mu} - g W^{\mu} \times W^{\nu}$$

(1.18)

The non-Abelian nature of SU(2) reveals itself in the $g W^{\mu} \times W^{\nu}$ term. The presence of $g$, the gauge coupling constant, implies that the gauge fields carry SU(2) charge. There are interactions between the gauge fields directly.

The final form for the SU(2) invariant Dirac Lagrangian is then:

$$\mathcal{L} = \bar{\chi} (\gamma D_\mu + m) \chi + \frac{1}{4} F^{\mu\nu} F_{\mu\nu}$$

(1.19)

As in the U(1) case, inserting mass terms for the gauge fields into the Lagrangian would violate gauge invariance. While in the case of QED this is desirable as the photon is massless, the weak gauge bosons are massive (of the order of 100 GeV/c$^2$). In the Glashow-Salam-Weinberg model, this problem is tackled by the Higgs mechanism.

## 1.5 The Glashow-Salam-Weinberg Model

Glashow, Salam and Weinberg state that the symmetry group of the electroweak theory is:

$$SU(2)_L \otimes U(1)_Y$$

The subscript L is sometimes substituted by “W”, which stands for weak. The subscript L here is to remind us that the weak isospin current couples only left-handed fermions - those with negative helicity (this is observed in experiments). “Y” stands for hypercharge.

The Dirac spinors $\psi$ can be considered as superpositions of left- and right-handed fermions and can be projected onto left $\psi_L$ and right $\psi_R$ components by the projection operators:
\[ \psi_L = \hat{P}_L \psi = \frac{1 - \gamma_5}{2} \psi \quad \psi_R = \hat{P}_R \psi = \frac{1 + \gamma_5}{2} \psi \] (1.20)

The Dirac spinors representing the left-handed fermion fields are combined in SU(2) doublets and assigned the following isospin quantum numbers:

\[ (\nu_e, e^-)_L, (\nu_\mu, \mu^-)_L, (\nu_\tau, \tau^-)_L, (u, d)_L, (c, s)_L, (t, b)_L \] (1.21)

the upper component of the doublet having \( t_3 = +\frac{1}{2} \) and the bottom component \( t_3 = -\frac{1}{2} \).

Several aspects of the weak interaction are encapsulated in these assignments. The electric charge of the doublet members differs by 1, so the gauge field responsible for transitions between them must have charge \( \pm 1 \). Moreover, for leptons these transitions are strictly bound to remain within generations (as listed in Table 1). This is a statement that the lepton numbers are conserved. In the case of the quarks however, the states within these weak isospin multiplets are not the QCD eigenvectors, but rather are combinations of them. These combinations are given by the Cabibbo-Kobayashi-Maskawa matrix:

\[
\begin{pmatrix}
V_{ud} & V_{us} & V_{ub} \\
V_{cd} & V_{cs} & V_{cb} \\
V_{td} & V_{ts} & V_{tb}
\end{pmatrix}
\] (1.22)

The resulting cross-generational couplings in the quark sector permit the weak decay of hadrons such as the \( K^\pm \), which would otherwise be stable within the Standard Model. These couplings are also the source of CP (combined Charge-Parity symmetry) violation in the Standard Model.

The right handed fermion fields are SU(2) singlets \((t = 0, t_3 = 0)\):

\[ e^-_R, \mu^-_R, \tau^-_R, u_R, d'_R, c_R, s'_R, t_R, b'_R \] (1.23)

Within the GSW framework neutrinos are considered massless and so only have one definite helicity - found experimentally to be left-handed[21].

The separation of the left- and right-handed fermions into separate \( SU(2)_L \) multiplets means that the incorporation of a mass term as in (1.1) would break the symmetry, since \( m\bar{\psi}\psi \) couples the left- and right-handed components of the spinors:

\[ m\bar{\psi}\psi = m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \] (1.24)

If the \( SU(2)_L \) transformations (1.15) act only on the left-handed components of the \( \chi \) SU(2) doublet, then clearly these mass terms are not gauge invariant. The masses of the fermions will be shown to arise through couplings to the Higgs sector in Section 1.6.

The weak hypercharge, \( Y_W \), is the generator of the U(1) symmetry and is calculated as:

\(^3\)All three quark colors are included in the multiplets.
Chapter 1. The Electroweak Theory

\[ Y_W = 2(Q - t_3) \quad (1.25) \]

Requiring the Lagrangian to be invariant under the the combined \( SU(2)_L \otimes U(1)_Y \) group gives rise to four gauge bosons – the number that are observed in nature. The Lagrangian for these gauge bosons is:

\[ \mathcal{L}_G = -\frac{1}{4}F_{\mu \nu}F^{\mu \nu} - \frac{1}{4}B_{\mu \nu}B^{\mu \nu} \quad (1.26) \]

where \( F_{\mu \nu} \) is the energy-momentum tensor for the three \( SU(2) \) gauge bosons \( W_\mu = (W_\mu^+, W_\mu^-, W_\mu^3) \) and \( A_\mu \) that for the \( U(1) \) gauge boson.

These four bosons are however all massless and do not yet correspond to the physical states. In GSW theory the \( SU(2)_L \otimes U(1)_Y \) symmetry is eventually “broken”: the two neutral states, \( W_3^2 \) and \( B \) “mix” producing one massless linear combination (the photon), and an orthogonal massive combination (the \( Z_0 \)) [23].

An obvious question arises: why do the \( B \) and the \( W_3^2 \) states “mix” to form a massive particle, the \( Z_0 \), and a massless particle, the photon? Why is the underlying symmetry of the electroweak interactions “broken”? How can we account for the masses of the gauge bosons and the fermions without “ourselves” breaking the symmetry by adding the mass terms?

The Higgs mechanism has been introduced as a solution to the spontaneous breaking of the electro-weak symmetry.

1.6 Spontaneous Symmetry Breaking: “Hidden” Symmetry

Since the Feynman calculus is really a perturbation procedure, in which we start from the ground state (the “vacuum”), and treat the fields as fluctuations about that state, it is crucial for a given Lagrangian to identify its ground state. The ground state is obtained by finding the value of the field for which the potential \( V(\phi) \) in the Lagrangian has a minimum. The Feynman calculus is then formulated for that Lagrangian in terms of deviations from one or the other of these ground states.

Suppose we have the following Lagrangian for the scalar field \( \phi \):

\[ \mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) + \frac{1}{2}\mu^2 \phi^2 - \frac{1}{4}\lambda^2 \phi^4 \quad (1.27) \]

where \( \mu \) and \( \lambda \) are (real and positive) constants. The second term looks like a mass, and the third like an interaction, but we see that if the second term is truly a mass, then \( m \) is imaginary, which is nonsense. Before any speculation, we should, as said before, re-express \( \mathcal{L} \) as a function of the deviation from its ground state.

In the present case:

\[ V(\phi) = -\frac{1}{2}\mu^2 \phi^2 + \frac{1}{4}\lambda^2 \phi^4 \quad (1.28) \]
and the minimum occurs at

$$\phi = \pm \mu/\lambda$$

(1.29)

Perturbative calculations should involve expansions around the classical minimum $\phi = \mu/\lambda$ or $\phi = -\mu/\lambda$.

We therefore write

$$\phi(x) = \frac{\mu}{\lambda} + \eta(x)$$

(1.30)

where the new variable field $\eta(x)$ represents the quantum fluctuations about this minimum. We have chosen to translate the field to $\phi(x) = +\mu/\lambda$, but this does not imply any loss of generality since $\phi = -\mu/\lambda$ can always be reached by reflection symmetry. (Nature has also to make such a choice).

In terms of $\eta$ the Lagrangian reads:

$$L' = \frac{1}{2} (\partial_\mu \eta)(\partial^\mu \eta) - \mu^2 \eta^2 \pm \mu \lambda \eta^3 - \frac{1}{4} \lambda^2 \eta^4 + \text{const}$$

(1.31)

The second quantity is now a mass term with the correct sign. Comparing it to the Klein-Gordon Lagrangian for a scalar field $\phi$:

$$L = \frac{1}{2} (\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2} \left( \frac{mc}{\hbar} \right)^2 \phi^2$$

(1.32)

we discover that the mass of the particle is $m = \sqrt{2} \mu \hbar / c$.

The Lagrangians (1.27) and (1.31) represent the same physical system, in the latter the notation only has changed.

So the recipe to find the (hidden) mass term in a Lagrangian is to locate the ground state; re-express $L$ as a function of the deviation, $\eta$, from this minimum; expand in powers of $\eta$ and obtain the mass from the coefficient of the $\eta^2$ term.

In doing this we gained clarity in the identification of the mass term, but we lost the original symmetry of the Lagrangian under the exchange of $\phi$ into $-\phi$. The spontaneous symmetry-breaking phenomenon is a way to account for the mass of the particles without adding them by hand.

### 1.7 The Goldstone Boson

Having introduced the concept of spontaneous breaking symmetry, we now repeat the above procedure for a complex scalar field $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$ described by the Lagrangian:

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

(1.33)

where $\mu$ and $\lambda$ are real and positive constants. $L$ is invariant under the global transformation $\phi \to e^{i \alpha} \phi$, that is, $L$ possesses a $U(1)$ global gauge symmetry.

The “potential energy” function, written in terms of $\phi_1$ and $\phi_2$ reads:
\[ V = \frac{1}{2} \mu^2 (\phi_1^2 + \phi_2^2) + \frac{1}{4} \lambda^2 (\phi_1^2 + \phi_2^2)^2 \] (1.34)

and the minima lie on a circle of radius \( \mu/\lambda \):

\[ \phi_1^2 + \phi_2^2 = \frac{\mu^2}{\lambda^2} \] (1.35)

Again we translate the field \( \phi \) to a minimum energy position, which without loss of generality we may take as the point

\[ \phi_1 = \frac{\mu^2}{\lambda^2}, \phi_2 = 0 \] (1.36)

As before, we introduce new fields, \( \eta \) and \( \xi \), which are the fluctuations about this vacuum state:

\[ \eta \equiv \phi_1 - \mu/\lambda; \quad \xi \equiv \phi_2 \] (1.37)

so that

\[ \eta \equiv \phi_1 - \mu/\lambda; \quad \xi \equiv \phi_2 \] (1.38)

Rewriting the Lagrangian in terms of these new field variables we obtain:

\[ \mathcal{L} = \frac{1}{2} (\partial_\mu \eta)^2 - \mu^2 \eta^2 + \frac{1}{2} (\partial_\mu \xi)^2 + \text{const.} + \text{cubic and quadratic terms in } \eta, \xi \] (1.39)

The first and the second term represent respectively the kinetic energy and the mass for the \( \eta \)-field. Similarly, the third term represents the kinetic energy for the \( \xi \)-field, but this time the field has no corresponding mass term. That is, the theory also contains a massless scalar, which is known as a Goldstone boson.

It can be shown (Goldstone’s theorem) that spontaneous breaking of a continuous local symmetry is always accompanied by the appearance of one or more massless scalar particles (called “Goldstone’s bosons”).

The \( \eta \)-field corresponds to a particle that oscillates in the direction of the slope (see Fig. 1.2) with non-zero frequency; on the contrary, the \( \xi \)-field corresponds to oscillations along the valley of the minima with zero frequency (massless mode), that is for which there is no resistance.

We hoped to use the mechanism of spontaneous symmetry-breaking to account for the mass of the weak interactions gauge fields, but now we find that this introduces a massless scalar boson, which we don’t know to exist in nature.

We will see in the next paragraph that the key that leads to the final solution is to proceed from a global to a local gauge theory.
where the classical potential term $V$ is chosen to have the form:

$$V = 4 \mu^2 \phi^2.$$  

Note the negative mass term $\mu^2 \phi^2$. If $\mu > 0$ and $\mu^2 > 0$, this potential will not be the typical ‘well’ with a unique minimum at $\phi = 0$, but rather will have a unstable maximum there. The minima satisfy the condition $|\phi| = q^2 \mu^2$. These can be considered as an annulus of degenerate classical potential ground states in the $SU(2)_L \times U(1)_Y$ space as illustrated in Figure 1.3. In field theoretic terms the classical potential minimum can be interpreted as a set of degenerate vacua, $|\phi|$, such that:

$$V = \frac{1}{4} \lambda^2 (\phi_1^2 + \phi_2^2)^2 \mu^2.$$  

As already stated, (1.23) with (1.24) is invariant under $SU(2)_L \times U(1)_Y$ and thus this set of vacua is too. However, any particular vacuum in which the system will actually lie is an unstable maximum and there is a continuum of degenerate ground states which satisfy $\phi_1^2 + \phi_2^2 = \mu^2 / \lambda^2$.

Figure 1.1 – For an SU(2) doublet of scalars, $\phi$, the classical potential $V = -\frac{1}{2} \mu^2 (\phi_1^2 + \phi_2^2) + \frac{1}{2} \lambda^2 (\phi_1^2 + \phi_2^2)^2$ is represented here. $\phi = (0, 0)$ is an unstable maximum and there is a continuum of degenerate ground states which satisfy $\phi_1^2 + \phi_2^2 = \mu^2 / \lambda^2$.

Figure 1.2 – Oscillation of a particle in the direction of the slope.
1.8 Introduction of the massive Higgs boson field

If we want the Lagrangian in Eq. (1.33) to be invariant under a $U(1)$ local gauge transformation

$$\phi \to e^{i\theta(x)}\phi$$

(1.40)

$\partial_\mu$ requires to be replaced by the covariant derivative, $D_\mu = \partial_\mu - ieA_\mu$.

The gauge invariant Lagrangian written as a function of $\eta$ and $\xi$ is:

$$\mathcal{L}' = \frac{1}{2}(\partial_\mu \xi)^2 + \left[\frac{1}{2}(\partial_\mu \eta)^2 - \mu^2 \eta^2\right]$$

$$+ \left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\left(\frac{e \mu}{\lambda}\right)^2 A_\mu A^\mu\right]$$

$$+ 2i\left(\frac{e \mu}{\lambda}\right)A_\mu \partial^\mu \xi + \text{interaction terms}$$

(1.41)

The particle spectrum of $\mathcal{L}'$ appears to be a massless Goldstone boson $\xi$, a massive scalar $\eta$, and a massive vector $A_\mu$. Before asking ourselves what to do with the unphysical massless Goldstone boson, we notice that the last term couples the $A_\mu$-field with the $\xi$-field. If we read it as an interaction, it leads to a vertex of the form

in which $\xi$ turns into an $A$. This means that neither $\xi$ nor $A$ exists as an independent free particle and that we have therefore incorrectly identified the fundamental particles in the theory. We can solve this problem and get rid of the Goldstone's boson in only one step making use of the local gauge invariance of $\mathcal{L}$ in the original form.

Choosing in (1.40) $\theta$ to be

$$\theta = -\tan^{-1}(\phi_2/\phi_1)$$

(1.42)

will render $\phi'$ real, which is to say that $\phi'_2 = 0$, with $\xi$ now being equal to zero. In this particular gauge the Lagrangian reduces to:

$$\mathcal{L}' = \left[\frac{1}{2}(\partial_\mu \eta)^2 - \mu^2 \eta^2\right]$$

$$+ \left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\left(\frac{e \mu}{\lambda}\right)^2 A_\mu A^\mu\right]$$

$$+ 2i\left(\frac{e \mu}{\lambda}\right)A_\mu \partial^\mu \xi + \text{quadratic and cubic terms in } \eta$$

(1.43)
By an astute choice of gauge, we have eliminated the Goldstone boson and the offending term in $\mathcal{L}$; we are left with a massive scalar $\eta$ (the “Higgs” boson) and a massive gauge field $A_\mu$.

1.9 Choice of the Higgs Field in the EW theory

The electroweak Lagrangian resulted from imposing $SU(2)_L \otimes U(1)_Y$ local gauge invariance for the electron-neutrino pair is:

$$
\mathcal{L}_{EW} = \bar{\chi}_L \gamma^\mu \left[ i \partial_\mu - g W_\mu - g_Y B_\mu \right] \chi_L \\
+ \bar{e}_R \gamma^\mu \left[ i \partial_\mu - g' Y_R B_\mu \right] e_R - \frac{1}{4} W^\mu W_\mu - \frac{1}{4} B^\mu B_\mu
$$

(1.44)

Where $\chi_L$ is one of the first three doublets in (1.21) and $t_2 = T$. $\mathcal{L}_{EW}$ embodies the weak isospin and hypercharge interactions. The final two terms are the kinetic energy and self coupling of the $W_\mu$ fields and the kinetic energy of the $B_\mu$ field. Remember that the mass terms for both the gauge bosons and the fermions do not appear in the Lagrangian as they would otherwise break the symmetry.

We want to introduce the Higgs mechanism so that fermions, $W^\pm$, and $Z^0$ become massive and the photon remains massless.

To do this we introduce four real scalar fields and add to $\mathcal{L}_{EW}$ an $SU(2)_L \otimes U(1)_Y$ gauge invariant Lagrangian for these scalar fields:

$$
\mathcal{L}_\phi = \left| \left( i \partial_\mu - g T \cdot W_\mu - g' Y B_\mu \right) \phi \right|^2 - V(\phi)
$$

(1.45)

To keep $\mathcal{L}_\phi$ gauge invariant, the $\phi_i$ must belong to $SU(2)_L \otimes U(1)_Y$ multiplets. The most economical choice is to arrange the four fields in an isospin doublet with $Y = 1$:

$$
\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad \text{with} \quad \begin{aligned}
\phi^+ &= (\phi_1 + i \phi_2) / \sqrt{2} \\
\phi^0 &= (\phi_3 + i \phi_4) / \sqrt{2}
\end{aligned}
$$

(1.46)

Note that this scalar field $\phi$ couples to the gauge bosons through the covariant derivative $\partial_\mu - ie A_\mu + ig T \cdot W_\mu$.

To generate the gauge boson masses we use the Higgs potential:

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

(1.47)

Since all the vacua are equivalent, there is freedom to choose one which is convenient to work with:

---

4As well as quarks
This isospin doublet has $Y = 1$, $\tau_3 = -\frac{1}{2}$ and $Q = 0$. This is the appropriate choice to make in order to break both $SU(2)$ and $U(1)_Y$ gauge symmetries, while preserving $U(1)_{EM}$ (being $\phi_0$ neutral).

The gauge boson masses are identified by substituting the vacuum expectation value $\phi_0$ for $\phi(x)$ in $\mathcal{L}_\phi$.

The Lagrangian is then left with two terms: one, diagonal in the $W^3_\mu$ and $B_\mu$ basis, corresponds to the mass term for the charged bosons, the other, which is off-diagonal in the same basis, becomes diagonalized if read in terms of the new physical fields $Z_\mu$ and $A_\mu$ [24]. $Z_\mu$ and $A_\mu$ are defined as follows:

$$A_\mu = \cos\theta_W B_\mu + \sin\theta_W W^3_\mu$$

$$Z_\mu = -\sin\theta_W B_\mu + \cos\theta_W W^3_\mu$$

where $\theta_W$ is:

$$\tan\theta_W = \frac{g'_W}{g_W}$$

The same Higgs doublet which generates $W^\pm$ and $Z$ masses is also sufficient to give masses to the leptons and quarks.

Due to gauge invariance, rather than using $\phi_0$, we can apply the transformation $\phi \rightarrow \phi' = U(\xi)\phi = e^{i\xi(x) T/2v}\phi$, so that:

$$\phi'(x) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$$

The neutral Higgs field $h(x)$ is the only remnant of the Higgs initial doublet (composed of four fields) after the spontaneous symmetry breaking has taken place and the other three fields have been gauged away.

Substituting (1.52) in the Lagrangian that describes the coupling between lepton doublets, the required lepton mass is finally generated. Analogously for the quark doublets.
Chapter 2

The ATLAS Detector

“I think the primary justification for this sort of science that we do is fundamental human curiosity. [...] It’s true, of course, that every previous generation that’s made some breakthrough in understanding nature has seen those discoveries translated into new technologies, new possibilities for the human race. That may well happen with the Higgs boson. Quite frankly, at the moment I don’t see how you can use the Higgs boson for anything useful.”

— John Ellis

2.1 Overview

The Large Hadron Collider (LHC) is a 27 km circumference accelerator installed in the LEP (Large Electron Positron Collider) tunnel, at CERN. It is the highest-energy accelerator ever built. Inside the LHC, bunches of more than $10^{11}$ protons collide up to 40 million times per second to provide 7 TeV proton-proton collisions at luminosities of up to $10^{34} \text{cm}^{-2}\text{s}^{-1}$.

There are four collision points in the LHC ring, where high energy particle physics experiments are installed. The LHC hosts six distinct experiments, each of which is characterized by its unique particle detector.

The two large experiments, ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid), are multi-purpose experiments. Searches of the SM Higgs boson as well as measurements of the EW parameters of the SM are carried out. The two detectors vary in their magnetic system and the technology used in their sub-detectors. In CMS the tracking and calorimeter (EM and hadronic) systems are contained in a superconducting solenoid delivering a magnetic field of 4 T. In ATLAS only the Inner Detector is surrounded by a 2 T solenoid. Having two independently designed detectors is vital for cross-confirmation of any new discoveries made.
Two medium-size experiments, ALICE (A Large Ion Collider Experiment) and LHCb (Large Hadron Collider beauty), have specialized detectors for analyzing the LHC collisions in relation to specific phenomena. ALICE is a detector dedicated to heavy ions collisions. Its aim is to study the quark-gluon plasma (QGP) in order to understand the matter confinement at early time of the Universe. LHCb experiment was designed to perform measurements on physics phenomena involving B mesons in order to understand the matter-antimatter asymmetry.

Two further experiments, TOTEM and LHCf (Large Hadron Collider forward), are much smaller in size. They are designed to focus on "forward particles" (protons or heavy ions). The TOTEM, for TOTal Elastic and diffractive cross section Measurements, studies diffractive physics. Studying elastic diffractive scattering at low momentum transfer enables to measure the absolute luminosity of the LHC and the total cross section of proton-proton collision. The LHCf detector is a special-purpose LHC experiment for astrophysics studies at high energy within the particles produced at large pseudo rapidity. It consists of two small detectors located at 140 m on each side of the interaction point.

The ATLAS, CMS, ALICE and LHCb detectors are installed in four huge underground caverns located around the ring of the LHC. The detectors used by the TOTEM experiment are positioned near the CMS detector, whereas those used by LHCf are near the ATLAS detector.

The ATLAS detector at LHC is 46 m long, 22 m in diameter and weighs 7000 tons. It is the largest detector installed at CERN. Fig. 2.1 shows the ATLAS subdetectors:

- the Inner Detector (ID) (surrounded by a solenoid magnet)
- the electromagnetic (EM) and hadronic calorimeters
- the muon spectrometer

Only the inner detector and the calorimeters will be here described. The muon chamber will not be illustrated as muons were not part of the analysis.

2.2 The ATLAS Coordinate System

The coordinate system of ATLAS is a right-handed coordinate system with the x-axis pointing towards the centre of the LHC tunnel, and the z-axis along the tunnel, in the beam direction. The y-axis is thus oriented upwards and together with the x-axis it defines a transverse plane with respect to the beam. Since the detector is by construction symmetric around \(z = 0\), a \(A\) side and a \(C\) side are defined, corresponding to positive and negative z domain respectively. The origin of
the coordinate system corresponds to the nominal beam spot, which is the collision point of the intersecting proton bunches.

The word “spot” is a bit of a misnomer, since the luminous region, i.e. the spacial region where collisions takes place, has rather the dimensions of a thin hair: along the beam direction it is a few centimeters long, while in the transverse direction is of the order of tens of micrometers.

The azimuthal angle \( \phi \) of a particle is the angle measured in the transverse plane from the \( x \)-axis to the particle direction in this plane.

The polar angle \( \theta \) is defined from the beam axis to the particle direction. The pseudorapidity \( \eta \) is related to the polar angle by the following relation:

\[
\eta = -ln(tan(\frac{\theta}{2}))
\]

where \( \eta = 0 \) corresponds to a particle going upward along the \( y \)-axis (\( \theta = 90^\circ \)), while larger values of \( \eta \) correspond to directions closer to the beam axis (\( \lim_{\theta \to 0^\circ} \eta \to \infty \)).

In terms of the momentum, the pseudorapidity variable can be written as:

\[
\eta = \frac{1}{2} ln \left( \frac{|p|+p_z}{|p|-p_z} \right)
\]

where \( p_z \) is the component of the momentum along the beam axis. In the limit where the particle mass is negligible with respect to its momentum, pseudorapidity is numerically close to the definition of rapidity:

\[
\eta = \frac{1}{2} ln \left( \frac{E+p_z}{E-p_z} \right)
\]

In the approximation of a null transverse momentum for the partons involved in the proton-proton collision, pseudorapidity is given by:
\[ E = \frac{1}{2} (x_1 + x_2) \sqrt{s}; \quad P_z = \left( \frac{1}{2} (x_1 - x_2) \right) \Rightarrow y = \frac{1}{2} \ln \left( \frac{x_1}{x_2} \right) \]

\[ x_i \text{ being the fraction of momentum of a proton carried by the parton } i \text{ involved in the process.} \]

Since the relativistically invariant phase space for a single particle of mass \( m \), momentum \( p \) and energy \( E \) is:

\[ d^4 \delta(p^2 - m^2) = d^3 p / E = \pi d \eta dp_T^2 \]

the particles produced in an inelastic reaction are expected to be found uniformly distributed in rapidity and \( \phi \) [22]. This suggests to use \( \eta \) and \( \phi \) in describing the detector geometry. The size of a cell in this plane is \( \Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} \) where \( \Delta \eta \) (\( \Delta \phi \)) is the length of the cell in the \( \eta \) (\( \phi \)) direction (\( \Delta \eta \sim \Delta \theta / \sin \theta \) when \( \Delta \eta \) is small).

In proton-proton collisions at LHC, the momentum carried by the two partons is not equal, \( x_1 \neq x_2 \), due both to the motions of partons inside the proton. This results in a longitudinal boost in the partonic interaction (a boost that the proton interaction does not have), which is small, but unknown \textit{a priori}. It follows that to pass from the the proton-proton center of mass frame (which coincides with the lab frame) to the partons center of mass frame, a Lorentz boost along \( z \) needs to be performed. Variables which are invariant passing from the center of mass system of the protons (which coincides with the laboratory system) to the center of mass system of the partons are therefore preferred. The \textbf{transverse momentum} \( p_T \), defined as the momentum perpendicular to the beam axis, as well as the \textbf{transverse energy}, \( E_T \), are quantities invariant passing from one system to the other.

For the reasons mentioned above, particles are described by the following coordinates:

\[ (E, p_T, \eta, \phi) \]

which are equivalent to the four-momentum:

\[ p = (E, p_x, p_y, p_z) \]

### 2.3 The Inner Detector

The basic function of the Inner Detector (ID) is to track charged particles by detecting their interaction with material at discrete points.

From the track reconstruction one can extract the charge of a particle (negative or positive) and, by combining all the tracks together, identify primary vertices as well as secondary vertices. The ID also provides for high-precision measurements of charged particles momentum and direction. The 2 T \textit{magnetic field} generated by the \textit{central solenoid} [11] surrounding the entire inner detector causes particles to curve:
The tracking system consists of three different subsystems:

- the Pixel Detector
- the Semi-Conductor Tracker (SCT)
- the Transition Radiation Tracker (TRT)

The Pixel Detector

The precision tracking detectors (pixel and SCT) cover the region $|\eta| < 2.5$. In the barrel (central) region they are arranged on concentric cylinders around the beam axis while in the end-cap regions, they are located on disks perpendicular to the beam axis.

The Pixel Detector provides an accurate measurement of charged particle tracks close to the interaction point. It is required to have a fast and accurate pattern recognition in a high multiplicity environment, an accurate vertexing algorithm and a good resolution on the transverse impact parameter.

The Pixel Detector consists of three layers of detectors in the barrel region. The end-caps are made of three wheels installed at a $z$ position between 49 and 65 cm from the interaction point, with a radius ranging from approximately 90 to 150 mm.

The pixel layers are segmented in $R - \phi$ and $z$ with typically three pixel layers crossed by each track. All pixel modules are identical in size, $\Delta R - \phi \times z = 50 \times$.

$^3$The transverse momentum of the colliding partons is essentially null, being, on average, 200-200 MeV/c.
Figure 2.3 – Schematic view of the helical track of a charged particle projected onto the x-y plane.

400 \mu m^2. The resolution are 10 \mu m (R - \phi) and 115 \mu m (z) both in in the barrel and in the disks.

The innermost layer of the Pixel detector is the so called B-layer, as it is used to tag b-jets coming from the hadronization of b-quarks.

The identification of b-quarks takes mainly advantage of the relatively long lifetime of hadrons containing a b-quark, which is of the order of 1.5 ps (c\tau \approx 450\mu m). A b-hadron in a jet with p_T = 50GeV will therefore have a significant flight path length, traveling on average about 3 mm in the transverse plane before decaying, leading to a secondary vertex distinct from the primary vertex.

The transverse impact parameter, d_0, is the distance of closest approach of the track to the primary vertex point, in the R - \phi projection. The longitudinal impact parameter, z_0, is the z coordinate of the track at the point of closest approach in R - \phi . The tracks from b-hadron decay products tend to have rather large impact parameters which can be distinguished from tracks stemming from the primary vertex. The other more demanding option is to reconstruct explicitly the displaced vertices. Both of these two approaches require a high accuracy in the measure of the tracks near the interaction point.

The SCT

The Semiconductor Tracker (SCT) is located after the Pixel Detector. It is made of Silicon microstrip sensors. It consists of eight strip layers in the barrel region, each of which provides a 2-dimensional coordinate per charged particle track. The end-caps are made of eighteen forward and backward disks, with strips running radially, located a distance from 80 to 280 cm from z = 0. The microstrip sensors installed in SCT modules consist of 126 mm long chained sensors with a strip pitch of 80 \mu m. Each module is made of two detectors slightly tilted by 40 mrad stereo angle to measure the z position. 4088 modules are assembled to form the barrel and end-caps for a total area covered by the SCT of 63 m. In both the barrel and end-cap regions, a resolution \sigma_{R-\phi} \times \sigma_z = 580 \times 17 \mu m^2 is obtained.
The TRT

The Transition Radiation Tracker (TRT) system is made of straw tubes of 4 mm diameter filled with a gas made of 70% Xenon, 20% Methane and 10% CO₂. A 30 µm diameter (Au-covered) Tungsten string, placed at the center of each tube, constitutes the anode of the system, that collects electrons after the gas has been ionized by the passage of a charged particle. The barrel part is made of 3 layers of 150 cm height long straws while the end-caps are made of eighteen wheels containing 39 to 55 cm height long straws oriented radially. The TRT coverage enables track trajectory reconstruction up to |\eta| \sim 2.0 – 2.1, as shown in 2.4. It provides a measurement of the charged particle trajectory with typically 30-36 hits per track. The TRT is also used to discriminate electrons from heavier charged particle, such as pions.

2.4 Calorimetry

The electromagnetic (EM) calorimeter of ATLAS is located after the ID and measures the energy and direction of photons and electrons in a large energy dynamic range (from few tens of GeV to 2-3 TeV). The hadronic calorimeter follows the EM calorimeter and measures the energy and direction of hadronic jets originating from quarks and gluons. The two calorimeters comprises of a barrel covering |\eta| \lessapprox 1.5 and two end-caps covering the region 1.5 \lessapprox |\eta| \lessapprox 3.2. Dedicated forward calorimeters are used at larger \eta, namely 3.1 \lessapprox |\eta| \lessapprox 4.9, to ensure a full coverage by the calorimetry in the |\eta| \lessapprox 5 region. The calorimeters are shown in Fig. 2.5. The description for both the EM and hadronic calorimeter will consists of three parts:

1. the physics of the particle shower
2. the resolution of the calorimeter
Figure 2.5 – View of the ATLAS calorimeter systems. The barrel and end-cap parts of the EM and hadronic calorimeters are shown. The forward calorimeters are illustrated as well.

3. the description of the detector

2.4.1 The Electromagnetic Calorimeter

Physics of the electromagnetic cascade

As suggested by its name, a calorimeter measures energy. High energy electrons, positrons, and photons traversing matter interact through well-known electromagnetic processes described by QED processes [29].

The average energy lost by electrons in lead and the photon interaction cross-section are shown in Fig. 2.6 and 2.7 as a function of energy respectively. Two main regimes can be identified.

For energies larger than $\sim 10$ MeV, the main source of electron energy loss is bremsstrahlung. In this energy range, photon interactions produce mainly electron–positron pairs. For energies above 1 GeV both these processes become roughly energy independent. At low energies, on the other hand, electrons lose their energy mainly through collisions with the atoms and molecules of the material thus giving rise to ionization and thermal excitation; photons lose their energy through Compton scattering and the photoelectric effect.

As a consequence, electrons and photons of sufficiently high energy ($\geq 1$ GeV) incident on a block of material produce secondary photons by bremsstrahlung, or secondary electrons and positrons by pair production. These secondary particles in turn produce other particles by the same mechanisms, thus giving rise to a cascade (shower) of particles with progressively degraded energies. The number of particles in the shower increases until the energy of the electron component falls below a critical energy $E_c$, where energy is mainly dissipated by ionization and excitation and not in the generation of other particles. $E_c$ is defined as the energy at which the electron ionization losses and bremsstrahlung losses become equal.
The longitudinal and lateral developments of electromagnetic showers can be described in terms of one parameter, the radiation length $X_0$, which depends on the characteristics of the material:

$$X_0 (g/cm^2) \simeq \frac{716 g cm^{-2} A}{Z(Z+1) \ln(287/\sqrt{Z})}$$

(2.1)

where $Z$ and $A$ are the atomic number and weight of the material, respectively.

At high energy, when electrons lose energy mainly by bremsstrahlung, the radiation length is equal to the path traveled by an electron to reduce its energy to $1/e$ of its original energy $E_0$:

$$E(x) = E_0 e^{-x/X_0}$$

(2.2)

Similarly, a photon beam of initial intensity $I_0$ traversing a block of material is absorbed mainly through pair production. After traveling a distance $x = \frac{9}{7} X_0$, its intensity is reduced to $1/e$ of the original intensity:

$$I(x) = I_0 e^{-\frac{7}{9} x/X_0}$$

(2.3)

Another definition for $E_c$, which makes use of the parameter $X_0$, is the following:

$$E_c : \frac{dE}{dx_{\text{ionization}}} = \frac{E}{X_0}$$

(2.4)

This definition and the one given above are equivalent in the approximation:

$$\frac{dE}{dx_{\text{bremsstrahlung}}} \simeq \frac{E}{X_0}$$

(2.5)

Equations (2.2) and (2.3) show that the physical scale over which a shower develops is similar for incident electrons and photons, and is independent of the material type if expressed in terms of $X_0$. Therefore electromagnetic showers can be described in a universal way by using simple functions of the radiation length.

For instance, the mean longitudinal profile can be described (Longo and Sestili, 1975):

$$t_{\text{max}} \simeq \ln \frac{E_0}{E_c} + t_0$$

(2.6)

where $t = x/X_0$ is the depth inside the material in radiation lengths and $a$ and $b$ are parameters related to the nature of the incident particle ($e^\pm$ or $\gamma$). The shower maximum, i.e., the depth at which the largest number of secondary particles is produced, is approximately located at:

$$t_{\text{max}} \simeq \ln \frac{E_0}{E_c} + t_0$$

(2.7)

where $t_{\text{max}}$ is measured in radiation lengths, $E_0$ is the incident particle energy, and $t_0 = -0.5 (+0.5)$ for electrons (photons). This formula shows the logarithmic dependence of the shower length, and therefore of the detector thickness needed to
Figure 2.6 – Fractional energy lost in lead by electrons and positrons as a function of energy (Particle Data Group, 2008).

absorb a shower, on the incident particle energy. The calorimeter thickness containing 95% of the shower energy is approximately given by:

\[
t_{95\%} \approx t_{\text{max}} + 0.08Z + 9.6
\]  

(2.8)

where \( t_{\text{max}} \) and \( t_{95\%} \) are measured in radiation lengths. In calorimeters with thickness \( \approx 25X_0 \), the shower longitudinal leakage beyond the end of the active detector is much less than 1% up to incident electron energies of \( \sim 300 \text{ GeV} \). Therefore, even at the particle energies of the LHC, electromagnetic calorimeters are very compact devices: the ATLAS lead-liquid argon calorimeter has a thickness of \( \approx 45 \text{ cm} \), with a radiation length of \( \approx 1.8 \text{ cm} \).

The transverse size of an electromagnetic shower is mainly due to multiple scattering of electrons and positrons away from the shower axis. Bremsstrahlung photons emitted by these electrons and positrons can also contribute to the shower spread. A measurement of the transverse size, integrated over the full shower depth, is given by the Molière radius \( (R_M) \), which can be approximated by:

\[
R_M(\text{g/cm}^2) \approx 21 \text{ MeV} \frac{X_0}{E_c(\text{MeV})}
\]  

(2.9)

It represents the average lateral deflection of electrons at the critical energy after traversing one radiation length.

On average, about 90% of the shower energy is contained in a cylinder of radius \( \sim 1R_M \). Since for most calorimeters \( R_M \) is of the order of a few centimeters, electromagnetic showers are quite narrow [18].

Energy resolution of electromagnetic calorimeters

The measurement of energy with an electromagnetic calorimeter is based on the principle that the energy released in the detector material by the charged particles
2.4. Calorimetry

Figure 2.7 – Photon interaction cross-section in lead as a function of energy (Fabjan, 1987).

of the shower, mainly through ionization and excitation, is proportional to the energy of the incident particle.

The energy deposited is then converted into a number of “carriers”, that are collected and originate the measurement. This number is subject to random fluctuations and gives origin to the intrinsic resolution of an ideal calorimeter:

\[ \sigma(E) \sim \sqrt{E} \rightarrow \frac{\sigma(E)}{E} \sim \frac{a}{\sqrt{E}} \quad (2.10) \]

\( a \) is called the sampling term. Since in a sampling calorimeter only a fraction of the energy is converted into “carriers”, its resolution is worse than in an homogeneous one. The real sampling term of a calorimeter is also related to the fraction of carriers that are converted in signal (e.g. photons converted in a pulse etc.).

The energy resolution of the EM calorimeter is also affected by the noise introduced by the read out chain (photomultipliers, preamps etc...), which is constant and important at low energies (term \( b \) in Eq. (2.11)). Any instrumental effect, which produces response variations in the detector (e.g. detector geometry, imperfections in the mechanics or readout, temperature gradients, non-uniform aging, radiation damage etc...) might also affect the detector performance. Such additional term dominates at high energy (term \( c \) in Eq. (2.11)).

Including all these terms, the calorimeter resolutions reads as follows:

\[ \frac{\sigma(E)}{E} \sim \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (2.11) \]

where the symbol \( \oplus \) indicates a quadratic sum. The typical energy resolution of sampling electromagnetic calorimeters is in the range \( 5 - 20\%/\sqrt{E} \text{[GeV]} \). The noise contribution is usually required to be much smaller than 100 MeV per electronic channel. The constant term should instead be kept at the level of 1% or smaller.
The ATLAS EM Calorimeter

The ATLAS EM calorimeter is a sampling calorimeter made of lead absorbers interleaved with Liquid Argon (LAr). In a sampling calorimeter, as opposed to a homogeneous calorimeter, the material that produces the particle shower is distinct from the material that measures the deposited energy. One advantage of this is that each material can be well-suited to its task; for example, a very dense material (e.g. lead) can be used to produce a shower that evolves quickly in a limited space, even if the material is unsuitable for measuring the energy deposited by the shower. A disadvantage is that some of the energy might be deposited in the passive material and cannot be measured. Corrections to the energy measured must be applied.

An electrode is placed at the center between two absorbers to collect the LAr ionization signal generated by the passage of a charged particle. The EM barrel calorimeter (EMB) and the two end-caps (EMEC) are in different cryostats at the temperature of the LAr (88.5 K).

The EMB ($|\eta| < 1.4$) is made of two half-barrels, symmetric in $z$, which have an inner and an outer radius of $R_{in} = 1.15 m$ and $R_{out} = 2.25 m$, respectively. Each half-barrel is made of 16 modules along $\phi$ (of size $2\pi/16$) containing 64 $3.2 m$ long absorbers. The modules are made of two types of electrodes, commonly named electrodes $A$ and $B$, with the $2.5 mm$ transition region arising at $|\eta| = 0.8$. A high voltage of 2000 V is applied on both sides of the electrode.

The EMEC is made of two 63 cm thick wheels, covering a range in rapidity $1.475 < |\eta| < 3.2$ and having inner and outer radius of $R_{in} = 30 cm$ and $R_{out} = 2.1 m$, respectively.

2.4.2 The Hadronic Calorimeter

Physics of the hadronic cascade

Hadrons interact with nuclei through the strong interaction and produce charged and neutral particles (mainly $\pi^0$). The process goes on up to an energy of secondary hadrons below the threshold energy for $\pi$ production ($\sim 240 \text{ MeV}$). The particles produced by the multiple interactions with the material form what is called a hadronic shower. A hadronic shower can be described by the nuclear interaction length of the hadron, $\lambda_I$, which is mean free path between two interactions.

Hadronic showers are more complex than EM showers:

- Up to 30% of incident energy may be lost (invisible energy) due to:
  - nuclear excitation and break-up
  - spallation or "evaporation" of slow neutrons and protons
  - production of muons (recovered with the $\mu$ spectrometer) and neutrinos which escape from the calorimeter.
• The inelastic cross section, and hence the nuclear interaction length, is a function of both the energy and type of incoming particle.

• Fluctuations in the amount of energy deposited are largely due to the variable fraction of the shower which is converted into an electromagnetic shower, by the production of fast neutral pions and their subsequent rapid decay into energetic photons.

As the number of energetic hadronic interactions increases with incident energy, so will the fraction of the electromagnetic cascade. For the hadronic fraction one finds $F_h = (E/E_0)^k$ with $k = ln \alpha/ln m$. The parameter $E_0$ denotes a cutoff for further hadronic production, typically $E_0 \approx 1-2 GeV$; $m$ is the multiplicity of fast hadrons produced in a hadronic collision; the parameter $\alpha$ gives the fraction of hadrons not decaying electromagnetically; the value of $k$ is $\approx -0.2$. Values of $F_h$ are of order 0.5 (0.3) for a 100 (1000) GeV shower. As the energy of the incident hadron increases, it is doomed to dissipate its energy in a flash of photons [18].

Let $\eta_e$ be the efficiency for observing a signal $E_{\text{vis}}^{e}$ (visible energy) from an electromagnetic shower, i.e., $E_{\text{vis}}^{e} \equiv \eta_e E^{(em)}$; let $\eta_h$ be the corresponding efficiency for purely hadronic energy to provide visible energy in the detector. Therefore, for a pion-induced shower the visible energy $E_{\text{vis}}^{\pi}$ is:

\[
E_{\text{vis}}^{\pi} = \eta_e F_{\pi} E + \eta_h F_h E
\]

\[
= \eta_e \left( F_{\pi} + \frac{\eta_h}{\eta_e} F_h \right) E
\]

where $E$ is the incident pion energy. The ratio of observable, i.e. 2visible”, signals induced by electromagnetic and hadronic showers, usually denoted $e/\pi$, is therefore:

\[
\frac{E_{\text{vis}}^{\pi}}{E_{\text{vis}}^{e}} = \left( \frac{e}{\pi} \right)^{-1} = 1 - \left( 1 - \frac{\eta_h}{\eta_e} \right) F_h
\]

The relative response $e/\pi$ turns out to be the most important yardstick for gauging the performance of a hadronic calorimeter.

**Energy resolution of hadron calorimeters**

The properties of the hadronic cascade determine the intrinsic fluctuations and hence the energy resolution. Inescapably, hadronic cascades imply nuclear interactions with their correlated invisible energy. With less energy measurable from a hadronic shower than from an electromagnetic shower, we expect that on average for particles with the same incident energy the signal response to hadrons will be lower, i.e., $e/\pi > 1$. Event by event the visible energy will fluctuate between two extremes: fully electromagnetic, yielding the same signal as an electron, or fully hadronic with a maximum of invisible energy.

If $e/\pi \neq 1$ then:
1 Introduction

Many collaborations, including CDF \cite{1}, CMS \cite{2} and ALEPH \cite{3}, have been able to improve the resolution of the measured particle energies by replacing the energy deposited by electrically charged particles measured in the calorimeter detectors with the momentum measured in the charged particle tracking detectors. This method makes use of the fact that the resolution in the charged particle tracking systems is better than the energy measurements in the calorimeters at low energies. Therefore "energy flow" (also known as "particle flow") algorithms are expected to improve the energy resolution of hadronic objects, such as jets and taus, as well as the Missing Transverse Energy ($E_T$). The calorimeter resolution (equation 1) and the tracking resolution (equation 2) for single pions at $d_\eta = 0$ in the ATLAS detector are shown in Figure 1 \cite{4}. The curves cross over at around 150 GeV, above which the calorimeter has superior resolution compared to the charged particle tracking system.

\begin{align}
\frac{\sigma(E)}{E} &\approx 50\% \oplus 3\% \quad (2.16) \\
\frac{\sigma(p_t)}{p_t} &\approx 0.036p_t\% \oplus 1.3\% \quad (2.17)
\end{align}

This means that the resolution in the charged particle tracking systems is worse than the energy measurements in the calorimeters at high energies.

Fig. 2.8 \cite{10} shows that the curves describing the calorimeter and the tracking resolution cross over at around 150 GeV, above which the calorimeter has superior resolution.

The energy resolution, which no longer scales with $1/\sqrt{E}$, is usually approximated by:

\[ \frac{\sigma}{E} = \frac{a_1}{\sqrt{E}} \oplus a_2 \quad (2.15) \]

where a "constant" term $a_2$ is added quadratically.

The energy resolution for the hadronic calorimeter for single pions at $\eta = 0$ is:

\[ \frac{\sigma(E)}{E} \approx 50\% \oplus 3\% \]

The $p_t$ resolution for the tracking system for single pions at $\eta = 0$ is instead:

\[ \frac{\sigma(p_t)}{p_t} \approx 0.036p_t\% \oplus 1.3\% \]

fluctuations in $F_{\eta 0}$ are a major component of the energy resolution;

the average value for $F_{\eta 0}$ is energy dependent and therefore calorimeters have a response to hadrons which is non-linear with energy;

the above-mentioned fluctuations are non-Gaussian and therefore the energy resolution scales weaker than $1/\sqrt{E}$.

Figure 2.8 – Resolution of Single Pions at $\eta = 0$ in Calorimeter and Charged Particle Tracking Detectors
2.4. Calorimetry

The Tile Calorimeter

As can be seen from Fig. 2.9, the tile calorimeter is located outside the EM calorimeter, extending radially from 2.28 m to 4.25 m.

The nuclear interaction length of "active" detector material (the material that is designed to absorb the energy of the incident particle) is so large that ATLAS calorimeters were chosen to be sampling devices.

The ATLAS sampling calorimeter uses steel as the absorbing material and scintillating tiles as the active material (from which its name “tile” calorimeter. The hadronic barrel covers \( \eta < 1.0 \) and is complemented by two extended barrels covering \( 0.8 < |\eta| < 1.7 \); each of these subdetectors is divided azimuthally into 64 modules in \( \phi \).

The barrel and extended barrel tile, have a radial depth of approximatively 7.4 \( \lambda_I \). Accounting for the contribution of the EM calorimeter, the calorimeters amount to about 11 \( \lambda_I \) in total.

The end-cap hadronic calorimeter (HEC) is a Liquid Argon-copper sampling calorimeter covering the range \( 1.5 < |\eta| < 3.2 \). It is therefore contained in the end-cap cryostat. The HEC consists of an inner wheel, made of 25 copper absorbers of 25 mm thickness separated by a Liquid Argon 8.5 mm thickness gap, and an outer wheel, made of 17 copper absorbers of twice the thickness of gap. A high voltage of 1800 V is applied in the end-cap wheels.

The Forward Calorimeter

The Forward calorimeters (FCALs) are relatively far from the interaction point (4.7 m). They are located in the same cryostat where the EMEC and HEC calorimeters are placed, and cover a range in rapidity \( 3.1 < |\eta| < 4.9 \). These detectors are made of three modules, which differ in their absorber component and gap size. The gaps are smaller than for the other calorimeters to compensate the high multiplicity of charged particles at large \( \eta \).
2.4.3 Calorimeter Jets

The ATLAS calorimeter system has about 200,000 individual cells of various size and different readout technologies and electrode geometries. For jet finding it is necessary to first combine these cell signals into larger signal objects, which are regarded as reconstructed energy deposits, with physically meaningful four-momenta. The two concepts available are calorimeter signal towers and topological cell clusters. Since the jet algorithm used in the present analysis (Cambridge-Aachen) makes use of the latter, a brief description will only be done for it.

Topological Jet Clusters

The Topological Cluster algorithm starts from calorimeter cells with a signal-to-noise ratio, or signal significance \( \Gamma = \frac{E_{\text{cell}}}{\sigma_{\text{noise, cell}}} \), above a certain threshold \( S \), i.e. \( \Gamma > S = 4 \). These cells are used as seeds for the topological clusters. All contiguous cells, in all three dimensions, with \( \Gamma \) greater than a secondary threshold \( N \), i.e. \( \Gamma > S = 2 \), are added to the seed cells. Finally, a ring of guard cells with signal significance above the basic threshold \( \Gamma > P = 0 \) is added to the cluster. The energy cluster at this point is measured at the electromagnetic scale. This is the raw signal from the ATLAS calorimeters, without accounting for any correction. The nomenclature “electromagnetic scale” is used both for EM and hadronic calorimeters in ATLAS [11].

For the EM calorimeter signals, electromagnetic scale indicates that this scale has been derived from electron signals, but it lacks all corrections applied in high precision electron or photon reconstruction, such as electronic corrections and the geometrically motivated corrections for high voltage problems, like inactive electrode sub-gaps and similar.

For the hadronic calorimeter, the situation is slightly more complicated, since hadrons, as seen above, do not interact in a way that their energy is all deposited in the detector (oppositely to electrons and photons). Production of neutrons that escape the detector, energy loss in inactive materials, secondary particles that cannot be measured, non-interacting neutrinos being produced or energy being used to create nuclear states that are not observable may contribute to measure the wrong hadronic energy. Thus, apart from the corrections already mentioned for the EM calorimeter (mainly due to the electronics), in ATLAS, Local Hadron Calibration is provided to correct topological clusters to the hadronic particle energy scale. Then if such clusters are used in jet reconstruction further corrections must be applied to the jets to account for algorithm specific effects (e.g. low momentum particles not reaching the calorimeter). Alternatively one can input clusters at the uncalibrated electromagnetic scale to a jet reconstruction algorithm. Then, at the end, a correction can be applied to the jet (or jet constituents, the topological clusters that make up the jet) to account for both the non-compensation of the calorimeter and jet algorithm effects [26].

Fig. 2.10 shows the average number of particles in Monte Carlo (see Appendix A) generated jets from QCD di-jet production together with the number of topologi-
2.5. Trigger System

At the design luminosity of $10^{34} \text{cm}^{-2}\text{s}^{-1}$, the proton-proton interaction rate is approximately 1 GHz. However, the rate at which data can be recorded is limited to approximately 200 Hz. This means that some of the events have to be discarded, namely the uninteresting ones. Defining the readout timing and only recording interesting bunch crossing (while avoiding biases) at a given rate based on what the detector observed are two functions that the trigger system controls.

The ATLAS trigger system has three distinct levels: Level 1 (L1), Level 2 (L2), and the event filter (EF). Each level of the trigger refines the decision made by previous levels by applying additional selection criteria where appropriate.

Initially, event information is accepted from the readout electronics and buffered. The L1 trigger is a hardware-based system which uses a subset of the available detector information to reject events; budgeting 2.5 ms of processing time per event, it is able to reduce the overall rate to 75 kHz.

These events pass to L2 and EF, collectively known as the high-level trigger

---

**Figure 2.10** – Average number of particles in MC generated jets from QCD di-jet production together with the number of topological cell clusters in the corresponding simulated calorimeter jets as a function of the jet pseudorapidity.
Each HLT trigger is seeded from a specific lower level trigger and is able to examine relevant features of the event in greater detail in order to make an overall trigger decision. Events passing the triggers selection are passed to the data acquisition system. Collision events that are not triggered on are lost forever.

The central trigger processor implements a ‘trigger menu’, comprising a list of trigger selection criteria at all three levels, together with relevant prescales. The trigger accepts an event only if one or more selection criteria are satisfied.

The prescale is factor associated with a trigger at each level that indicated what fraction of events that could pass this trigger selection is actually accepted. In an ideal world of unlimited data storage, triggers would run un-prescaled. However, prescaling triggers is unfortunately unavoidable. 'Primary' triggers, that apply tight selections and are used for physics measurements and searches, are ensured to run un-prescaled.

**L1 Trigger**

Due to the limited time available, the L1 trigger only uses a small amount of the available detector information - it uses only the calorimeter and muon systems. The L1 system aims to identify high $p_T$ leptons, photons and jets as well as large $E_{\text{miss}}$ or $E_{\text{total}}$. High $p_T$ muons are identified using trigger chambers in the barrel and endcap regions of the muon spectrometer. Calorimeter selections are based on reduced-granularity information from all parts of the calorimeter system.

Events passing the L1 trigger selection are transferred to the next stages of the detector-specific electronics and subsequently to the data acquisition via point-to-point links. In each event, the L1 trigger also defines one or more Regions-of-Interest (RoI’s): the geographical coordinates in $\eta - \phi$ space of those regions within the detector where its selection process has identified interesting features. The RoI data includes information on the type of feature identified and the criteria passed. This information is passed to the HLT for later use.

**High-level trigger**

The RoI information from L1 is used to seed the L2 trigger. L2 triggers are able to use all of the available detector data in the region covered by the RoIs at full granularity and precision; this corresponds to approximately 2% of the total event data. An average event processing time of 40 ms is allowed at this stage, resulting in an output event rate of approximately 3.5 kHz. The event filter performs the final stage of event selection, implementing analysis procedures similar to those performed offline to reduce the final event rate to roughly 400 Hz.
Chapter 3

Boson production at the LHC

3.1 W and Z boson production and decay modes

W and Z production

The W and Z bosons can be singly produced in pp collisions via the Drell-Yan process $\bar{q}q \rightarrow W, Z$.

While in $\bar{p}p$ collisions (like at Tevatron) both quark and anti-quark are in most cases valence quarks, in pp collisions, at LHC, the anti-quark is coming from the sea.

First let us look at the W boson production. Charge conservation requires an up-type (u, c, t) and a down-type (d, s, b) antiquark to interact. In the case of resonant scattering, the scale of the process is directly related to the momenta of the incoming partons and the W mass:

$$Q^2 = s x_1 x_2 = M_W^2$$  \hspace{1cm} (3.1)

where $\sqrt{s}$ is the center-of-mass energy of the colliding beams and $x_1, x_2$ the fractions of momentum carried by the colliding quarks. Since u quarks in the proton carry on average more momentum than d quarks, due to the presence of an additional valence quark, $x(u) > x(d)$, it is kinematically more likely for a $ud$ combination to satisfy Eq. (3.1) than a $d\bar{u}$ combination. As a result more $W^+$ relative to $W^-$ are produced.

The parton decomposition of the $W^+$ and $W^-$ total cross sections as a percentage of the total cross section is shown in Fig. (3.2).

Both figures 3.2 and 3.3 show how the relative contributions of the various $\bar{q}q$ processes change with collider energy. We split the collider energy range at $\sqrt{s} = 4TeV$, and assume proton–antiproton and proton–proton collisions below and above this value.

The dominant mechanism for W production in pp collisions is valence-sea scattering of u and d quarks, where $ud \rightarrow W^+$ and $ud \rightarrow W^-$. At the LHC energy of $\sqrt{s} = 7TeV$, about 10% of the total is associated with charm-strange scattering. These sea-sea processes dominate over the valence-sea contributions like up-strange,
that are across families and hence Cabibbo suppressed, and contribute at the percentage level. This shows the increased importance of the sea in the LHC kinematic region, compared to the Tevatron ($p\bar{p}$ collisions), where the charm-strange contribution is only 5%.

The corresponding situation for the $Z^0$ boson is shown in Fig. 3.3. At LHC energies we have $u\bar{u} \sim d\bar{d} \gg s\bar{s} \gg c\bar{c}$, in line with the ordering of the parton distributions inside the parton.

The $W$ and $Z$ decay widths are small ($\Gamma_W = 2.08\text{ GeV}$ and $\Gamma_Z = 2.50\text{ GeV}$ in the SM) compared to their masses, and so it is sufficient to consider the production of effectively stable particles, multiplying the cross sections by the appropriate final-state branching ratios. At leading order, the $\bar{q}q \rightarrow W, Z$ cross sections are obtained from the couplings of the gauge bosons to fermions:

\[
\hat{\sigma}^{\bar{q}q \rightarrow W} = \frac{\pi}{3} \sqrt{2} G_F M_W^2 |V_{q\bar{q}}|^2 \delta(\hat{s} - M_W^2) \\
\hat{\sigma}^{\bar{q}q \rightarrow Z} = \frac{\pi}{3} \sqrt{2} G_F M_Z^2 |V_q|^2 + A_q^2 |2(\hat{s} - M_Z^2) | (3.3)
\]

where $V_{q\bar{q}}$ is the appropriate CKM matrix element [15].

The leading - and next-to-leading-order diagrams for the Drell-Yan process are shown in Fig. 3.1. There are three classes of contributions:

(a) virtual gluon corrections to the leading-order contribution, (b) real gluon corrections from $q(p_1) + \bar{q}(p_2) \rightarrow W + g(k)$ and (c) the quark-gluon scattering process $q(p_1) + g(p_2) \rightarrow W + q(k)$ together with the corresponding $\bar{q}g$ contribution [15].

W and Z decay modes

The main decay modes for the $Z^0$ and the $W^+$ boson are listed in Table (3.2) and (3.1) respectively. $W^-$ modes are charge conjugated of the $W^+$ modes.
3.1. W and Z boson production and decay modes

**Figure 3.2** – Decomposition of the $W^+$ (solid line) and $W^-$ (dashed line) total cross sections in $\bar{p}p$ and $pp$ collisions by parent quark flavors. We split the collider energy range at $\sqrt{s} = 4$ TeV, and assume proton–antiproton and proton–proton collisions below and above this value. Due to the symmetry of $\bar{p}p$ collisions, the $W^+$ and $W^-$ have the same flavor parents, while differences are present for $pp$. The differing contributions of each flavor result from the differing CKM coefficients for each quark pairing and the composition of the proton [27].

<table>
<thead>
<tr>
<th>$W^+$ DECAY MODES</th>
<th>Fraction ($\Gamma_i/\Gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l^+\nu$</td>
<td>$(10.80 \pm 0.09)$%</td>
</tr>
<tr>
<td>$e^+\nu$</td>
<td>$(10.75 \pm 0.13)$%</td>
</tr>
<tr>
<td>$\mu^+\nu$</td>
<td>$(10.57 \pm 0.15)$%</td>
</tr>
<tr>
<td>$\tau^+\nu$</td>
<td>$(11.25 \pm 0.20)$%</td>
</tr>
<tr>
<td>hadrons</td>
<td>$(67.60 \pm 0.27)$%</td>
</tr>
</tbody>
</table>

**Table 3.1** – $W^+$ decay modes. Values taken from the Particle Data Book.
Figure 3.3 – Parton decomposition of the $Z^0$ total cross sections in $\bar{p}p$ and $pp$ collisions. Individual contributions are shown as a percentage of the total cross section.

<table>
<thead>
<tr>
<th>$Z^0$ DECAY MODES</th>
<th>Fraction ($\Gamma_i/\Gamma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l^+l^-$</td>
<td>(3.3658 ± 0.0023)%</td>
</tr>
<tr>
<td>$e^+e^-$</td>
<td>(3.363 ± 0.004)%</td>
</tr>
<tr>
<td>$\mu^+\mu^-$</td>
<td>(3.366 ± 0.007)%</td>
</tr>
<tr>
<td>$\tau^+\tau^-$</td>
<td>(3.367 ± 0.008)%</td>
</tr>
<tr>
<td>$\nu\bar{\nu}$</td>
<td>(20.00 ± 0.06)%</td>
</tr>
<tr>
<td>hadrons</td>
<td>(69.91 ± 0.06)%</td>
</tr>
</tbody>
</table>

Table 3.2 – $Z^0$ decay modes. Values taken from the Particle Data Book.
3.2. The Higgs production at LHC and its decay modes

**Higgs production at LHC**

The Higgs mass $M_H$ is the only unknown parameter of the SM. Once this parameter is fixed, the entire profile of the SM Higgs boson is uniquely determined. In particular, its couplings to the other particles, its production and decay rates can be calculated [7].

Fig. 3.4 display the production cross section of the SM Higgs boson at LHC as a function of its (unknown) mass, at a centre-of-mass energy of 7 and 14 TeV respectively.

The main process for the Higgs production is the **gluon fusion**, represented by the diagram 3.5(a). The Higgs boson couples only indirectly to gluons via a triangular loop of quarks, mainly the top. As shown in Fig. 3.4, this production process is dominant by several order of magnitude, in the whole range of Higgs mass that will be scanned at LHC.

The sub-leading Higgs production mode is the **Vector Boson Fusion** (VBF).
The Higgs boson is also produced in association with a W or Z boson, and also with a top quark pair, as shown in 2.1(c) and 2.1(d) respectively. The VBF and associated production modes are different compared to other processes since the Higgs is produced in association with something else in the final state. This allows a better discrimination between signal and background although the sensitivity is typically reduced at low luminosities.

All these production modes are considered inclusively in the study dedicated to the search of the Higgs boson through its decay into two photons.

2.2 The Higgs Boson Decay Mode in the SM

In the SM, the Higgs mass is a free parameter of the theory. It fixes the Higgs coupling to gauge bosons and fermions. The Higgs boson branching ratios (BR) are shown as a function of its mass in Figure 2.2. The Feynman diagrams of the different Higgs decay modes are illustrated in the same Figure. Three mass regions appear:

- In the low mass range ($100 \, \text{GeV} < m_H < 130 \, \text{GeV}$), the Higgs mainly decays into a $b\bar{b}$ pair, $\text{BR}_{b\bar{b}} \approx 70\%$, due to the Higgs coupling to fermions proportional to their mass. The discovery potential of Higgs search in this channel at LHC is however small due to large QCD background.

- The $H \to \tau^+\tau^-$ channel has a smaller branching ratio, about 8%, and suffers from large background from the $Z \to \tau^+\tau^-$ decay process. However, when searching the Higgs in the VBF production mode, this is a promising channel at low masses.

- The $H \to \tau^+\tau^-$ and $Z \to \tau^+\tau^-$ decay modes also occur in the low mass range and are rare due to the indirect Higgs coupling to the photon. The Higgs decay into two photons is made through fermions, mainly top quark, and W boson loops.

Figure 3.6 – The dominant decay processes of the Higgs boson including the loop and three-body decays (right); Standard Model Higgs decay Branching Ratio as a function of $m_H$ (left).

This process, illustrated by the diagram 3.5(b) presents a particular topology. The quarks interact via a W or a Z boson, which means that this is an electro-weak process and does not imply a color exchange between initial-state quarks. The quarks in the final state generate forward jets opposite in $\eta$, leading to a central rapidity gap where no QCD activity is expected, but the Higgs decay products.

The Higgs is also produced in association with a vector boson (W or Z), and also with a top quark pair, as shown in 3.5(c) and 3.5(d) respectively.

Higgs decay modes

Since the Higgs boson couples to particles proportionally to their masses, it will have the tendency to decay into the heaviest particle allowed by kinematics.

In the low mass range ($100 \, \text{GeV}/c^2 \leq m_H \leq 130 \, \text{GeV}/c^2$), the Higgs boson will prefer to decay into a pair of bottom quarks and, to a lesser extent, a pair of charm quarks or $\tau$ leptons, which have smaller masses. The hierarchy of the decay rates is given by the mass squared of these particles. The probability for these decays to occur, or the branching ratios (BR), are shown in Fig. 3.6 (left) and as can be seen, for Higgs masses below 130 $\text{GeV}/c^2$, the decays into bottom quark final states are by far dominant with a probability of the order of 80%, while the probability for decays into charm quarks and $\tau$ lepton pairs is of the order of a few percent. For a low mass Higgs boson, the kinematic of the decay into a pair of bottom quarks would be similar to that of the $Z \to b\bar{b}$. The development of a technique that could allow the detection of a $Z \to b\bar{b}$ would then be a test-bed for the search of the Higgs boson in the channel with the highest BR, but also with the highest background, in the low mass range.

The $\gamma\gamma$ and the $\gamma Z$ decay modes are very rare, a consequence of the fact that
3.2. The Higgs production at LHC and its decay modes

the electromagnetic coupling is much smaller than the strong interaction coupling. Below 130 GeV/c$^2$, the probability for these two decay modes to occur is at the few permille level. Despite its small branching ratio and the large associated background, the $H \to \gamma\gamma$ is a very important channel in the low mass region since it has a narrow width and a very clean signature.

In the intermediate mass range ($130$ GeV/c$^2 \leq m_H \leq 200$ GeV/c$^2$) the BR of the aforementioned decay modes rapidly decreases. The decay channels into WW and ZZ become kinematically open. For values of the Higgs mass above $130$ GeV/c$^2$, but below $2M_W$ ($2M_Z$) the rate for the three-body Higgs decay into a real and a virtual W/Z becomes comparable and even larger than the otherwise dominating two-body decay into a pair of bottom-antibottom quarks. This is due to the fact that the virtuality of the gauge boson is partially compensated by the stronger coupling of the Higgs boson to the W/Z bosons compared to bottom quarks.

For Higgs boson masses of the order of $180$ GeV/c$^2$ and beyond, the Higgs decays into two pairs of real WW and ZZ bosons largely dominate with branching fractions of 2:1 in favor of the WW channel. The $ZZ \to llll$ channel, (where $l = e, \mu$), is a very important channel at medium to high masses ($m_H \geq 200$ GeV/c$^2$). It presents a clean signature with four electrons or muons. At low mass, it is also promising although it requires an accurate reconstruction of low energy leptons momentum and direction.

When the Higgs mass reaches $2m_{\text{top}}$, the top pair decay opens, although the decay modes in WW and ZZ remain dominant thanks to the longitudinal components of the W,Z bosons which significantly enhance the rates [14].

Finally, one should note that the total decay width of the Higgs boson, $\Gamma$ ($\tau = h/\Gamma$) is of the order of only a few MeV for Higgs masses in the low mass region, but it considerably increases with the Higgs mass to reach the GeV range for $m_H \propto 180$ GeV/c$^2$ and becomes of the same order of $m_H$ when the Higgs mass approaches 1 TeV (see Fig. 3.7). Thus, the Higgs boson is a very narrow resonance for small masses but the resonance becomes very wide for a very heavy Higgs particle.

In Fig. 3.8 the plot shows the production cross section times the decay branching ratio of a SM Higgs as a function of his mass.
Figure 3.7 – Higgs decay width as a function of $m_H$.

Figure 3.8 – Production cross section times the decay branching ratio of a SM Higgs as a function of his mass.
Chapter 4

Vector bosons hadronic decay modes

4.1 Background in hadronically-decaying \( W,Z \) bosons

As seen before, there are several ways in which vector bosons \( Z \) and \( W \) may decay. However, for both of them the predominant decay is into two quarks.

As colored states are forbidden in nature, free quarks (as well as free gluons) do not exist. Once partons are produced they first undergo a process called parton showering, after which they recombine into hadrons, that are color singlets and so allowed to exist.

As partons are emitted in the hard scattering process (or in the vector boson decay), the strong coupling constant increases with their separation. Part of the kinetic energy of the quarks turns into potential energy which is then used to produce pairs of \( q\bar{q} \) and gluons. This increases the probability for QCD radiation, which is predominantly low-angled with respect to the originating parton. The initial stage of this process is called fragmentation process (when the two partons are at relatively close distance, so that the strong coupling constant is relatively small, i.e. at less than 1 fm) can be accounted for perturbatively. The non-perturbative latter stage of jet fragmentation, when the fragmented partons combine together to give hadrons, is called hadronization (see Fig. 4.1). Fragmentation ends when hadronization complete. Hadronization of a final state parton shower is the dynamical color neutralization process, which bleaches colored partons into color-singlet hadrons [3]. As at this point the strong coupling constant is large, a perturbative approach is not applicable and only a phenomenological description can be given.

The collimated bunch of hadrons (which will then may decay into stable particles) flying roughly in the same direction forms what is called hadronic-jet. With this intuitive definition, a jet can be thought as a cone made of particles.

The major problem in studying the \( W/Z \) decaying hadronically is the massive background of this channel. In QCD, two-jets events result when an incoming parton from one hadron scatters off an incoming parton from the other hadron to produce two high-transverse-momentum partons which are observed as jets (see Fig. 4.4). From momentum conservation, the two final-state partons are produced with equal
Figure 4.1 – Scheme of hadron production in quark-quark annihilation. The initial stage of the process that leads partons to color singlets hadrons is called fragmentation, and can be accounted for perturbatively. Hadronization can only instead be given a phenomenological description.

and opposite momenta in the subprocess centre-of-mass frame. If only two partons are produced, and the relatively small intrinsic transverse momentum of the incoming partons is neglected, then the two jets will be back-to-back in azimuth and balanced in transverse momentum in the laboratory frame [15].

Since the W jets’ $p_t$ spectrum falls exponentially much slower than the QCD di-jets’ $p_t$ spectrum (see Fig. 4.3), if a cut is made at high $p_t$, the ratio signal/background increases. The high $p_t$ boson will then decay into two neighboring jets, which would be reconstructed as a single fat jet. The filtering method allows one to recognize the jet sub-structure and resolve the fat jet on a finer angular scale. The processes that can fake a boosted W,Z are still the ones in Fig. 4.4, where one of the two partons in the final state emits a gluon. If these two partons are enough energetic they can easily make up the mass of the W or the Z.

Since the two incoming partons collide back to back, a singly produced vector boson cannot acquire any transverse momentum. Production of a boosted W or Z is accompanied by radiation of another particle (usually a jet) in the opposite direction, to balance the transverse momentum. These are next-to-leading-order diagrams for the Drell-Yan process (see Fig. 3.1 (b) and (c)).

Another possibility to produce boosted W (Z) bosons would be through W (Z) pair production.

Fig. 4.2 represents the event topology for a highly-boosted W boson decaying into two quarks. The figure shows three high $p_T$ jets: one that balances the W and two jets coming from the weak boson decay.
4.2. Jets Reconstruction

As said before, a jet can be qualitatively thought as a bunch of particles flying roughly in the same direction. A more precise and correct definition can be given describing how a jet is detected and reconstructed in an experiment.

A detector can only account for the properties (energy, momentum, and direction) of the single particles that compose the jet, but what is that we regard as a jet is more arbitrarily than objective. Different definitions (algorithms) applied to the same particles energy deposits in the calorimeter can in fact lead to different results. Algorithms for jet reconstruction may be divided into two classes: cone and sequential recombination jet-algorithms. The historical cone jet-algorithms have been almost completely abandoned, so they will not be described here.

The one that was used in the present work, the Cambridge-Aachen, belongs to the sequential recombination jet-algorithms. All these algorithms act on reconstructed hadronic energy deposits in the calorimeter detectors.

4.2.1 Sequential Recombination algorithms: Cambridge Aachen, Kt and AntiKt

A jet algorithm maps the momenta of the final state particles measured by the detector into the momenta of a certain number of jets:

\[ \text{Figure 4.2} \quad \text{– Event display for a highly-boosted W boson decaying into two quarks.} \]

The figure shows three high $p_T$ jets: one that balances the W and two jets coming from the weak boson decay.
Figure 4.3 – MC: Differential cross section with respect to the unfiltered jet $p_t$ for background (blue) and signal (red) normalized to 1. In green the ratio signal/background (normalized to 1 too). $S/B$ increases with $p_t$. 
4.2. Jets Reconstruction

\( \{p_i\} \rightarrow \text{jet algorithm} \rightarrow \{j_k\} \)  \hspace{1cm} (4.1)

\( p_i \) is the momentum of the i-th particle, or, better\(^1\), the momentum of the i-th topological cluster, while \( j_k \) is the momentum of the k-th jet, which has been built using “some” topological clusters combined together through a jet algorithm.

A good jet algorithm reconstructs jets which are close to the direction of the partons produced by the hard scattering, even if partons are not real objects, and their definition is model-dependent.

Cambridge-Aachen, \( k_t \) and anti-\( k_t \) algorithms proceed by sequentially combining pairs of particles\(^2\) according to the measure \( d_{ij} \) defined below. The procedure may be summarized as follows:

1. Calculate the distances between the particles: \( d_{ij} \)
2. Calculate the beam distances: \( d_{iB} \)
3. Combine particles with smallest distance \( d_{ij} \) or, if \( d_{iB} \) is smallest; call it a jet
4. Find again smallest distance and repeat procedure until no particles are left

\(^1\)\( p_i \) is truly the momentum of the i-th particle if the jet-reconstruction is performed at the Monte Carlo generation level.

\(^2\)As already discussed above, the particles, in which the partons hadronize, are reconstructed as topoclusters. Combining the topoclusters together through some algorithm gives a jet.
Figure 4.5 – A sample parton-level event (generated with Herwig), together with many random soft “ghosts” (particles with infinitesimally small momentum with respect to all other particles in the event), clustered with three different jets algorithms, illustrating the “active” catchment areas of the resulting hard jets. For $k_t$ and Cam/Aachen the detailed shapes are in part determined by the specific set of ghosts used, and change when the ghosts are modified[9].
The definition of the distance, which all these three algorithms rely on, is the following:

\[ d_{ij} = \min(k_{t_1}^{2p}, k_{t_2}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \]  
\[ d_{iB} = k_{t_1}^{2p} \]  

where \( \Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \) and \( k_{t_1}, y_i \) and \( \phi_i \) are respectively the transverse momentum, rapidity and azimuth of particle \( i \). The parameter \( p \) governs the relative power of the energy versus geometrical (\( \Delta_{ij} \)) scales; the parameter \( R \) is the radius of the reconstructed jet in the \( \eta - \phi \) plane (\( \frac{\Delta_{ij}^2}{R^2} \) measures the geometrical distance between topoclusters in units of the jet radius)\(^9\).

For \( p = 1 \) one recovers the inclusive \( k_t \) algorithm, while for \( p = -1 \) the inclusive \textit{anti} – \( k_t \) algorithm. The case of \( p = 0 \) is special and it corresponds to the inclusive Cambridge-Aachen algorithm. The default jet-clustering algorithm in ATLAS is \textit{anti}-kt, implemented using FastJet \cite{20}, a software package for jet finding in pp and e+e- collisions.

\textit{Anti-kt} will be here described in detail to illustrate how a recombination algorithm works.

The functionality of the \textit{anti} – \( k_t \) algorithm can be understood by considering an event with a few well-separated hard particles with transverse momenta \( k_{t_1}, k_{t_2}, \ldots \) and many soft particles.

The \( d_{i1} = \min(1/k_{t_1}^{2p}, k_{t_2}^{2p}) \Delta_{i1}^2/R^2 \) between a hard particle 1 and a soft particle \( i \) is exclusively determined by the transverse momentum of the \textit{hard} particle and the \( \Delta_{i1} \) separation, as opposed to the \( k_t \) algorithm where \( d_{i1} \) is exclusively determined by the transverse momentum of the \textit{soft} particle.

The \( d_{i1} \) between similarly separated soft particles will instead be much larger. Therefore soft particles will tend to cluster with hard ones long before they cluster among themselves.

If a hard particle has no hard neighbors within a distance \( 2R \), then it will simply accumulate all the soft particles within a circle of radius \( R \), resulting in a perfectly conical jet.

If another hard particle 2 is present such that \( R < \Delta_{12} < 2R \) then there will be two hard jets. It is not possible for both to be perfectly conical.

If \( k_{t_1} \gg k_{t_2} \) then jet 1 will be conical and jet 2 will be partly conical, since it will miss the part overlapping with jet 1.

Instead, if \( k_{t_1} = k_{t_2} \) neither jet will be conical and the overlapping part will simply be divided by a straight line equally between the two.

For a general situation, \( k_{t_1} \sim k_{t_2} \), both cones will be clipped, with the boundary \( b \) between them defined by \( \Delta R_{1b}/k_{t_1} = \Delta_{2b}/k_{t_2} \).

A third possibility occurs when \( \Delta_{12} < R \). Here particles 1 and 2 will cluster to form a single jet. If \( k_{t_1} \gg k_{t_2} \) then it will be a conical jet centered on \( k_1 \). For

\(^9\)Note that the intuitive picture of a jet being a cone of radius \( R \) is wrong.
Figure 4.6 – For a given $p_t$ of the vector boson, the more the two quarks are asymmetric in momentum the larger is the angular separation.

$k_{t1} \sim k_{t2}$ the shape will instead be more complex, being the union of cones (radius $< R$) around each hard particle plus a cone (of radius $R$) centered on the final jet. The key feature above is that the soft particles do not modify the shape of the jet, while hard particles do. i.e. the jet boundary in this algorithm is resilient with respect to soft radiation, but flexible with respect to hard radiation.

Conversely, the $k_t$ algorithm sweeps soft particles into jets first, creating irregular jet structures.

The Cambridge-Aachen, instead, is completely independent of the transverse momentum of the particles, being only based on the disposal of the topoclusters[9].

The jet $E_t$ and axis are calculated from the vector sum of all energy depositions in the cone.

### 4.2.2 Subjet Reconstruction and Filtering

In this section the filtering method used to reconstruct the jet-substructure of highly boosted hadronically decaying vector bosons will be presented.

When a highly boosted vector boson decays hadronically, it produces a single fat jet containing two quarks. The $q\bar{q}$ angular separation will vary significantly with the boson $p_t$ and the directions of the two decaying quarks (which depends on how they share the $p_t$ of the “mother” boson), roughly:

$$R_{q\bar{q}} \approx \sqrt{\frac{1}{z(1-z)} \frac{m_V}{p_t}}, \quad (p_t \gg m_V) \quad (4.4)$$

where $z, 1-z$ are the momentum fractions carried by the two quarks. It is apparent from the formula that for a given $p_t$ the more the two quarks are asymmetric in momentum the larger is the angular separation; and the larger is the $p_t$ (i.e. larger boosting), the smaller is the angular separation between the two quarks (see Fig. 4.6).

In order to maximize the resolution on the mass of the reconstructed jet, one
4.2. Jets Reconstruction

should capture $q, \bar{q}$, and any gluon they emit, while discarding as much contamination as possible from the underlying event (UE), as well as from hard gluons emitted before the Z or W.

For the jet definition, the filtering procedure uses the Cambridge-Aachen algorithm, since it is the one that works the best (explanation for this will be given later).

The initial radius chosen for the jet clustering is relatively large (usually $R = 1.2$), so that the jet obtained will for sure contain both quarks and their radiation.

Given a reconstructed hard jet with radius $R$ and mass $m_j$, the following iterative decomposition is performed (see Fig. 4.7):

1. Undo last step of clustering. Calculate the distance, $d$, between the remaining jet and the part removed. If they are not too close, that is $d > R_{\text{min}}^{\text{split}}$ ($R_{\text{min}}^{\text{split}}$ being an arbitrary parameter) then proceed with step 2. Otherwise stop and set the four-momentum of the filtered jet equal to $(0,0,0,0)$.

2. Label the two subjet $j_1, j_2$ such that $m_{j_1} > m_{j_2}$.
   Check how the mass splits between the two subjets.
   **If there was** a significant mass drop so that $\frac{m_{\text{min}}}{m_j} > \mu$, where $\mu$ is an arbitrary parameter $\in (0,1)$
   and the splitting is not too asymmetric$^4$, $y = \frac{\text{min}(p_{t,j_1}^2, p_{t,j_2}^2)}{m_{j}} \Delta R_{j_1,j_2}^2 > y_{\text{cut}}$, then stop and recombine these two jets ($j_1, j_2$).
   **If there was not** a significant mass drop or the splitting is too symmetric repeat the procedure of point 1.

3. The process of recombining jets $j_1, j_2$ is called filtering. This basically means to recluster the constituents of $j_1, j_2$ on a finer angular scale, $R_{\text{filt}} < R_{q\bar{q}}$, until only three (or less) hard objects remain. $R_{\text{filt}} = \text{min}(0.3, R_{q\bar{q}}/2)$ is found to be a good choice.
   These three objects are associated to the (sub-) jets coming from the two quarks and the one coming form a gluon. Thus, one captures the dominant $O(\alpha_s)$ radiation from the vector boson, while eliminating much of the UE contamination[8].
   Conventionally, the jet that comprises the three subjects, is called filtered jet.

To be noticed that the jet-definition used for both the initial fat jet and the filtering can be anyone. The reasons to use the C/A are mainly the following:

- the fact that C/A provides a hierarchical structure purely based on geometry for the clustering (v. fig. 4.8)
- the meaningfulness of the sequence (C/A recombines closest topoclusters first, easy to picture)
- the smaller natural sensitivity to soft background (as opposed to $k_t$)
Chapter 4. Vector bosons hadronic decay modes

Figure 4.7 – Start from a hard fat jet. Undo last step. If there is a significant mass drop and the splitting is not too asymmetric, then stop and recombine this two jets using $R_{filt}$ as the parameter for the jet definition. Recluster until three or less hard objects remain.

Hierarchical Structure

Figure 4.8 – Hierarchical structure for anti-$k_t$, $k_t$ and Cambridge-Aachen algorithms.

Apart from the considerations above, it must been add that tests performed on filtering algorithms based on $k_t$ or anti-$k_t$ provided worse results than CA.

From now on, the word “jet” will implicitly refer to Cambridge-Aachen jets. Jets are reconstructed using the FastJet package.

4Usually, signal splitting is more symmetric than background.
4.2. Jets Reconstruction

A. Davison in Arimis Meeting 04/07-08 of 19

Jet Substructure as a new Higgs Search Channel at the LH C

These slides pirated from an excellent talk by G. Salam at SUSY 08

Matteo Cacciari - LPTHE Milan - June 2011

Boosted Higgs tagger

Butterworth, Davison, Rubin, Salam, 2008

pp

ZH

.. bb

Start with the hardest jet.

m = 150 GeV

Use C/A with R=1.2 to recluster jets.

Start with the hardest jet.

m = 150 GeV

Undo last step of reclustering.

Check how the mass splits between the two subjets (m_1 = 139 GeV, m_2 = 5 GeV)

If max(m_1, m_2) > µ repeat

Figure 4.9 – Filtering, first part: undo reclustering until there is a significant and not too asymmetric mass drop.
Stop when a large mass drop is observed and recombine these two jets.

\[
m_1 = 52 \text{ GeV}, \ m_2 = 28 \text{ GeV}
\]

Take the constituents of the two jets and recluster them using \( R_{\text{filt}} = 0.3 \)

Final filtered result

**Figure 4.10** – Filtering, second part: once a significant mass drop is found, stop and recombine the constituents of the jets left using \( R_{\text{filt}} \).
Chapter 5

Experimental results on hadronically-decaying $W,Z$

“Data! Data! Data!” he cried impatiently. “I can’t make bricks without clay.”
— Sir Arthur Conan Doyle

5.1 MC samples

Since the aim is to select highly boosted vector bosons, only events with leading jet\(^1\) of $p_T \geq 300 \text{ GeV}$ have been accepted. The filtering has then been applied to the unfiltered leading jets only (we therefore speak of leading filtered jets).

The Monte Carlo (MC) simulations for the signal (vector bosons $W$ and $Z$ decaying hadronically) have been generated using HERWIG, with multiple interactions accounted for by the JIMMY package, while the ones for the background using PYTHIA.

A cut at the MC generation level had already been applied. The MC sample for the signal ($W, Z \rightarrow q\bar{q}$) has been produced keeping events with leading jet of $p_T \geq 250 \text{ GeV}$, whereas the background sample has been produced keeping events with leading jet of $p_T \geq 150 \text{ GeV}$. With this choice, it follows the $p_T$ cut can never be lower than 200 GeV.

The background sample is divided into five ranges, J4-J8, each of which corresponds to a specific range in $\hat{p}_t$, the transverse momentum at parton level in the Leading Order (LO) approximation. For each range the same number of events has been generated.

Since the $\hat{p}_t$ follows an exponential-type distribution, a large number of events would have to be produced to populate the high-$\hat{p}_t$ region if background was pro- 

\(^1\)The leading jet of an event is defined as the jet with the highest $p_T$ among the jets in that event.
duced in a single sample\(^2\). This would eventually affect the speed of the simulation process, making it slow. This explains why the background is produced into different \(p_t\) ranges.

Each JX sample shows a jet \(p_t\) spectrum with jets between the upper and the lower \(p_t\) boundaries, but without hard edges, because reconstruction, hadronisation, fragmentation, jet clustering etc... smear the energy of the jet coming from the generated parton.

Fig. 5.1 (left) shows the jet \(p_t\) distribution for the unfiltered jets belonging to the J5 sample. The bump around 320 GeV is due to di-jets coming from hard-scattering collisions (di-jet production being the process specified). Lower \(p_t\) jets come from underlying events (UE), these being associated to Initial/Final State Radiation (ISR/FSR), Multiple Parton Interactions (MPI), additional proton-proton collisions (pile-up) occuring in the same bunch-crossing, Beam-Beam Remnants\(^3\) (BBR), noise...

Fig. 5.2 (left) shows the \(p_t\) spectrum for leading unfiltered jets of J4 sample, which is the one corresponding to the lowest \(p_t\) range we have. It is here apparent the cut at 150 GeV applied on the jet \(p_t\). The scatter plot in 5.2 (right) shows on the x-axis the pt of second unfiltered leading jets and on the y-axis the pt of leading unfiltered jets. Only events with at least two jets have been selected. No entries below \(y = 150\) GeV are present, showing correctly that only events with leading jets with \(p_t \geq 150\) GeV are selected. No sub-leading jets are found below the line \(y = x\), according to the definition of leading jet.

Fig. 5.3 shows, instead, the \(p_t\) spectrum for filtered jets. Lower \(p_t\) jets have been reconstructed using “soft” constituents from initial fat-unfiltered jets and thus correspond to soft radiation (remember that the radius used in the filtering, \(R_{filt}\) is much smaller than the initial radius used in the Cambridge-Aachen definition for the jet reconstruction). For future plots a \(p_t\) cut at 300 GeV will also be applied to filtered jets, so to reject jets coming from soft radiation. The statistics gets consequently smaller passing from unfiltered to filtered jets.

The jets that make up the bump are, instead, hard filtered jets (coming from hard-scattering collisions between partons).

---

\(^2\)Remember that the fraction of accepted points is equal to the fraction of the box’s area under the curve in question.

\(^3\)Each incoming beam particle may leave behind a beam remnant, which does not take active part in the initial-state radiation or hard-scattering process.

In that case, the initial-state radiation algorithm reconstructs one shower initiator in each beam. This initiator only takes some fraction of the total beam energy, leaving behind a beam remnant which takes the rest. For a proton beam, a quark initiator (e.g. \(u\)) would leave behind a diquark (\(ud\)) beam remnant, which has a color charge. Thus, these remnants need to be put together and color connected to the rest of the event. This extra component gives a non-negligible contribution to the underlying event.
5.1. MC samples

Figure 5.1 – J5 background sample.

Right: Distribution of unfiltered CA jets with respect to their $p_t$, expressed in GeV.

Left: Distribution of leading unfiltered CA jets (by definition there is only one leading jet per event) with respect to their $p_t$, expressed in GeV.

To produce these plot no cuts on the $p_t$ (apart the ones at the MC generation level) have been applied.
Figure 5.2 – J4 background sample.  
Left: Distribution of leading unfiltered CA jets (by definition there is only one leading jet per event) with respect to their $p_t$, expressed in MeV. No entries are found below 150 GeV, showing that a $p_t$ cut at this energy is applied at the generation level.  
Right: scatter plot - leading unfiltered jets on the y-axis, second leading jets on the x-axis. Only events with at least two jets have been selected. No entries below $y = 150\, GeV$ are present, showing correctly that only events with leading jets with $p_t \geq 150\, GeV$ are selected. No sub-leading jets are found below the line $y=x$, according to the definition of leading jet.

Figure 5.3 – J4 background sample: Number of leading filtered CA jets with respect to their $p_t$, expressed in GeV. Low $p_t$ jets exceed the scale of the vertical axis.  
Jets at low $p_t$ have been reconstructed using “soft” constituents from initial fat-unfiltered jets and thus correspond to soft radiation.
5.2 MC-Data unfiltered jets distributions

Results for background and signal have been merged together, taking into account their cross sections and the number of events produced for the specific sample at the generation level. The code, run on the MC samples, gives a number of entries as a function of the jet mass. The histograms resulted for the W, the Z and the five background ranges, have been added one to other applying for each entry in the \( i \)-th bin a weight \( w_i \) (\( i = W, Z, JX \)):

\[
  w_i = \frac{\sigma_i}{N_i^{\text{gen lev}}}
\]

where \( N_i^{\text{gen lev}} \) is the number of events for the \( i \)-th bin produced at the generation level.

In this way the total cross section, \( \sigma_{\text{tot}} \), will be scaled by a factor \( N_{\text{tot}}/N_{\text{tot}}^{\text{gen lev}} \), where \( N_{\text{tot}} \) is the total number of events that passed all the selections and have finally been collected. \( N_{\text{tot}}^{\text{gen lev}} \) are instead the events originally produced. The factor \( N_{\text{tot}}/N_{\text{tot}}^{\text{gen lev}} \) can thus be regarded as a sort of efficiency.

While MC simulations predict a cross section \( \sigma \), the detector simply counts event rates (\( \dot{N} \)). The proportionality factor between a number of events over a period of time \( T \) and a cross section is the integrated luminosity over that same period of time (for more details on luminosity see Appendix B).

The MC cross section can then be converted into a number of events if multiplied by a luminosity:

\[
  N = \sigma \cdot \mathcal{L}
\]

Fig. 5.4 shows the unfiltered leading jets distribution with respect to the jet \( p_T \), \( \eta \), \( \phi \) and mass for MC and data.

Since detection is particularly difficult in the forward region and the resolution power of the clustering algorithm is here reduced, a cut at \( |\eta| = 2.8 \) has been applied.

The data analyzed in this thesis correspond to a total luminosity of \( 1.56 \text{ fb}^{-1} \). The error in the counts for a given bin follows a Poisson distribution. If \( N_i \) is the number of entries in the \( i \)-th bin, its error is \( \sigma_i = \sqrt{N_i} \).

MC and data are in disagreement. The \( p_T \) distribution in logarithmic scale indicates that data and MC do not agree in shape, with the MC being above data at low \( p_T \) and below at high \( p_T \).

Looking at the \( \phi \) distribution, it is particularly evident that MC and data are not normalized to equal area. The discrepancy between the two areas is less than 40%:

\[
  1 - \frac{\int_{\text{all range}} dN^\text{data}/dx \, dx}{\int_{\text{all range}} dN^\text{mc}/dx \, dx} \simeq 38\% \quad x = p_T, \eta, \phi, m
\]

This disagreement is not surprising, since PYTHIA MC generator provides predictions of signal and background processes only at the Leading Order in perturbative QCD. This is why systematic theoretical uncertainties associated to PYTHIA cross
sections are large, of the order of the disagreement observed between MC and data normalizations.

Another source of disagreement comes from the absence of a trigger selection. For technical reasons trigger menus could not be stored in the samples used. Any trigger requirement could impact on the jet distributions shapes and thus could potentially explain the disagreement observed\(^4\). Trigger information are currently being implemented in a new production.

### 5.3 MC reweighting

In MC and data, each event contributes to the histogram with a weight of 1. In order to make the MC model better data, a weight to MC events has been applied. This procedure goes under the name of “reweighting”. In the present analysis the weight was chosen to be based on the unfiltered jets \( p_T \) distribution.

As a first step MC and data unfiltered \( p_T \)-jet histograms have been normalized

---

\(^4\)If the trigger used is not 100\% efficient, then there will be a relative bias between data (where the trigger is applied) and MC (where it is not), leading to a higher number of MC events.
5.4. Effect of filtering on jet mass

to equal area, namely to the data integral over the whole $p_T$ range (see Fig. 5.5).

The second step is to get the following ratio for each of the plotted bins:

$$\frac{dN_{Data}}{d\eta} \div \frac{dN_{MC}}{d\eta}$$ (5.3)

Fitting the ratio with a cubic polynomial:

$$f(p_t) = a p_t + b p_t^2 + c p_t^3 + d p_t^4$$ (5.4)

gives the following parameters:

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>-3.29788 e-01</td>
<td>4.14478 e-01</td>
</tr>
<tr>
<td>$b$</td>
<td>6.39008 e-03</td>
<td>2.44705 e-03</td>
</tr>
<tr>
<td>$c$</td>
<td>-9.18873 e-06</td>
<td>4.56210 e-06</td>
</tr>
<tr>
<td>$d$</td>
<td>4.98486 e-09</td>
<td>2.69136 e-09</td>
</tr>
</tbody>
</table>

The parameters were obtained minimizing the value of the chi-square between
the fitting function (cubic polynomial) and the data for given set of parameters.
Fig. 5.6 (left) shows the fitted ratio.

The reweighting acts mostly on the shape of the curves. Nevertheless, changes
in the shape can slightly affect the normalization.

Histograms in Fig. 5.7 were normalized to data. The weight applied to each
event is equal to the value of the function in Eq. (5.4) evaluated for the $p_t$ for that
event. The weight is below 1 when MC is higher than data and above 1 viceversa.
Thus, the reweighting procedure force MC to model data better.

Even if not explicitly stated, all the plots from now on will be reweighted.

5.4 Effect of filtering on jet mass

The parameters here used for the filtering are:

$\mu = 0.3$, $y_{cut} = 0.2$, $R_{min,split} = 0.1$

Fig. 5.8 shows the $p_T$ and the mass distributions before and after the filtering at
MC level. The reconstructed background is shown in blue, the reconstructed signal
in azure and the signal at true $^5$level in green.

After the filtering, the reconstructed jet $p_T$ distribution is in better agreement
with the true level.

From the comparison between the bottom plots it is apparent that the filtering
method enables a better resolution on the jet mass, discarding as much contami-
nation as possible from underlying events. Another very important feature of the

\[ ^5 \text{By true level one means the bare generated event, without any simulated interaction with the detector.} \]
Figure 5.5 – MC and data differential number of leading unfiltered jets with respect to the jet-$p_t$. MC has been normalized to data.

Figure 5.6 – Ratio data over MC fitted with cubic polynomial.
Figure 5.7 – Leading unfiltered jet $p_t$ distribution after applying the reweighting.
filtering is that, by a proper tuning of its parameters ($\mu$, $y$, ...), it shapes the background and the signal differently, so that the former is flattened and the latter is peaked.

In the mass region of interest for W,Z, the ratio of signal over background goes from 0.3\% to 2\%. For the luminosity in our data, $\sim 1.56 fb^{-1}$ the ratio $N_s/\sqrt{N_b}$ goes from $\sim 1.28$ to $\sim 5.61$ in the same mass range. For the luminosity collected in 2001, $\sim 5 fb^{-1}$ the ratio $N_s/\sqrt{N_b}$ goes from $\sim 2.3$ to $\sim 10$ in the same mass range.
5.5 Filtering: MC-Data comparison

Fig. 5.9 shows the comparison between MC and Data for filtered jets. The agreement in $p_T$, $\eta$, $\phi$ is not as good as for the unfiltered jets. An important disagreement is observed especially at the low mass range.

As a different tuning of the filtering parameters (especially the $R_{\text{min}}$) can deeply modify the background distribution, MC-Data distribution needs to be further investigated using different filtering parameters.

Figure 5.9 – Filtered jet $p_T$ distribution for MC and data.
Chapter 6

Multi Variate Analysis

“If your experiment needs statistics, you ought to have done a better experiment.”
— Ernest Rutherford

In order to increase the signal over background (S/B) ratio, the filtering method has been combined with the use of discriminating variables. Eight variables have been used in the present analysis. Each variable $x_i$ has an assigned probability distribution function (PDF) $f(x_i)$, which is different for background (QCD di-jets) and signal (W,Z di-jets) hypotheses. By convention, the PDF for $x_i$ in the hypothesis $H_i$ is defined as $f(x_i|H_i)$.

Each jet (event) has an associated value for each of the discriminating variables. Therefore, each event is represented by a point in an eight-dimensional space $x$ formed by the variables used.

6.1 Test statistic, hypotheses, efficiency, purity

In order to make a statement about how well the observed event stands in agreement with a given hypothesis (the observed event is signal-like or background-like), one constructs a function of the measured variables, $x$, called a test statistic $t(x)$. The dimension of the test statistic $t$ can be lower or equal to the one of $x$, and each hypothesis $H_i$ has an associated PDF, $g(t|H_i)$. The null hypothesis $H_0$ is the one being tested. In this analysis, $H_0$ is the signal hypothesis, while $H_1$ is the background one.

A test statistic can be constructed in different ways, both linear or non-linear, but the ultimate scope is always to increase the distinction between signal and background. This is done by rejecting most of the background while retaining most of the signal.

The name Multi-Variate Analysis (MVA) is used to describe an analysis which makes use of more than one variable to construct $t$. A MVA employs different methods, called classifiers, to achieve this. There is no immediate way to know a
which method will give the best results for a particular analysis. For this reason, three classifiers have been tested.

Suppose for the moment to choose a scalar function \( t(x) \), which has PDFs \( g(t|H_0) \) and \( g(t|H_1) \). A cut on \( t, \tau_{cut} \), is chosen to achieve the maximum signal purity \([13]\). The cut will delimitate an acceptance region, in which the signal \((H_0)\) is accepted and a critical region in which it is rejected. Since a cut will not be able to reject all the background and accept all the signal, two variables are defined to quantify these errors, the Type I and Type II errors, respectively labelled \( \alpha \) and \( \beta \).

\[
\alpha = \int_{t_{cut}}^{\infty} g(t|H_1) \, dt \quad \beta = \int_{-\infty}^{t_{cut}} g(t|H_0) \, dt
\]

Both of them are preferred to be small. These quantities also define background and signal efficiencies. Background efficiency is a measure of how much background is accepted, which is equivalent to \( \alpha \). Signal efficiency is a measure of how much signal is accepted, equivalent to \( 1 - \beta \), also called power of the test statistic. It is important not to confuse background efficiency with background rejection efficiency, which is \( 1 - \alpha \), also called selection efficiency.

A better discrimination will be given by PDFs which mean values for signal and background hypotheses fall as much apart as possible, having the smallest possible overlap. This will allow to apply a cut with maximum S/B ratio and purity. The definition of these quantities depend on Bayesian statistics. Purity is defined as

\[
P = \frac{\epsilon_{sg}\pi_{sg}}{\epsilon_{sg}\pi_{sg} + \epsilon_{bg}\pi_{bg}}
\]

where \( \epsilon_{sg/bg} \) are the efficiencies for signal and background and \( \pi_{sg/bg} \) are the prior probabilities for signal and background. The prior probabilities have been considered as the MC truth level in this analysis while the efficiencies are at the MC reconstructed level. The previous equation can be rearranged and expressed in terms of \( \alpha \) and \( \beta \) as

\[
P = \frac{1 - \beta}{1 - \alpha - \beta}
\]

Note that the purity takes into account both Type I and II errors. This means that purity gives an indication of how much contamination there is in the accepted region and how much signal has been lost. Signal over background is the ratio of the reconstructed PDFs and then can be expressed as

\[
S/B = \frac{1 - \beta}{1 - \alpha}
\]

which also depends on \( \alpha \) and \( \beta \).
6.2. Shape variables

Several variables have been investigated to discriminate signal from background. In this section each of them will be defined.

6.2.1 Sphericity

In order to determine how jet-like an event is, one may calculate a quantity which is called sphericity:

\[ S = \frac{\sum_i p_i^2}{\sum_i |p|^2} \]  

(6.5)

The sum runs over the number of particles and the subscript \( t \) refers to the transverse component of the particle’s momentum with respect to the axis which minimizes the sum over all the particles transverse momenta (also called sphericity axis).

\( S \) approaches 0 for events with bounded transverse momenta and approaches 1 for events with large multiplicity and isotropic phase-space particle distributions [25].

Historically, sphericity was first introduced by Bjorken and Brodsky [6] to detect jet formation in \( e^+e^- \) collisions. Events with small sphericity corresponded to two-jet like events (with the two jets being back-to-back as the collision was in the \( e^+e^- \) mass frame) and those with larger sphericity to more spherical events which contained more than three jets.

Figure 6.1 – Probability densities for the test statistic \( t \) under the assumption of the hypotheses \( H_0 \) and \( H_1 \). \( H_0 \) is rejected if \( t \) is observed in the critical region, here shown as \( t > t_{\text{cut}} \).
An equivalent and more practical definition for this variable makes use of the following sphericity tensor:

$$S_{\alpha\beta} = \frac{\sum_i p_i^{\alpha} p_i^{\beta}}{\sum_i |\vec{p}_i|^2}$$

(6.6)

where $\alpha, \beta = 1, 2, 3$ correspond to x, y, z components respectively.

By diagonalizing $S_{\alpha\beta}$ one may find three eigenvalues. Let the eigenvalues be named $\lambda_1, \lambda_2, \lambda_3$, so that $\lambda_1 \geq \lambda_2 \geq \lambda_3$ (note also that, by the above definition, $\text{Tr}(S_{\alpha\beta}) = 1$, i.e. $\lambda_1 + \lambda_2 + \lambda_3 = 1$). The axis which minimizes the sum over all the particles transverse momenta, also called sphericity axis of the event, is defined to be the direction of the eigenvector with the largest eigenvalue, namely $\lambda_1$. It follows directly that the sphericity is defined as:

$$S = \frac{3}{2} (\lambda_2 + \lambda_3)$$

(6.7)

Clearly $S$ is not invariant under Lorentz transformations, therefore $S$ changes if calculated in different frames of reference.

Suppose to calculate $S$ in the lab frame for a reconstructed boosted fat jet coming from a vector boson decaying hadronically. The sum over $i$ in Eq. (6.5) would then run over the topological clusters that make up the jet. The sphericity axis for this event would lay between the two subjets inside the fat jet. This would also hold if the fat jet comes from ordinary QCD background.

If, instead, sphericity is calculated in the center-of-mass (CM) frame of the boosted jet (i.e. where the jet is at rest, so that its four momentum is equal to $p_{\mu}^{\text{rest}} = (m_{\text{jet}}, 0, 0, 0)$), one would get different results depending if the fat jet comes from W,Z vector bosons or QCD background.

In the rest frame of a hadronically decaying W boson, the constituents particles look like a back-to-back di-jet event (the rest frame of the boson being identical to the CM frame of its decay products). On the other hand, the particle distribution of a QCD jet in its rest frame does not correspond to any physical state and is therefore more likely to be random (see Fig. 6.2).

Thus, sphericity calculated in the boosted fat jet CM frame tends to 0 if the two subjets that make up the jet come from an actual particle and tends to 1 if the subjets are originated from QCD radiation. In the form case the energy clusters will be back-to-back in the fat jet CM, while in the latter they will favor an isotropic distribution.

Sphericity may be computed using the constituents (energy topoclusters) of unfiltered and filtered jets. Since the latter give already a rejection of UE, filtered jets will be preferred. From now on, even if not specified, sphericity will be calculated in the jet CM frame.

### 6.2.2 Aplanarity

A complementary shape variable to sphericity, that is also defined from the sphericity tensor of Eq. (6.6), is aplanarity.
6.2. Shape variables

Aplanarity measures the amount of momentum “out of” the event plane, this being defined by the sphericity axis and the direction of the second largest eigenvalue of $S^{a\beta}$. Aplanarity is then simply:

$$A = \frac{3}{2} \lambda_3$$

(6.8)

and is constrained to the range $0 \leq A \leq \frac{1}{2}$. If calculated in the boosted fat jet CM frame, $A$ is expected to be 0 for planar events and $\frac{1}{2}$ for spherical events.

Suppose to have a filtered jet coming from a W. By construction the number of subjets are at most three, two quarks and a gluon. The event in the W CM frame will then be planar, thus giving $A=0$. In the boosted fat jet CM frame a QCD event should instead give $A = \frac{1}{2}$, hence showing not to be planar.

6.2.3 Thrust Major and Thrust Minor

Another way to determine how jet-like an event is by calculating thrust. Thrust is defined as:

$$T = \frac{\sum_i |\vec{p}_L^i|}{\sum_i |\vec{p}_i|}$$

(6.9)

As above for sphericity, the sum runs over the number of particles. The subscript $L$ here refers to the longitudinal component of the particle’s momentum with respect to the axis which maximizes the sum over all the particles longitudinal momenta (also called thrust axis). Another way to write the above equation is:

$$T = \max_{\vec{n}_T} \frac{\sum_i |\vec{p}_i \cdot \vec{n}_T|}{\sum_i |\vec{p}_i|}$$

(6.10)

For the same reasons explained above for sphericity, thrust will calculated in the boosted fat jet CM frame. $T$ approaches 1 if the di-jet event is pencil-like, that is if all particles, namely the topoclusters that make up the jet, are aligned parallel or antiparallel to a the thrust axis $\vec{n}_T$. $T$ will be equal to $\frac{1}{2}$ for an isotropic particles distribution.
Thrust major and thrust minor characterize the jet energy flow in the plane orthogonal to the thrust axis \( \vec{n}_T \).

Thrust major is defined as the projection of all particle momenta on the direction \( \vec{n} \), which is orthogonal to \( \vec{n}_T \) (i.e. \( \vec{n}_T \cdot \vec{n} = 0 \)) and along which the momentum flow is maximal:

\[
T_{maj} = \max_{\vec{n} \cdot \vec{n}_T = 0} \frac{\sum_i |\vec{p}_i \cdot \vec{n}|}{\sum_i |\vec{p}_i|} \tag{6.11}
\]

Thrust minor sums up the components of \( \vec{p}_i \) which are orthogonal to the plane defined by \( \vec{n} \) and \( \vec{n}_T \):

\[
T_{min} = \frac{\sum_i |\vec{p}_i \cdot (\vec{n} \times \vec{n}_T)|}{\sum_i |\vec{p}_i|} \tag{6.12}
\]

Thrust major and minor close to 0 correspond to a highly directional distribution on the energy clusters. If they are instead close to 0.5, they correspond to an isotropic distribution.

### 6.2.4 Angularity

Angularities are a class of jet shapes depending on a real parameter \( a \). They were originally introduced for two-jets events in \( e^+e^- \) annihilation with center-of-mass energy \( \sqrt{s} \), generating a final state \( N \), which was taken to consist of massless particles\[4, 5\]. The angularity, with weight \( a \), of the state \( N \), was then originally defined as:

\[
\tau_a(N) = \frac{1}{\sqrt{s}} \sum_{i \in N} p_{i \perp} e^{-|\eta_i|(1-a)} \tag{6.13}
\]

\[
= \frac{1}{\sqrt{s}} \sum_{i \in N} \omega_i (\sin \theta_i)^a (a - |\cos \theta_i|)^{1-a}
\]

with \( p_{i \perp} \) the transverse momentum of the particle \( i \) relative to the thrust axis of the event, \( \eta_i \) the corresponding pseudorapidity, \( \eta_i = \ln(\cot(\theta_i/2)) \), \( \theta_i \) the angle with respect to the thrust axis, and \( \omega_i \) the energy of particle \( i \). The two definitions in the above equation are equivalent for massless particles.

A natural generalization of these jet shapes to single-cone jets of large mass \( m_J \) is the following\[2\]:

\[
\tilde{\tau}_a(m_J, R, p_T) = \frac{1}{m_J} \sum_{i \in \text{jet}} \omega_i (\sin \theta_i)^a (a - |\cos \theta_i|)^{1-a}
\]

\[
\approx \frac{2a-1}{m_J} \sum_{i \in \text{jet}} \omega_i \theta_i^{2-a} \tag{6.14}
\]
with \(m_J\) the mass of the jet, \(\omega_i\) the energy of the \(i\)-th jet energy topocluster, \(\theta_i\) the angle between the \(i\)-th topocluster and the jet axis (as defined by the used jet reconstruction algorithm).

Limiting the parameter \(a < 2\) ensures IR safety, because the contribution of any constituent \(i\) to the jet shape behaves as \(\theta_i^{2-a}\) in the collinear limit (\(\theta_i \to 0\)). This can be directly seen from the expression on the right hand side of the equation, which is the Taylor expansion to the second order in \(\theta_i\), valid in the limit of small angle, \(\theta_i \to 0\).

As the parameter \(a\) decreases, the angularities give more weight to particles at the edge of the cone compared to those near the center (in the present analysis \(a = -2\)), thus emphasizing energy near the edge of the jet.

For \(1 < a < 2\), contributions to angularity from energetic particles near the jet axis are generically larger than contributions from soft, wide-angle radiation, or equal for \(a = 1\)[4].

Since angularity basically measures the energy distribution inside the jet, it is particularly sensitive to the degree of symmetry in the energy deposition.

For the reasons explained below, angularity can distinguish jets originating from QCD production of light quarks and gluons from boosted heavy particle decay.

For QCD background high mass jets (like the ones that fake a \(W\) or a \(Z\) mass), the distribution of the leading parton and the emitted gluon is expected to be peaked around a symmetric \(p_T\) configuration, where both partons are at the same angular distance from the jet axis. This is also true for jets originated from boosted hadronically decaying \(W, Z\) bosons. Thus, the angularity distribution of the signal and background are similar in shapes. However, the comparison is only qualitative, as the QCD distribution has a broader tail towards larger angularity values. In fact in signal events, energy is deposited in lobes, while for QCD events it is expected to be more evenly distributed from the center to the edge of the jet.

### 6.2.5 Width

Jet width is defined in the lab frame as the \(p_T\) weighted sum of the jet constituents falling into the ring \(dR\) in pseudorapidity \(\eta\) and azimuthal angle \(\phi\) centered with respect to the jet axis. The mathematical definition is:

\[
w = \sum_{i \in \text{jet}} \frac{dR^i p_T^i}{p_T^i} \quad (6.15)
\]

where \(dR^i = \sqrt{d\phi^i + d\eta^i}\) defines a ring positioned at the jet center, with \(d\phi_i\) and \(d\eta_i\) being the distances in the azimuthal angle and pseudorapidity of a jet constituent \(i\) with a transverse momentum \(p_T^i\) from the jet center. The sum runs over all particles inside the jet.

Large values for the width means that energy is distributed away from the center of the jet, viceversa for small values.

In a boosted massive particle decay energy is distributed in symmetric lobes away from the jet axis, thus leading to a larger width; in QCD background events,
energy is instead expected to be more evenly spread around the center of the jet, resulting in a smaller width.

6.2.6 Subjettiness

N-subjettiness, \( \tau_N \), is a variable especially defined in the lab frame to identify boosted hadronically-decaying objects and rejecting the background of QCD jets with large invariant mass.

First, one reconstructs a candidate W jet using some jet algorithm. Then, one identifies \( N \) candidate subjets. With these candidate subjets in hand, \( \tau_N \) is calculated via:

\[
\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min \{ \Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k} \}
\]

(6.16)

Here, \( k \) runs over the constituent particles in a given jet, \( p_{T,k} \) are their transverse momenta and \( \Delta R_{J,k} = \sqrt{\Delta \eta^2 + \Delta \phi^2} \) is the distance in the rapidity-azimuth plane between a candidate subjet \( J \) and a constituent particle \( k \). The normalization factor \( d_0 \) is taken as:

\[
d_0 = \sum_k p_{T,k} R_0
\]

(6.17)

where \( R_0 \) is the characteristic jet radius used in the original jet clustering algorithm.

It is straightforward to see that \( \tau_N \) quantifies how N-subjettiness a particular jet is, or in other words, to what degree it can be regarded as a jet composed of \( N \) subjets.

Jets with \( \tau_N \approx 0 \) have all their radiation aligned with the candidate subjet directions and therefore have \( N \) (or fewer) subjets. Jets with \( \tau_N \gg 0 \) have a large fraction of their energy distributed away from the candidate subject directions and therefore have at least \( N+1 \) subjets. The ratio \( \tau_2/\tau_1 \) is used as a discriminating variable to distinguish W,Z jets from QCD jets. As a W,Z hadronically-decaying boson should originate two hard nearby jets, \( \tau_2/\tau_1 \) is expected to have smaller values than QCD jets (for signal events \( \tau_1 \gg 0 \), hence indicating that there is more than one subjet, and \( \tau_2 \approx 0 \), showing that there are at most two subjets).

6.2.7 Eccentricity

Jet eccentricity is a commonly used jet shape variable in the lab frame and it is defined as follows:

\[
\epsilon = 1 - \frac{v_{\text{max}}}{v_{\text{min}}}
\]

(6.18)

where \( v_{\text{max}} \) (\( v_{\text{min}} \)) is the maximal (minimal) values of variances of jet constituents along the jet axis.
6.3. MVA Classifiers

It is a geometrical measure of the jet elongation of the jet energy deposit in the calorimeter, ranging from 0 (for perfectly circularly symmetric jet shapes) to 1 (for infinitely elongated jet shapes).

Since fat jets coming from hadronically decaying vector bosons are made of two hard, well separated, symmetric subjets, jet eccentricity is expected to be peaked towards high values (~ 1). The background is instead expected to be more evenly distributed in the whole range.

6.3 MVA Classifiers

The performance of the methods used in a multi-variate analysis strongly depends on the variables used. It is difficult to predict the outcome of each method, but they all have some general features which make a certain method more feasible in a certain situation. The performance of most of the methods used in this analysis will depend on correlation between the variables. The speed at which a classifier will produce a cut will depend on its training, response speed and often on the number of variables used (the so called curse of dimensionality). Problems such as overtraining or the use of weak variables (variables without good discriminating power) may affect the robustness of the method [16]. Each classifier responds differently to these problems. Each of the ones used in the present work will be discussed in the following paragraphs.

**Fisher**  The Fisher discriminant is a linear combination of the variables:

\[ y_{Fi}(x) = F_0 + \sum_{i=1}^{n} a_i x_i = F_0 + a^T x \]  \hspace{1cm} (6.19)

where \( a^T \) is the transposed vector of coefficients, \( x \) is the vector of variables and \( F_0 \) an offset that centers the sample mean \( \bar{y}_{Fi} \) of all \( N_S + N_B \) events at zero [19]. In the multidimensional space \( x \), the Fisher discriminant can be represented as a hyper-plane of the same dimension of \( x \). The coefficient vector will be computed to obtain the hyper-plane giving the best discrimination boundary between signal and background events.

The coefficients \( a^T \) are found by maximizing the following quantity:

\[ J(a) = \frac{(\tau_{sg} - \tau_{bg})^2}{\Sigma_{sg}^2 + \Sigma_{bg}^2} \]  \hspace{1cm} (6.20)

where \( \tau_{sg/bg} \) are the means for signal and background PDFs respectively and \( \Sigma_{sg/bg}^2 \) are the variances for signal and background PDFs respectively. In order to maximize \( J(a) \) signal and background PDFs need to be as most separated as possible, with the lowest variance. The latter giving the smallest overlap between the two curves.
Note that this expression is not considering the information about the shape of the curves. Furthermore, if the mean of one curve falls on the mean of the other, that will give \( J(a) = 0 \) with zero discrimination between signal and background.

The Fisher method is a lot affected by linear correlations, while it does not take into account non-linear correlations. Training for the Fisher method is relatively fast with respect to other classifiers and the risk of overtraining is generally little[16].

**Likelihood ratio** The likelihood ratio is a non-linear classifier. This test statistic is defined as the ratio of the joint signal PDF for all the variables over the sum of the signal and the background joint PDFs. The likelihood ratio \( y_L(i) \) for the event \( i \) is then defined as:

\[
y_L(i) = \frac{L_S(i)}{L_S(i) + L_B(i)}
\]  

(6.21)

Since the construction of the joint PDF costs an important computational effort, the likelihood ratio is usually constructed with the assumption that all the variables are independent, so that the joint PDF is simply the product of the single PDFs for each of the \( N \) variables:

\[
L_{S(B)} = \prod_{j=1}^{N} f_{S(B),j}(x_j)
\]

(6.22)

The PDFs are normalized:

\[
\int_{-\infty}^{+\infty} f_{S(B),j}(x_j) \, dx_j = 1, \quad \forall k.
\]

(6.23)

For the event \( i \), Eq. (6.22) would simply become \( L_{S(B)}(i) = \prod_{j=1}^{N} f_{S(B),j}(x_j(i)) \), with each of the variables taking the corresponding value in the event \( i \).

Independency among variables implies uncorrelation at all order. In most cases, linearly uncorrelated variables are considered a good approximation for MVA, since higher order correlations do not greatly affect the output of the classifier.

The chance to overtrain the sample using the likelihood ratio is higher than using the Fisher discriminant. This classifier is also more affected by the use of weak variables.

**Boosted Decision Tree** A decision tree is a binary tree structured classifier. Accept/reject operations will be performed in chain on a single variable at a time, until a stop criterion is fulfilled. The event space will then be split into many regions which are eventually classified as signal or background. The major drawback of decision trees is that they can lead to very different results with a small change in the training sample. This problem is solved using *boosting* algorithms. The same tree is trained several times using a successively boosted (reweighted) training event sample. The final classifier is given by a weighted average of the individual decision trees [16].
6.4 Discrimination power of the variables

Boosted Decision Tree (BDT) is a method that strongly depends on higher order correlations between variables. Training can be particularly difficult for BDT since it is slow due to the high computational power required to perform the binary acceptance/rejection operations on each event. There is high risk of overtraining for this method, compared to the previous ones, because the classifier can be very sensitive to changes in the training sample.

6.4 Discrimination power of the variables

In order to visualize the validity of each variable described in Section 6.2, a series of plots has been produced for each of them. All of these plots, as mentioned before, have a cut at 300 GeV in the jet $p_T$. A further cut on the mass has been applied at this point, selecting jets in the mass range between 70 GeV and 100 GeV. This was moved from the knowledge of the mass of the signal, being centered around 80-90 GeV. The full mass range is then not needed to study the topology of jets, as jet discrimination is only wanted around the mass of the W and the Z.

The plots are divided into two sets. The first set uses filtered jets to calculate the variables, while the second makes use of unfiltered jets.

For each set four plots (which will be described below) have been produced. The variables are in order:

1. Sphericity (CoM)
2. Aplanarity (CoM)
3. Thrust Major (CoM)
4. Thrust Minor (CoM)
5. Angularity (lab frame)
6. Width (lab frame)
7. Subjetiness $\tau_2/\tau_1$ (lab frame)
8. Eccentricity (lab frame)

Signal over background ratio The plot of signal on background (top left) is the most relevant plot to understand the discriminating power of a variable. Since the number of events for signal is expected to be significantly lower than the number of background events, the plot is normalized to unity. This will allow to give a better judgement of the discriminating power of the two variables, although any information about purity and efficiencies will be lost. This ought not to be considered as a problem, because the plots are figurative and analysis will be further progressed using TMVA to combine the most discriminating variables to compute the highest purity.
**W and Z signals**  A plot of signal only (bottom left), distinguishing between signal from W and Z bosons, normalized to one can be useful to compare the shape of the variable for different particle’s mass. No difference is expected to be found between W and Z for variables calculated in the boson CM frame. On the other hand, variables calculated in the lab frame are expected to show differences if they are correlated with the particle mass.

**MC and real data**  A comparison of Monte Carlo simulated signal plus background and real data (top right) is essential to determine if a variable can be used or not. There should be agreement between the two curves before using a variable in a MVA, because the cut is applied in the light of the Monte Carlo signal and background spectra. A difference in real data and simulation would mean that a cut that ameliorates background efficiency in simulations could instead not provide an improvement on real data. This regard applies only for the events which have not been rejected by a linear cut on the variable. This means that the agreement for MC and real data is essentially needed for the range in which the variable is kept - where there is an excess of signal. Hence, this plot can be interpreted in conjunction with the plot of Signal over Background. In general, though, a good agreement for the entire range would mirror a better reliability of the variable used in the test statistic.

**Ratio between data and MC Kolmogorov Probability**  In order to “quantify” the comparison between MC and real data, the ratio of real data over Monte Carlo has been plotted (bottom right top). This may be enough to give a feeling of how good the agreement is, but it does not provide a numerical method to solve this problem. The importance of quantifying the agreement relies on the fact that a MVA does not consider Data/MC agreement and cannot select variables on this grounds.

In order to achieve a quantitative solution, the Kolmogorov-Smirnov test has been used to assess the compatibility between MC and data distributions. The test is defined as:

\[
D_n = \max_{1 \leq n \leq N} \left( F(Y_n) - \frac{n - 1}{N}, \frac{n}{N} - F(Y_n) \right)
\] 

(6.24)

where \( F(Y) \) is the cumulative distribution for the \( i^{th} \) data point \( Y_i \), with a total of \( N \) data points. Small values of \( D_i \) are expected for a sample which resembles the cumulative distribution, that is if there is a match between data and MC [28]. The hypothesis that the MC is compatible with the data is rejected if the Kolmogorov-Smirnov test is below a certain threshold \( K_\alpha \):

\[
\sqrt{\frac{n_a n_b}{n_a + n_b}} D_{n_a, n_b} > K_\alpha
\]

(6.25)

where \( K_\alpha \) is called critical value of the Kolmogorov distribution [17]. The confidence level (CL) for the Kolmogorov probability has been calculated for each vari-
able. The CL is defined between zero and one and gives a measure of how much compatible the two distributions are. They are more likely to be compatible for bigger values of the Kolmogorov probability (close to 1), whereas small values (close to 0) show little compatibility.

The criteria through which a variable was used in TMVA did not eventually relay on MC-Data agreement, as such an agreement was not achieved in the filtered jet $p_T$, $\eta$, $\phi$ and mass distributions (see Section 5.5). A first selection on the variables was done starting from their discriminating power, that is looking at the plot Signal on Background. Each caption tells if the discriminating power is good at first sight. The ‘quality’ of the discriminating power of the variables is summarized in Table 6.1.

![Figure 6.3](image)

**Figure 6.3** – Plots of Sphericity in the center of mass for filtered constituents. Top left histogram is *signal over background ratio*. Top right histogram is *Monte Carlo and Data*. Bottom left histogram is *W and Z signal*. Bottom right plot is *Data over MC ratio*, with Kolmogorov Probability value shown on top. The following plots for variables will follow the same structure. This variable might be used in MVA, as it shows a good discriminating power between background and signal, with the latter being peaked in 0 (remember that in the CM frame of the W,Z sphericity is expected to be 0, thus indicating a back-to-back event).
Figure 6.4 – Plots of Aplanarity in the center of mass for filtered constituents. Signal is expected to peak in 0, while background is expected to show higher values. Since this is what is observed, aplanarity might be used in MVA. (For further considerations, concerning its correlation with other variables, it will not eventually be used.)
Figure 6.5 – Plots of Thrust-Major in the center of mass for filtered constituents. Signal is expected to peak in 1, while background is expected to show lower values. Since this is what is observed, thrust major might be used in MVA.
Figure 6.6 – Plots of Thrust-Minor in the center of mass for filtered constituents. The discrimination is good at first sight. This variable is thus considered.

Figure 6.7 – Plots of Angularity for filtered constituents. As expected, the angularity distribution of the signal and background are similar in shapes, but with a broader tail for QCD background towards larger angularity values. Angularity will therefore be considered.
6.4. Discrimination power of the variables

Figure 6.8 – Plots of Width for filtered constituents. This variable does not show the trend it was expected, but the opposite, also, it is strongly correlated to mass. This will be analyzed later, but hints are given by the disagreement of W and Z in the bottom-left plot. Eventually, width will not be considered.

Figure 6.9 – Plots of Subjetiness for filtered constituents. This variable will be considered in MVA.
Figure 6.10 – Plots of Eccentricity for filtered constituents. This variable was not regarded as sufficiently discriminating.

Figure 6.11 – Plots of Sphericity in the center of mass for unfiltered constituents. This variable was not regarded as sufficiently discriminating.
6.4. Discrimination power of the variables

Figure 6.12 – Plots of Aplanarity in the center of mass for unfiltered constituents. Data/MC agreement for this variable was not sufficiently good.

Figure 6.13 – Plots of Thrust-Major in the center of mass for unfiltered constituents. This variable was not regarded as sufficiently discriminating.
Figure 6.14 – Plots of Thrust-Minor in the center of mass for unfiltered constituents. This variable was not regarded as sufficiently discriminating.

Figure 6.15 – Plots of Angularity for unfiltered constituents. This variable was not regarded as sufficiently discriminating.
6.4. Discrimination power of the variables

Figure 6.16 – Plots of Width for unfiltered constituents. This variable was excluded for the same reasons explained in the caption of Fig. 6.8.

Figure 6.17 – Plots of Subjetiness for unfiltered constituents. This variable was not regarded as sufficiently discriminating.
Variables | Will it be considered?
--- | ---
Sphericity (CoM) | yes
Aplanarity (CoM) | yes
Thrust Major (CoM) | yes
Thrust Minor (CoM) | yes
Width (lab frame) | no
Subjetiness $\tau_2/\tau_1$ (lab frame) | yes
Eccentricity (lab frame) | yes

Table 6.1

**Figure 6.18** – Plots of Eccentricity for unfiltered constituents. This variable was not regarded as sufficiently discriminating.

### 6.4.1 Correlation between variables

A study of the correlation between all the variables has been performed. Correlation varies from -100% to 100%, for which 0% means no correlation and 100% (-100%) means full correlation (anti-correlation). A correlation matrix for both signal and background (Fig. 6.19 and 6.20 respectively) has been produced using TMVA with the previously selected variables. The criterion used to select uncorrelated variables was to accept any variable which had a correlation smaller than 50% with all the other variables. With correlation, it is meant also negative correlation (anti-correlation), for which the criterion to accept a variable would be to have an
6.4. Discrimination power of the variables

anticorrelation higher than -50%. Using this method, one could safely remove linearly correlated variables from the MVA. Higher order correlations have not been explored.

The pair of variables which resulted to be more strongly (anti)correlated are Thrust Major and Thrust Minor. This is why only the most discriminant can be kept in the MVA.

Below will be listed all the variables for which only one must be chosen before considering each group of them, in descending order of correlation. Where not specified, both unfiltered and filtered variables are intended.

- (U)Width - (F)Width
- Thrust Major - Thrust Minor
- Sphericity - (F)Subjetiness - Aplanarity - (F)Thrust Minor
- (U)Subjetiness - (U)Sphericity - (U)Angularity
- (U)Eccentricity - (F)Eccentricity

The effect of correlation with mass is also considered. From the plots it is clear that the correlation of the variables with the filtered mass is low. A correlation of a variable with mass can provide systematic errors, such as accumulation of events (background mostly) in a region where a cut is applied. This would possibly create a false peak in the mass region selected, which is unwanted [13]. It can be noted that Unfiltered Subjetiness, Sphericity and Angularity are related to the unfiltered mass. This does not affect the analysis, since every operation is performed on the filtered mass only.

6.4.2 Results

The variables have been selected according to the criteria of large discrimination power and low correlation, as discussed above. Five variables have been finally selected and are listed below. All of them have been taken in the mass region between 70 GeV and 100 GeV.

- Angularity (Filtered)
- Thrust Major CoM (Filtered)
- Sphericity CoM (Filtered)
- Eccentricity (Filtered)
- Subjetiness (tau2/tau1) (Unfiltered)
Figure 6.19 - Correlation plot for variables and mass, for signal. The (U) label denotes variables created using unfiltered constituents. Variables with the label (F) are taken in the rest frame of the object, as described in the variables definition.
6.4. Discrimination power of the variables

Figure 6.20 – Correlation plot for variables and mass, for background. The (U) label denotes variables created using unfiltered constituents, while the (F) label denotes variables created using filtered constituents. It has been noted that correlation was stronger on average for variables in background than of signal.
The correlation plot for the selected variables has been calculated. See Figure 6.22 and 6.21.

**Figure 6.21** – Correlation matrix of background for the selected variables

**Figure 6.22** – Correlation matrix of signal for the selected variables

For means of comparison, the three classifiers described above (Fisher, BDT and Likelihood) have been used to see which one gave the best signal over background efficiency. Four plots have been created for each classifier method, as described below.
6.4. Discrimination power of the variables

**Efficiency**  Signal efficiency is a measure of how much signal has passed the MVA discrimination. While high signal efficiency is wanted, in order to preserve most of the signal after the cuts, sometimes a smaller signal efficiency could provide a higher background rejection. These effects will be later studied in the last plots. Signal and background efficiency have been plotted over the potential cut applied for each method. The background efficiency tells the amount of background accepted. The best cut, would then be where the difference between the curves is highest. The selection has been initially based on keeping constant signal efficiency to 90%. Later, different cuts have been applied on the method which resulted to provide the best S/B.

**Mass distribution**  The mass distribution for filtered jets after the cut was applied has also been plotted at MC level. Both signal and background have been plotted separately in order to visualize the effect of the cut after the MVA. A 5 GeV binning has been used. $S/B$ and $S/\sqrt{B}$ are reported, the latter for an integrated luminosity of $1.56 fb^{-1}$ (the one corresponding to the data collected). Background and Signal events have been generated as cross sections. In order to obtain a plot in number of events, they have been multiplied by the integrated luminosity. Due to the low number of generated events at MC level, a fit of the background has been performed. On top of it has been added the signal. For each method, S/B has been calculated for the signal on top of fitted background. This is expected to be different from the S/B calculated for background showed in the bottom left plot, because the fit eliminates statistical fluctuations. This value, compared with the S/B calculated in the bottom left plot, can be as expectation on the uncertainty of the S/B ratio.

**S/B**  The plot of S/B shows how the ratio of signal over background changes with the applied cut on the method. This plot is useful to spot the best cut to obtain the best selection efficiency. The classifier with the best S/B shape will reach higher values of S/B. This is ideal, since S/B takes in account type I and II errors, minimizing both. A balance on this must be still found, though, in order to keep enough events to have statistical significance. For the most discriminant classifier, different efficiency cuts will be performed to find the maximum S/B.

**Background Rejection - Signal Efficiency**  The best cut would have a high background rejection efficiency with a high signal efficiency (both close to 1). In the ideal case all background will be rejected and all signal kept. The closer the curve is to the top-right of the plot, the better the cut is.

Plots of the three classifiers have been produced as described. See Figure 6.23, 6.24, 6.25.
Figure 6.23 – Plots of Fisher classifier. The cut on the Fisher discriminant is required to give a signal efficiency of 90%. The mass has been plotted in a range between 30 GeV and 200 GeV, where the fit was performed. The top right plot shows in orange the MC signal, in grey the MC background, in blue the fitted background and in red the signal plus the background.
6.4. Discrimination power of the variables

Figure 6.24 – Plots of Likelihood classifier. The cut on the Likelihood is required to give a signal efficiency of 90%. The mass has been plotted in a range between 30 GeV and 200 GeV, where the fit was performed.
Figure 6.25 – Plots of Boosted Decision Tree method. The cut on the BDT is required to give a signal efficiency of 90%. The mass has been plotted in a range between 30 GeV and 200 GeV, where the fit was performed.

The classifier which gives the highest S/B for a 90% signal efficiency is the Fisher discriminant.

The classifier that, instead, achieves the highest S/B is the BDT, which, for a 20% signal efficiency, reaches a background rejection of \( \sim 93\% \), and \( S/B \approx 5\% \) (see Fig.). Harder cuts will provide a more pure sample. On the other hand, decreasing signal efficiency also decreases \( S/\sqrt{B} \), which is an alarm for how hard the selection is. With very low signal efficiency and high background rejection efficiencies, very few events pass the strict cut, giving a poor statistical significance for the result.

Depending on the analysis one may choose one classifier or the other, opt for high or low signal efficiency.

To illustrate how the mass distribution changes requiring different signal efficiencies, plots for the Fisher discriminant only have been made at signal efficiency of 70% 50% 30% 10%.
6.4. Discrimination power of the variables

Figure 6.26 – Plots for the Fisher discriminant for a 70% signal efficiency cut. The mass has been plotted in a range between 30 GeV and 200 GeV, where the fit was performed.
Figure 6.27 – Plots for the Fisher discriminant for a 50% signal efficiency cut. The mass has been plotted in a range between 30 GeV and 200 GeV, where the fit was performed.
6.4. Discrimination power of the variables

Figure 6.28 – Plots for the Fisher discriminant for a 30% signal efficiency cut. The mass has been plotted in a range between 30 GeV and 200 GeV, where the fit was performed.
Figure 6.29 – Plots for the Fisher discriminant for a 10% signal efficiency cut. The mass has been plotted in a range between 30 GeV and 200 GeV, where the fit was performed.
6.4. Discrimination power of the variables

Figure 6.30 – Fisher discriminant for a 80% signal efficiency. The top right plot shows the comparison between data and MC. As the agreement between data and MC has not been fully understood, it is not surprising that in the mass distribution MC does not give a good prediction on data.

These last plots show the mass distribution for both MC and data in the case of the Fisher discriminant and the BDT for a 80% signal efficiency. As the agreement between data and MC has not been fully understood, it is not surprising that in the mass distribution MC does not give a good prediction on data.
Figure 6.31 – BDT for a 80% signal efficiency. The top right plot shows the comparison between data and MC. As the agreement between data and MC has not been fully understood, it is not surprising that in the mass distribution MC does not give a good prediction on data.
Conclusions

The present work combined a recent jet reconstruction and decomposition technique, called filtering, to a Multi-Variate analysis and employed it to extract a signal for weak vector bosons with high transverse momentum (pt), decaying into light quarks, from a much larger QCD background.

The filtering procedure was applied to jets with large transverse momentum, reconstructed using the Cambridge-Aachen algorithm with radius of 1.2. An important part of the work consisted in comparing MC simulations and data recorded by ATLAS in 2011. Data used in the analysis were taken in the first half of the year and correspond to a luminosity of $\sim 1.5 \text{ fb}^{-1}$. A discrepancy of 38% between data and MC in the unfiltered jet-pt distribution (as well in the distribution of other typical kinematic variables) was found. Such a difference can be explained by the MC generator used, PYTHIA, that, being only a Leading Order code, does not account for NLO corrections. Another source of disagreement comes from the absence of a trigger selection. For technical reasons trigger menus could not be stored in the samples used. Any trigger requirement could impact on the jet distributions shapes and thus could potentially explain the disagreement observed. Trigger information are currently being implemented in a new production.

MC events were then reweighted according to the jet transverse momentum to better describe the data. The agreement for most kinematic variables improved after this procedure, even if differences were still observed. The filtered jet-mass distribution showed significant discrepancies between data and MC especially in the low mass range.

In order to increase the signal over background ratio, a Multi-Variate analysis was performed. Eight variables were investigated, four of which were calculated using the constituents of the boosted jet in its center of mass frame (CM). W,Z events are expected to show a back to back topology, while QCD jets an isotropic distribution.

Variables showing little discriminating power were discarded. Of those left those showing little correlation with the others were selected. Correlation between the filtered jet mass and all the variables was also studied, to avoid having a final discriminator that produces artificial peaks for the background, in the signal mass region. Only five variables were eventually kept. These were combined together using three different classifiers: likelihood ratio, Fisher discriminant and Boosted
Decision Tree. For a signal efficiency of 90%, the Fisher discriminant gives the highest S/B, namely 3%, with a background rejection of the 27%. The classifier that gives the maximum S/B is the BDT, which, for a signal efficiency of 20% and a background rejection of nearly 95%, gives S/B = 5%.

Since the statistics on MC is limited when applying both the filtering and a cut on the classifier, a fit was done on the QCD background using a polynomial function. This allowed plotting the signal on top of a smooth background distribution and getting a clearer picture.

The techniques employed look promising. Nevertheless more work needs to be done. In order to ameliorate the MC description for the data, trigger information must be added and used. More MC events should be produced, especially in the high-$p_T$ range. A new tuning of the filtering parameters might decrease the discrepancy in the low mass region for filtered jets. In particular, the parameter $R_{\text{minsplit}}$ needs to be further investigated. Systematic uncertainties need also to be evaluated.

This thesis work has shown the strength and weakness of this approach, and while not a final study, it did contribute to a wider effort in the exploration of what is possible to do with the Cambridge-Aachen algorithm, both in terms of filtering and of jet shape variables.
Bibliography


[12] The CDF Collaboration, the D0 Collaboration, the Tevatron New Physics, and Higgs Working group. Combined cdf and d0 upper limits on standard model higgs-boson production with 2.1 - 5.4 fb-1 of data, 2009.


[29] M. Thiioye and State University of New York at Stony Brook. Topics in the measurement of the electrons with the ATLAS detector at the LHC. State University of New York at Stony Brook, 2008.
Appendix A

Monte Carlo generators

The Monte Carlo method refers to any procedure that makes use of random numbers and uses probabilistic statistics to solve problems. Monte Carlo methods are used extensively in numerical analysis and simulation of natural phenomena. In the context of particle physics, Monte Carlo generators are used to produce theoretical simulations of real events.

Often a variety of different programs are used: feasibly one program could generate a hard process while another evolves it through a parton shower algorithm and a third hadronizes the coloured products of the shower. Different Monte Carlo generators often simulate different physics models, and can be used with different matrix elements, PDFs, evolution equations, parton showers or hadronization models. Accordingly, comparing a variety of Monte Carlo models to data provides a test of the compatibility of different theories with experimental results. For each of the Monte Carlo generators used in ATLAS, the particle four-vectors are passed through a full simulation of the ATLAS detector and trigger that is based on GEANT4. The simulated events are then reconstructed and jets are calibrated using the same reconstruction chain as the for the data.
Appendix B

Luminosity

The proportionality factor between the rate of a certain event and the cross section for that particular event is the **instantaneous luminosity** $\mathcal{L}_{\text{int}}$:

$$\dot{N} = \mathcal{L}_{\text{int}} \cdot \sigma$$  \hspace{1cm} (B.1)

The luminosity can also be expressed as a function of the parameters of the beams:

$$\mathcal{L}_{\text{int}} = \frac{N_{b1} N_{b2} f_{\text{rev}} k_b}{2\pi \sqrt{(\sigma_{x1}^2 + \sigma_{xy}^2)(\sigma_{y1}^2 + \sigma_{y2}^2)}} \cdot \exp \left[ -\frac{(\bar{x}_1 - \bar{x}_2)^2}{2(\sigma_{x1}^2 + \sigma_{xy}^2)} - \frac{(\bar{y}_1 - \bar{y}_2)^2}{2(\sigma_{y1}^2 + \sigma_{y2}^2)} \right]$$  \hspace{1cm} (B.2)

Where $N$ are the bunch populations, $f_{\text{rev}}$ is the revolution frequency, $k_b$ is the number of bunches per beam, are the beam sizes and the subscripts 1, 2, $x$ and $y$ indicate beam 1, beam 2, horizontal and vertical respectively. The nominal beam life time is about 22h although the luminosity is reduced by a factor 2 after 10h, imposing new beam injection to LHC. The luminosity needs to be measured frequently, notably to control the quality of data. The generic term of *lumi-block* is used and corresponds to a short *stable* beam period $\theta$ (min).

The **integrated luminosity** $\mathcal{L}$ is simply the integral of the instantaneous luminosity over a certain period of time $T$:

$$\int_T \dot{N} \cdot dt = N = \mathcal{L} \cdot \sigma$$  \hspace{1cm} (B.3)

Fig. B.1 shows the cumulative luminosity versus day delivered to (green), and recorded by ATLAS (yellow) during stable beams and for pp collisions at 7 TeV centre-of-mass energy in 2011. The delivered luminosity accounts for the luminosity delivered from the start of stable beams until the LHC requests ATLAS to turn the sensitive detector off to allow a beam dump or beam studies. Data were taken between March 2011 and December 2011. The first week of December was the last week of physics runs for the LHC during 2011 before the annual winter shutdown (Dec. 2011 - Feb. 2012).

Data are grouped in periods. Associated with each period are:
Appendix B. Luminosity

Figure B.1 – Cumulative luminosity versus day delivered to (green), and recorded by ATLAS (yellow) during stable beams and for pp collisions at 7 TeV centre-of-mass energy in 2011 (taken from https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults).

1. beginning and ending run numbers

2. special comments about the detector or other physics condition associated with the period

3. luminosity for that period, and other other information.

For each period certain run numbers are selected according to quality criteria. The ATLAS list of “good” runs is called GRL (Good Run List) and defines a list of runs and luminosity blocks to be considered.

Experimental data and MC information are stored in D3PDs. D3PD is the ATLAS name for a plain ROOT file, containing only ROOT TTree and/or histograms. The braches and/or leaves of these ROOT TTrees store all the events details (number of vertices in the event, number of jets, trigger information...).

Out of all the data collected, only the run numbers that match with the ones in the GRL are considered. The luminosity in Eq. (B.3) accounts only for this runs.