Name of the thesis

Konstantinos Papadimos

Contents

1 Theoritical Overview

1.1 Units in particle physics

Particle physics, study the properites of subatomic particles and their interactions. To describe such microscopic phenomena and quantities, an appropriate system of units must be adopted, for if the SI system of units was to be used, one would have to deal with large exponents. More over, it is more practical to utilize a system that is based on the typical length and time scales found in particle physics[modern particle phyics].

The foundamental constants of the special theory of relativity, as well as quantum mechanics are $\hbar$, c, and GeV and they can be used as a basis to form a system of units, the Natural Units. Natural units can be further simplified by chosing:

$$c = \hbar = 1$$

The table below sumarizes the relationship between natural units and S.I

1.2 Special Relativity

1.2.1 Four-Vectors - Lorentz transformations

Let S and S' be two inertial reference systems, with S' moving at (a relativistic) 2velocity u relative to S. The coordinates are chosen such that the motion is along the x-axis in both reference systems. The

Relationship between natural units and S.I		
Quanity	Natural units($\hbar = c = 1$)	S.I
Energy	GeV	Kgm^2s^{-2}
Momentum	GeV	Kgm^2s^{-2}
Mass	GeV	Kg
Time	GeV^{-1}	s
Length	GeV^{-1}	m

Table 1: Some basic quantites in S.I and in Natural units .

clocks of both S and S' are synchronized so that when $x = x' = 0$, $t = t' = 0$. For an event with coordinates (x, y, z, t) in S, the coordinates in S' are given by the Lorentz transformations:

$$\begin{aligned}(x')^0 &= \gamma(x^0 - \beta x^1)\\ (x')^1 &= \gamma(x^1 - \beta x^0)\\ (x')^2 &= x^2\\ (x')^3 &= x^3\end{aligned} \tag{1}$$

where $\beta \equiv \frac{u}{c}$, $x^0 \equiv ct$, $x^1 \equiv x$, $x^2 \equiv y$, $x^3 \equiv z$

The elements x^i, i = 0, 1, 2, 3 define the position four vector. Mathematicaly four vectors are 4 dimentional, rank 1 tensors that transform according to lorentz transformations

With the introduction of four vectors, using Einstein's summation convention, lorentz transformations can be written as:

$$(x')^i = \Lambda^i_j x^j \tag{2}$$

Where Λ, is the Lorentz transformation matrix, a rank 2 tensor:

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0\\ -\gamma\beta & \gamma & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix} \tag{3}$$

The following quantity does not change under lorentz transformation:

$$I^2 = -(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2 = -(x'^0)^2 + (x'^1)^2 + (x'^2)^2 + (x'^3)^2 \tag{4}$$

or written in a more compact form:

$$I = g_{\mu\nu}x^\mu x^\nu = x^\mu x_\mu \tag{5}$$

where $g_{\mu\nu}$ is the Minkowski (metric) tensor:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 1 & 0 & 1 & 0\\ 1 & 0 & 0 & 1\end{pmatrix} \tag{6}$$

Such a quantity as I is called **invariant**

The introduction of four vectors and the metric tensor, yield the Minkowski Space Time where points need 4 coordinates to be fully described(1, time like and 3 space like) and the distance between them is beeing defined by the Minkowski tensor. The necessity of four vectors, in order to describe non scalar quantities(such as velocity and momentum) in minkowski space time, is therefore evident.

1.2.2 Energy and Momentum

According to the principle of relativity, the laws of physics must be the same in all inertial reference systems. Hence, if momentum is conserved in one inertial frame of reference, it must also be conserved in all others. It is evident that the momentum of a moving particle must be defined in an appropriate manner, to satisfy the principle of relativity [3]. The four momentum is therefore defined as:

$$p^\mu = m\eta^\mu \tag{7}$$

Where $\eta^\mu = \frac{dx^\mu}{dt'}$, the four velocity of the particle.

The timelike component of four momentum, expresssed in natural units is $p^0 = \gamma m$. The 3 space like components, constitue the vector momentum, $\vec{p} = \gamma m\vec{\beta}$. The relativistic energy is defined as:

$$E = \gamma m = p^0 \tag{8}$$

Thus, the components of four momentum are:

$$p^\mu = (E, \vec{p}) \tag{9}$$

At this point we are able to calculate the invariant "interval" $p^\mu p_\mu$:

$$p^\mu p_\mu = E^2 - |\vec{p}|^2 = m^2 \tag{10}$$

What is invariant in the case of momentum four vector, is the particle's mass. This quantity is called **invariant mass** of a particle as all observers, in different frames of references, agree upon its value. It is the invariant mass of particles that we can(or we try to) measure, in CERN as well.

1.3 The standard model

The Standard Model (SM)of particle physics is a theoretical framework that describes the fundamental particles and their interactions through the strong, electromagnetic, and weak forces. Each force is described by a corresponding quantum field theory(QFT). Namely, electromagnetic and weak interactions are described by the electroweak theory and the strong interactions by quantum chromodynamics(QCD). The interactions between particles in each QFT, are described in terms of the exchange of a spin-1 gauge boson. The photon, is the gauge boson of QED, while the gluon, which like the photon has no mass, is the force-carrying particle in the strong interaction. The charged $W^+$and W^- bosons mediate the weak charged current interaction, which is responsible for β decay and fusion, while the weak neutral current interaction, is mediated by the electrically neutral Z boson. These interactions are also governed by the principles of symmetry and conservation, which dictate that certain properties, such as charge and energy, are conserved during particle interactions

The fundamental particles that comprise all matter according to SM are the already mentioned gauge bosons, quarks, and leptons. The quarks and leptons are organized into three generations, with each generation containing two types of leptons and two types of quarks. The leptons are either negatively charged, with a charge of -1, or electrically neutral. The quarks, on the other hand, have fractional charges of either -1/3 or +2/3, and are characterized by their color, which can be blue, green, or red. Additionally, for each elementary fermion, there is a corresponding antifermion with the same mass and spin but with an opposite electric charge.

Completing the picture of fundamental particles is the scalar boson, the Higgs boson, responsible for giving mass to the other particles. A brief summary of the fundamental particles is presented in Figure 1.

The SM has undergone extensive testing through high-energy experiments at CERN, with its predictions confirmed with a high degree of precision. However, the model has limitations, such as its inability to account for dark matter or the observed imbalance between matter and antimatter in the universe.

Despite its limitations, the Standard Model remains a cornerstone of modern physics, and its continued study and refinement is essential to advancing our understanding of the universe at its most fundamental level.

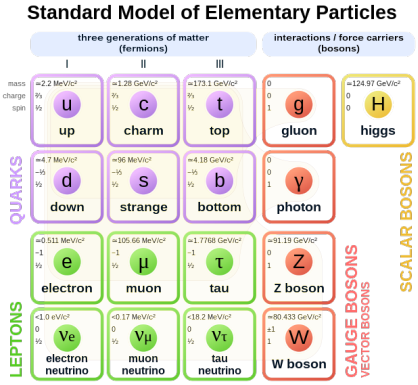

Figure 1: Summary of the elementary particles. All matter around us is made up by 12 fermions!

2 Accelerators and Detectors: the LHC and the CMS

2.1 The Large Hadron Collider

Our knowledge in the field of high energy physics has been largely obtained through fixed target experiments that used proton and electron accelerators. However, over the last decades, the significance of colliding beam experiments has been rising. Such experiments involve two particle beams that rotate in opposite directions and collide at multiple points around the ring. The key advantage of colliding beam machines is their ability to produce new particles due to the high center of mass energy created during the collision. This energy increases linearly as E, rather than as $E^{1/2}$, in fixed target experiments, and almost all of it is utilized in generating new particles[modern particle physics].

One great example of a colliding beam machine is The Large Hadron Collider (LHC). The LHC has been instrumental in many groundbreaking discoveries, with the most famous one beeing the Higgs boson, and has helped scientists to further our understanding of the fundamental nature of the universe. The LHC encompasses a 27-kilometre ring consisting of superconducting magnets with numerous accelerating structures along its length.

Within the accelerator, a strong magnetic field is accelerating the two counterrotating proton beams to velocities near that of the speed of light uppon collision. The thousands of superconducting magnets, responsible of the generation of the magnetic field, are of varying sizes and types. Dipole magnets, 1232 in total and 15 meters in length, are utilized to bend the beams and quadrupole magnets, 392 in total and 5-7 meters long, focus the beams. Prior to collision, another type of magnet is used to compress the particles, increasing the likelihood of collisions source.

2.2 The Compact Muon Solenoid

The main goal of the Compact Muon Solenoid (CMS), as a general purpose particle detector, is to to reconstruct the Feynman diagram associated with any interaction that might happen inside the LHC. The first and foremost interactions that happen are the collisions between the beams, which generate individual interactions known as *events*. Even though most of the particles associated with an event are unstable, their final decay products, are stable enough to reach the detector and be measured. In the rest of the chapter I will give a brief overview of the CMS detector and discus how does it detect particles.

2.2.1 Overview

The CMS detector consists of 5 compartments, each with unique functionality, that are organised in several coaxial layers. The Silicon Tracker, located in the innermost part of CMS, includes silicon pixel vertex detectors and silicon strip detectors, which trace the position and momentum of charged particles. The Electromagnetic Calorimeter (ECAL), the second layer, is composed of PbWO4 crystals and

intended to detect photons and electrons. The Hadronic Calorimeter (HCAL), the third layer, is designed to identify hadrons. The Superconducting Solenoid Magnet, the fourth layer, is an solenoid coil that generates a constant magnetic field of 4 T along the direction of the beam. Due to the deflection of the trajectories of charged particles by the magnet, it becomes possible to measure their momentum. The final, outer most layer, is responsible for the measurement of muon the tracks. Figure 2 provides a sectional view of the CMS detector[`https://cms.cern/news/cms-detector-design`].

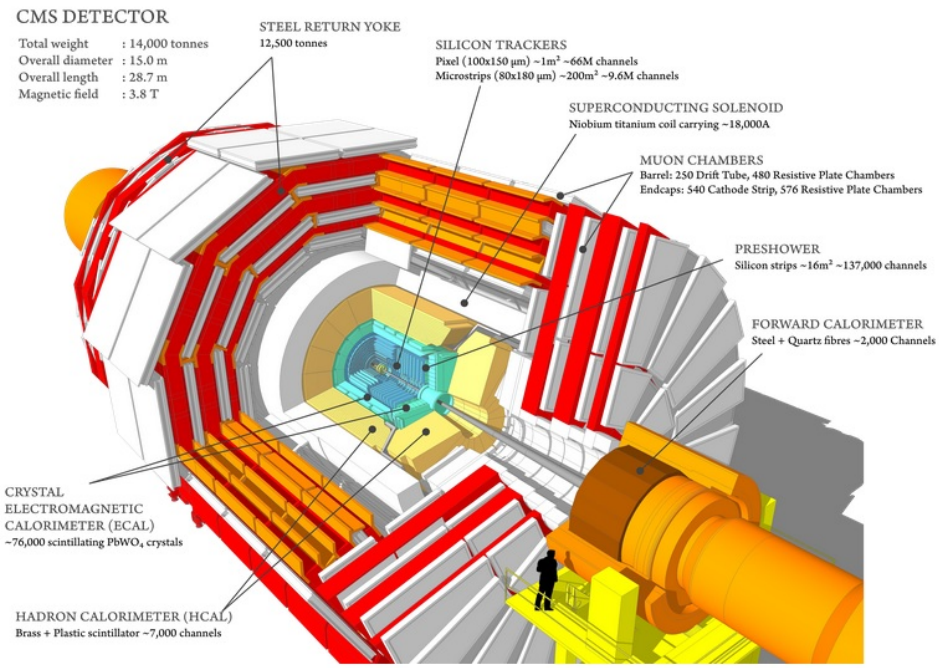

Figure 2: A cross-sectional perspective of the CMS detector

2.2.2 Coordinate convention at the CMS

Given the solenoid geometry of the CMS detector, it is more convenient to use a spherical type of coordinates(r, ϕ, θ). The origin is located at the collision point and the z axis is parallel to the beam as shown in figure 3. In this system, the momentum of a particle(or any other vector) can be analyzed in a component parallel to the z axis and one component perpendicular to the z axis(Transverse mometnum). Transverse mometnum is defined as follows:

$$|\vec{P_T}| = \sqrt{P_x^2 + P_y^2} = |\vec{P}| \sin\phi \tag{11}$$

Where $|\vec{P}| = \sqrt{P_x^2 + P_y^2 + P_z^2}$. The CMS detector, measures the transverse energy(source) of particles, and thus it is useful to work with the transverse momentum P_T. The azimuth angle $\phi \in [0, 2\pi)$ coordinate is the angle between P_t and x axis and the polar angle $\theta \in [0, \pi]$ is the angle between the momentum vector and the z axis.

Due to the relativistic nature of the phenomena taking place inside LHC, it is more usefull to work with lorentz invariant quantities [V. Chiochia (2010) Accelerators and Particle Detectors from University of Zurich]. Thus, instead of working with the polar angle it is more convenient to introduce the lorentz invariant *pseudorapidity* $\eta \in [-\infty, +\infty]$. Pseudorapidity is defined as:

$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{12}$$

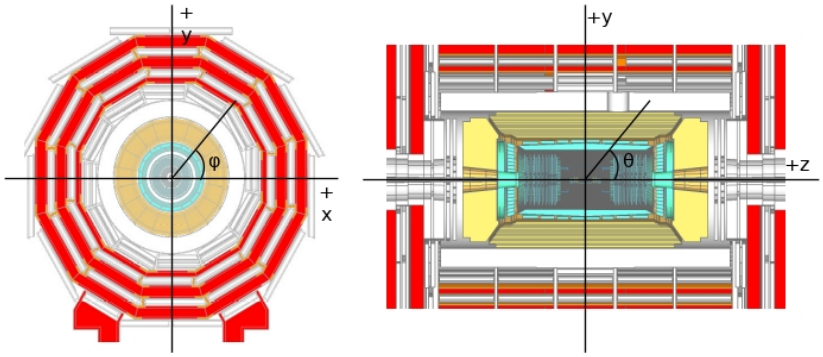

Figure 3: CMS coordinates

The cartesianp_x, p_y, p_z momentum components are related to the P_T, η, ϕ components by the following transformation relations:

$$\begin{aligned} p_x &= P_T \cos\phi \\ p_y &= P_T \sin\phi \\ p_z &= P_T \sinh\eta \\ |\vec{P}| &= P_T \cosh\eta \end{aligned} \tag{13}$$

2.2.3 Position tracking and momentum measurements: Silicon Tracker

The Silicon tracker measures the positions of charged particles at a number of points, thus it is able to record their trajectory. Given the radius of curvature of the particle's track(due to the 4T magnetic field of the super conducting solenoid), the tracker provides sufficient information, to reconstruct the momentum of the particle. More over, the geometrical location of the trajectory gives direct information regarding the position of the particle. Therefore, the silicon tracker measurements provide information regarding the P_T, η and ϕ of the particles that it detects.

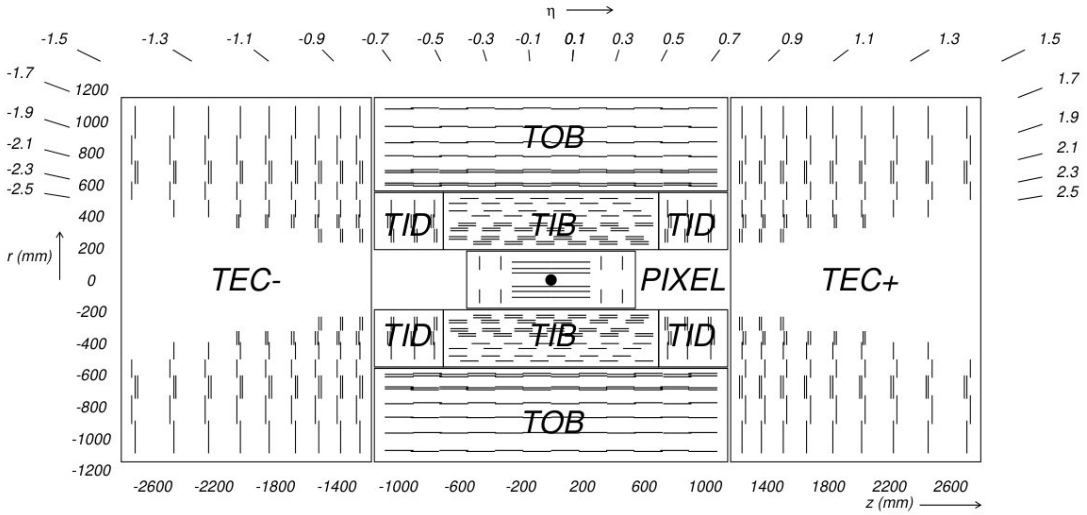

Figure 4: Schemtic illustration of a crossection of the CMS Tracker

A schematic representation of the Sillicon Tracker's corss section, can be viewed on figure 4. The tracker consists of a silicon pixel detector and a silicon strip detector. The silicon pixel detector is composed of two sub-detectors. Namely, the barrel which consists of three layers covering the region $|\eta| < 2.2$ and at $r = 4.4, 7.3$ and $10,2$cm. The end caps, are two discs of pixel modues, located one on each side, that complete the design of the silicon pixel detector. The pixel detector improrves the

trajectory and position measurements, by providing two-dimensional measurements of the charged particles' hit positions. (source).

The silicon strip detector, covers the radial region $r \in [20, 116]$ cm. and is comprised of four inner barrel (TIB) layers and two inner endcaps (TID). The TIBs are assembled in shells and each TID consists of three small discs. The outer barrel (TOB) encompasses both TIB and TID and contains six concentric layers. The tracker is closed off on either end by two endcaps (TEC). Measurements at the silicon strip detector give information regarding the path of each particle allows the distinction of separate particle trajectories.

2.2.4 Energy Measurements: Calorimeters

Apart from measuring position and momentum, determining the energy of particles produced in LHC collisions is crucial. In the Compact Muon Solenoid (CMS) experiment, this information is obtained from particle interactions with matter in the calorimeters. Particles that are stable enough to reach the detector without decaying are either leptons, photons, or hadrons. The interactions between electrons, photons, and matter are of electromagnetic nature, while those between hadrons (charged or neutral) and matter are strong interactions. Therefore, the CMS experiment employs two types of calorimeters: the Electromagnetic Calorimeter (ECAL), located at the innermost layer, which measures the energy of photons and electrons, and the Hadron Calorimeter (HCAL), situated at the outer shells of the calorimeter section.

- Electromagnetic Calorimeter (ECAL)

  Figure 5(source) provides a view of the Electromagnetic calorimeters inside the CMS. The ECAL is composed of lead tungstate (PbWO4) crystals and is designed with a central barrel section (EB) and two endcaps (EE) that cover a range of pseudorapidities up to $1.48 \leq |\eta| \leq 3.0$(source). The crystals are highly dense and scintillate when high-energy photons or electrons interact with them. When a particle passes through the ECAL, it deposits its energy in the form of electromagnetic showers, which cause the crystals to emit light. The emitted light is then captured and amplified in order to estimate the energy of the incoming particle. The high-density crystals of the ECAL make it possible to accurately measure the energy of photons and electrons with high precision and resolution.

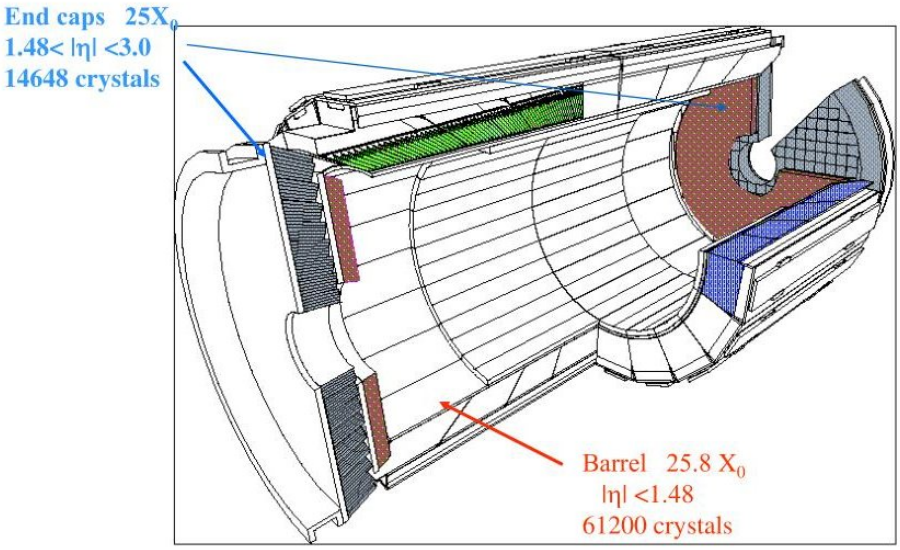

Figure 5: Schemtic illustration of the Ecal parts inside CMS

- Hadron Clorimeter (HCAL)

The hadrons that manage to reach the detector, fly off the ECAl and interact with the Hadron Calorimeter. The HCAL consists of alternating layers of absorber material and plastic scintillator tiles that detect particles generated by the hadrons as they interact with the absorber. When particles pass through the HCAL, they interact with the absorber material, producing showers of particles that create signals in the scintillator tiles. These signals are then read out and processed to measure the energy of the incoming hadrons. The HCAL has both a barrel section(HB), with pseudorapidity coverage at $|\eta| < 1.3$ and endcap(HE), covering a range of pseudorapidities $1.3 \leq |\eta| \leq 3.0$. The HCAL is highly effective at measuring the energy of hadrons due to its high-density absorber material and precise arrangement of scintillator tiles (source)

2.2.5 Detecting Muons

In the outer regions of the CMS detectors, are located the muon chambers. They are the final part of the detector and are designed soley for the detection of muons, which due to their large mass(207 times greater than the electron mass) muons travel a longer distance in matter than electrons. Thus, their energy cannot be measured in ECAL.

The muon chambers consist of 250 drift tubes (DTs) and 540 cathode strip chambers (CSCs), which track the positions of the particles. Additionally, there are 610 resistive plate chambers (RPCs) and 72 gas electron multiplier chambers (GEMs), making a total of 1400 chamber units . The use of multiple layers of detectors and different types of chambers makes the system robust and able to filter out background noise.

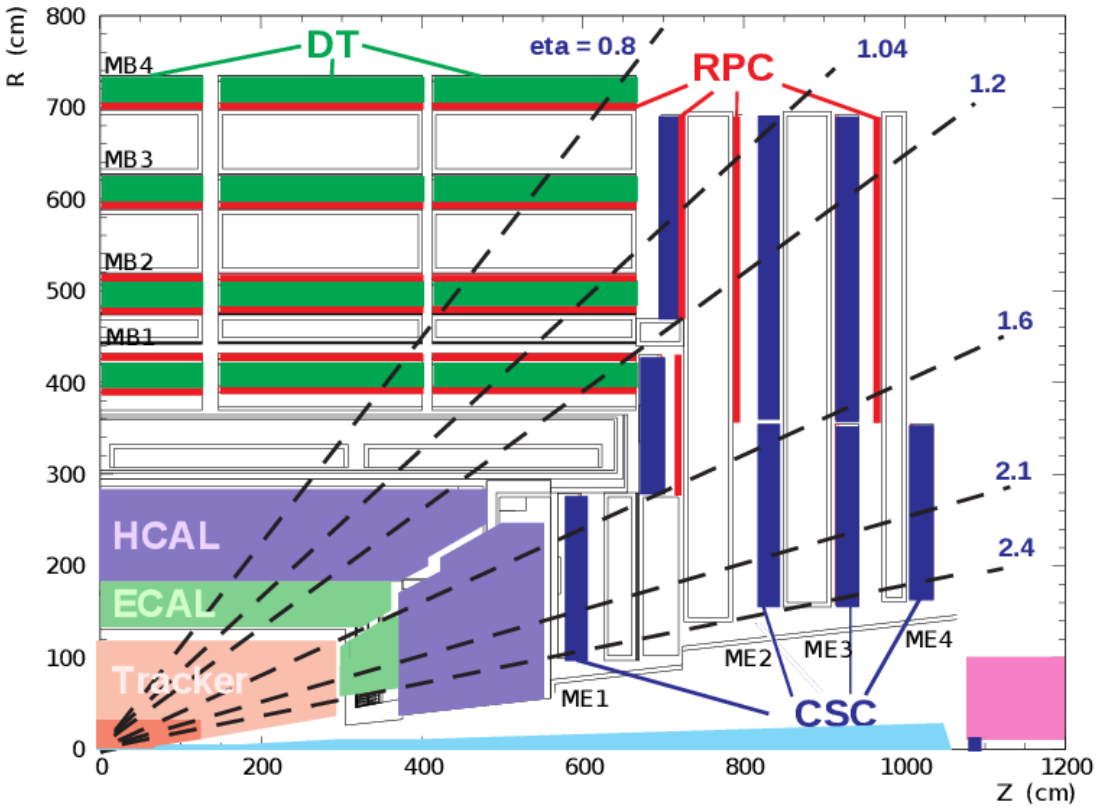

Figure 6: A quarter sectional view of the CMS muon chambers. The beamline is perpendicular to the plane of the page

Figure 6 illustrates the arrangement of the four different kinds of chambers. In the "barrel region," which surrounds the beam line, the DTs and square-shaped RPCs are grouped in coaxial cylinders. The CSCs, trapezoidal RPCs, and GEMs are located at the end cap region of the barrel . This arrangement

allows for accurate measurements of the muons' trajectories and momenta in different regions of the detector.(source)

3 Machine Learning for Classification

The main goal Machine Learning is set to achieve, is the development of algorythms equiped with the capabillity of learning from data automatically. In particular, an artificially intelligent system must have the ability to identify objects in its surroundings, as well as anticipate the actions of its environment, in order to make informed decisions. Due to this, machine learning techniques tend to be more oriented towards forecasting, rather than prediction.

3.1 Decision trees and Supervised Learning

The distinction between different particles, can be regarded as a classification problem where the target, is the prediction of a categorical output variable(i.e. lepton, boson), based on one or more input variables(i.e momenta components). Classification problems in machine learning can be solved with supervised learning. In such procedure, a training data set is being used for the development(training) of a model that is able to perform the classification task. The output of the model is then beeing tested and evaluated on previously unseen data.

Before presenting any specific method of solving classification problems It is important to present an overview of the key elements in supervised learning.

3.1.1 Supervised learning

Let us pose the following problem: Given a data set $D = (\vec{X}, \vec{y})$, where $\vec{X}$ is a matrix of the independet variables and $\vec{y}$ is a vector of dependent variables, we want to find a model $f(\vec{x}; \vec{\theta})$, that can predict an output from a set of input variables. Moreover, we want to be able to judge the performance of the model on a given data set. To do that we need to define a cost function $C(\vec{y}, f(\vec{X}; \vec{\theta}))$, such that the model will have to find the parameters θ that minimize the cost function.[5]

This the mathimatical pustulation of a supervised learning problem. I will now, in brief, discuss the role and interpretation of each of the 'ingredients' stated above.

- Model

  The model, is a mathematical function $f : \vec{x} \to y$ of the parameters θ. Given a set of parameters, the output of the function, the prediction y_i, is derived from the input variables $\vec{x}$. The parameters are undefined. The task of the training is to estimate the set of parameters from the training data set. In a classification problem(something is of type a or it is not), it is possible to use the logistic transformation of the function output, to obtain the probability of the positive class.

- Cost function

  The cost function, also known as an objective function, is represented by mathematical function and it measures how well a model fits the training data. The cost function is used to train the model by finding the best set of parameters θ that minimize the function. In machine learning, the objective function, usually consists of two parts: a training loss function (L) and a regularization term (Ω).

  $$obj(\theta) = L(\theta) + \Omega(\theta) \tag{14}$$

  The training loss function measures how predictive the model is with respect to the training data. A common choice of training loss function is the logistic loss, which is used for logistic regression(classification) and is given by

  $$L(\theta) = \sum_i [y_i \ln(1 + e^{-\hat{y}_i}) + (1 - y_i \ln(1 + e^{\hat{y}_i}))] \tag{15}$$

where y_i is the true label and $\hat{y}_i$ is the predicted label.

The regularization term, $\Omega(\theta)$, controls the complexity of the model, which helps to avoid overfitting. Overfitting occurs when a model is too complex and starts to extract local features from the training data. The model thus, looses its generalization power to new unseen data. Regularization helps to prevent overfitting by adding a penalty term to the cost function, which discourages the model from having too many parameters or too complex a structure.

The following figure gives an example of overfitting due to a very complex and very simple model.

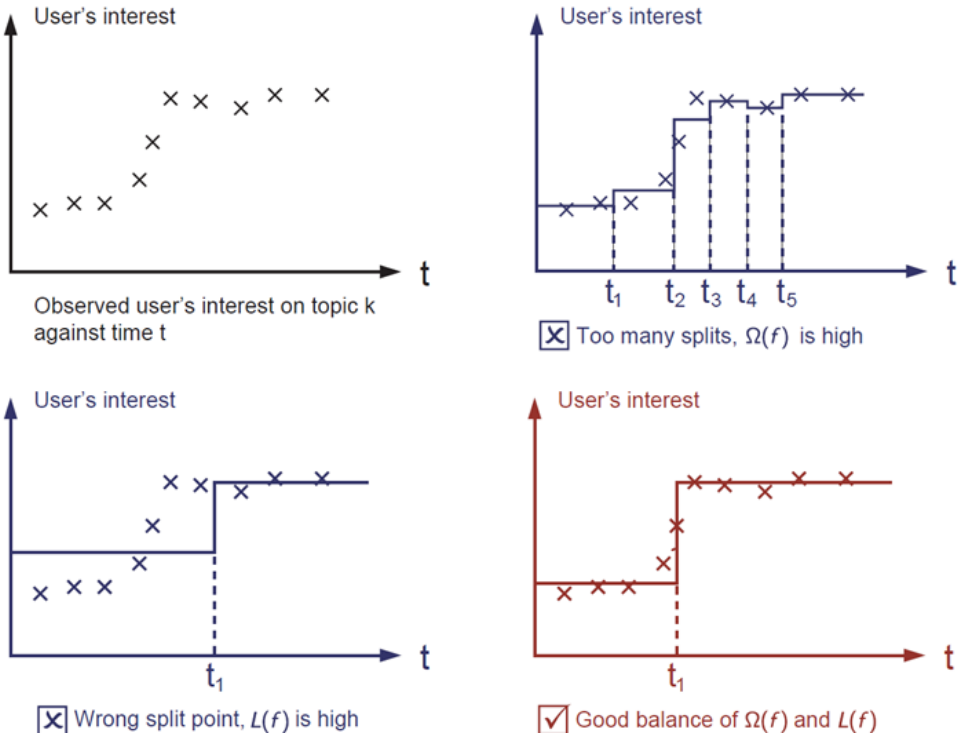

Figure 7: Examples of over fitting and under fitting. The top right model, places too many cuts. Even though it succesfully describes the trend, the splits seem to correspond only on the specific data set, therefore it is overfitted. The bottom left model places too few and impresice cuts. The bottom right model seems to succesfully describe the trend while its simplicity infers that It has not sacrificed its generalization power.

3.1.2 Decision Trees

A decision tree is a flowchart-like tree structure, where each internal node represents a feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome.

Formally, a decision tree can be represented as a set of rules or conditions in the form of:

```
f(X)={condition1,condition2,..condition_n}
```

where each condition is a tuple of the form (feature, threshold, comparison operator) and the final outcome is represented by the leaf node. For example, consider the decision tree of figure 8 that classify fruits based on color, shape, size, and taste. Let X be the input `X = {"red", "smal", "sour"}`. Then `f(X) = "grape"` [2]

- Decision Tree Ensembles

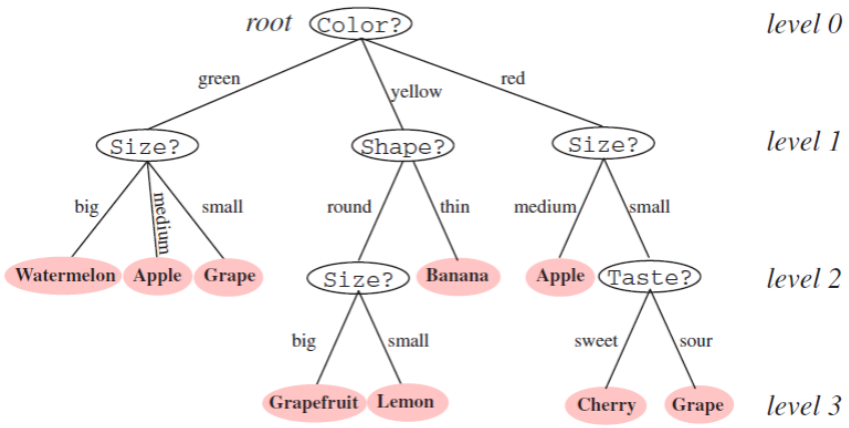

Figure 8: Example of a a decision tree that clasifies fruits

The tree ensemble model consists of a set of classification and regression trees (CART). Let $\mathcal{F}$ be the set of all possible CART's and $f_k \in \mathcal{F}$, a function that represents a CART. The model in discussion then, can be written as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \ f_k \in \mathcal{F} \tag{16}$$

If $\hat{y}_i$ represents the prediction of the tree, given an input variable x_i, the real label of x_i will be denoted as y_i . The objective function will be of the form:

$$obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{i=1}^{t} \omega(f_i) \tag{17}$$

where $\omega(f_i)$ is the complexity of a given tree and l is the loss function.

- Tree boosting

  As stated earlier, the model is beeing trained, to learn those trees f_k that minimize the objective. The resulting model then, will be an ensemble of those functions f_k. The optimization of the objective, is a problem that cannot be solved with the traditional methods. Instead, the model is being iteratively trained in an additive manner.[1] let the prediction value at the t-th iteration be $\hat{y}_i^{(t)}$. In the next iteration(t+1), the chosen function f_{t+1}, is such that if added to the model, the resulting prediction $\hat{y}_i^{(t+1)}$ will minimize the cost function:

$$\begin{aligned} \hat{y}_i^{(0)} &= 0 \\ \hat{y}_i^{(1)} &= \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(2)} &= \hat{y}_i^{(1)} + f_2(x_i) \\ &\dots \\ \hat{y}_i^{(t)} &= \hat{y}_i^{(t-1)} + f_t(x_i) = \textstyle\sum_{k=1}^{K} f_k(x_i) \end{aligned} \tag{18}$$

  The objective at step t is:

$$obj^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^{t} \omega(f_i) = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \omega(f_i(t)) \tag{19}$$

Taylor expanding the loss function $l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$, around f_t, up to the second order and neglecting terms, referring to previous rounds, the specific objective becomes:

$$\sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t) \tag{20}$$

Where

$$\begin{aligned} g_i &= \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \\ h_i &= \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \end{aligned} \tag{21}$$

This is the minimization goal for f_t . [4]

References

[1] Tianqi Chen and Carlos Guestrin. "XGBoost". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016. DOI: 10.1145/2939672.2939785. URL: https://doi.org/10.1145%2F2939672.2939785.

[2] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley, 2006. ISBN: 9788126511167. URL: https://books.google.com.cy/books?id=NR-SzW2t7WYC.

[3] D. Griffiths. *Introduction to Elementary Particles*. Physics textbook. Wiley, 2008. ISBN: 9783527618477. URL: https://books.google.com.cy/books?id=Wb9DYrjcoKAC.

[4] *Introdution to Boosted Trees*. URL: https://xgboost.readthedocs.io/en/stable/tutorials/model.html (visited on 01/28/2022).

[5] Pankaj Mehta et al. "A high-bias, low-variance introduction to Machine Learning for physicists". In: *Physics Reports* 810 (May 2019), pp. 1–124. DOI: 10.1016/j.physrep.2019.03.001. URL: https://doi.org/10.1016%2Fj.physrep.2019.03.001.