

SEARCH FOR RESONANT DOUBLE HIGGS PRODUCTION WITH bbZZ  
DECAYS IN THE  $b\bar{b}\ell\ell\nu\bar{\nu}$  FINAL STATE IN pp COLLISIONS AT  $\sqrt{s} = 13$  TeV

by

Rami Kamalieddin

A DISSERTATION

Presented to the Faculty of  
The Graduate College at the University of Nebraska  
In Partial Fulfilment of Requirements  
For the Degree of Doctor of Philosophy

Major: Physics and Astronomy

Under the Supervision of Professor Ilya Kravchenko

Lincoln, Nebraska

May, 2019

SEARCH FOR RESONANT DOUBLE HIGGS PRODUCTION WITH bbZZ  
DECAYS IN THE  $b\bar{b}\ell\ell\nu\bar{\nu}$  FINAL STATE IN pp COLLISIONS AT  $\sqrt{s} = 13$  TeV

Rami Kamalieddin, Ph.D.

University of Nebraska, 2019

Adviser: Ilya Kravchenko

Since the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments, most of the quantum mechanical properties that describe the long-awaited Higgs boson have been measured. Due to the outstanding work of the LHC, over a hundred of  $\text{fb}^{-1}$  of proton collisions data have been delivered to both experiments. Finally, it became sensible for analyses teams to start working with a very low cross section processes involving the Higgs boson, e.g., a recent success in observing ttH and VHbb processes. One of the main remaining untouched topics is a double Higgs boson production. However, an additional hundred of  $\text{fb}^{-1}$  per year from the HL-LHC will not necessarily help us much with the SM double Higgs physics, as the process may remain unseen even in the most optimistic scenarios. The solution is to work in parallel on new reconstruction and signal extraction methods as well as new analysis techniques to improve the sensitivity of measurements. This thesis is about both approaches: we have used the largest available dataset at the time the analysis has been performed and developed/used the most novel analysis methods. One such method is the new electron identification algorithm that we have developed in the CMS electron identification group, to which I have had a privilege to contribute during several years of my stay at CERN.

The majority of this thesis is devoted to techniques for the first search at the LHC for double Higgs boson production mediated by a heavy narrow-width resonance in

the  $b\bar{b}ZZ$  channel:  $X \rightarrow HH \rightarrow b\bar{b}ZZ^* \rightarrow b\bar{b}\ell\ell\nu\bar{\nu}$ . The measurement searches for the resonant production of a Higgs boson pair in the range of masses of the resonant parent particle from 250 to 1000 GeV using  $35.9 \text{ fb}^{-1}$  of data taken in 2016 at 13 TeV. Two spin scenarios of the resonance are considered: spin 0 and spin 2. In the absence of the evidence of the resonant double Higgs boson production from the previous searches, we proceed with setting the upper confidence limits.

“... a place for a smart quote”

*Zelenskiy, 2019.*

## ACKNOWLEDGMENTS

This will be a long list!

---

## Table of Contents

---

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>x</b>
<b>1 LHC and the CMS experiment</b>	<b>1</b>
1.1 The Large Hadron Collider . . . . .	1
1.1.1 The history of the LHC . . . . .	1
1.1.2 The layout of the LHC . . . . .	2
1.1.3 LHC operations . . . . .	3
1.1.4 Luminosity . . . . .	4
1.1.5 LHC infrastructure . . . . .	6
1.1.5.1 Magnets . . . . .	6
1.1.5.2 Cooling System . . . . .	9
1.1.5.3 Radio Frequency Cavities . . . . .	11
1.1.5.4 Vacuum System . . . . .	12
1.2 The CMS experiment . . . . .	14
1.2.1 The CMS coordinate system . . . . .	18
1.2.2 The Inner Tracker . . . . .	19
1.2.3 The ECAL . . . . .	21

1.2.4	The HCAL . . . . .	24
1.2.5	The Superconducting Solenoid . . . . .	26
1.2.6	The Muon Tracker . . . . .	27
1.2.7	The Triggers and DAQ . . . . .	29
1.2.7.1	The L1 Trigger . . . . .	30
1.2.7.2	The HLT Trigger . . . . .	32
1.2.7.3	The DAQ system . . . . .	33
1.2.8	The CMS design . . . . .	34
1.3	Physics Object Reconstruction in CMS . . . . .	35
1.4	Track Reconstruction . . . . .	36
1.4.1	Muon tracking . . . . .	38
1.4.2	Electron tracking . . . . .	38
1.4.3	Primary Vertex reconstruction . . . . .	39
1.5	Particle level objects . . . . .	41
1.5.1	Particle Flow links and blocks . . . . .	41
1.5.2	Muons . . . . .	42
1.5.3	Electrons and isolated photons . . . . .	43
1.5.4	Hadrons and non-isolated photons . . . . .	44
1.5.5	Jets and jet corrections . . . . .	45
1.6	Other important physics quantities and objects . . . . .	47
1.6.1	The b tagging and secondary vertices . . . . .	47
1.6.2	Missing transverse momentum . . . . .	49
1.6.3	Pileup interactions . . . . .	50
1.6.4	Lepton isolation . . . . .	50
1.6.5	Datasets and Trigger Paths . . . . .	51
1.6.6	Data-Monte Carlo corrections . . . . .	53



---

## List of Figures

---

1.1	Schematic layout of the LHC. . . . .	2
1.2	LHC dipoles . . . . .	7
1.3	LHC quadrupoles . . . . .	8
1.4	13 kA high-temperature superconducting current lead. . . . .	9
1.5	LHC cryogenic states and the temperature scale. . . . .	11
1.6	RF cavities module. . . . .	12
1.7	Beam screen. . . . .	13
1.8	CMS experiment with the main sub-detectors. . . . .	15
1.9	The logo of the CMS experiment that is showing curved trajectories of the emerging muons. . . . .	17
1.10	Coordinate system of the CMS detector . . . . .	19
1.11	The inner tracker . . . . .	20
1.12	The ECAL . . . . .	22
1.13	The HCAL . . . . .	25
1.14	The CMS superconducting solenoid . . . . .	26
1.15	The CMS muon tracker . . . . .	28
1.16	The CMS L1 trigger layout . . . . .	31

---

**List of Tables**

---

## CHAPTER 1

---

### **LHC and the CMS experiment**

---

CERN accelerator complex is a sequence of machines that produces and accelerates "bunches" of  $10^{11}$  protons to nearly the speed of light. In the Large Hadron Collider (LHC) the bunches collide at specific interaction points (IP), where the four main experiments are located: ALICE, ATLAS, CMS, and LHCb. We will start this section with the discussion of the LHC machine and then describe the CMS detector.

## **1.1 The Large Hadron Collider**

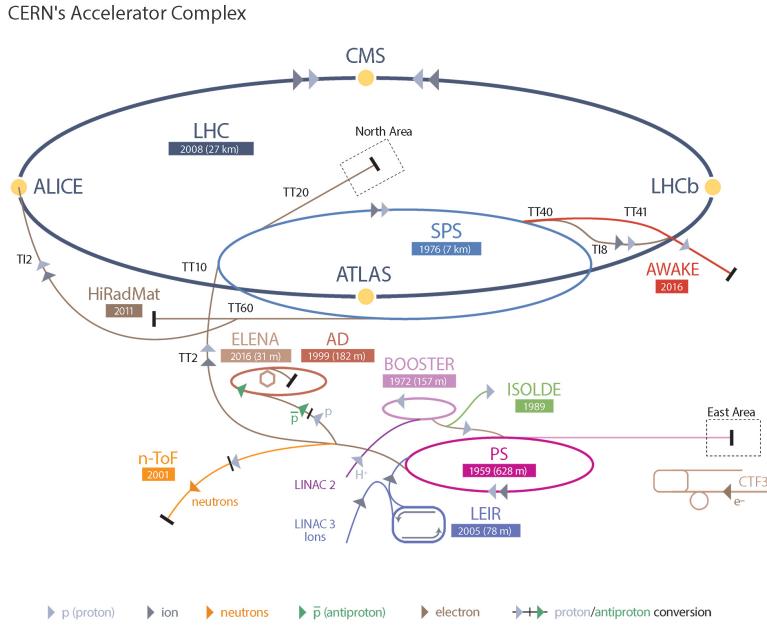
### **1.1.1 The history of the LHC**

The story of the LHC begins in 1977, when the CERN director general Sir John Adams suggested that the tunnel of the Large Electron-Positron Collider (LEP) can be reused to accommodate the future hadron collider of more than 3 TeV energies ??.

At the 1984 ECFA-CERN workshop on a "Large Hadron Collider in the LEP Tunnel" ??, the physics goals of the LHC were stated: confirmation of the BEH mechanism, search for the Higgs Boson, and exploration of the origin of masses of W and Z bosons.

The parameters of the proposed LHC were very ambitious: the centre-of-mass (COM) collision energy of 10 to 20 TeV, and a target instantaneous luminosity of  $10^{33-34} \frac{1}{cm^2 s}$ .

Large Hadron Collider (LHC) is the most powerful particle accelerator that has ever been built. It is located at the border of France and Switzerland at a depth from 50 to 175 m underground. LHC ring is 26.7 km in circumference and it is the final stage in a sequence of accelerators. We will discuss the whole sequence of accelerators in the following section.



**Figure 1.1:** Schematic layout of the LHC.

### 1.1.2 The layout of the LHC

It is a complex process to start proton-proton collision in the LHC at 13 TeV and, therefore, the process consists of several stages (see Fig. 1.1). Everything begins with the bottle of hydrogen. The hydrogen atoms from the bottle are fed into the source chamber of the Linear Accelerator (Linac). In the chamber the hydrogen is heated up to the plasma state until electrons are stripped off of the hydrogen atoms. Then electrons are removed and remaining protons are directed to the first acceleration stage which increases the energy of protons to 50 MeV. After Linac, the beam of

protons is injected into the Proton Synchrotron Booster (PSB). PSB contains four rings each accelerating a bunch of protons (a moving collection of protons of a narrow length) to 1.4 GeV. The third stage is the Proton Synchrotron (PS), which splits the incoming beam into 72 bunches separated by 7.5 m. The energy of the protons is increased to 25 GeV. After that, the protons are sent to the Super Proton Synchrotron (SPS), where they are accelerated to 450 GeV. SPS then fills the LHC ring with two beams each consisting of 2808 bunches of protons with nearly  $10^{11}$  protons in total. It takes SPS about  $O(10)$  minutes to fill each LHC ring with bunches. In the LHC two beams are circulating in opposite directions in two separate beam pipes. During standard data taking beams circulate for  $O(10)$  hours.

### 1.1.3 LHC operations

The first LHC budget plan was finalised in 1996 and the final cost was approved just a few years later. The first proton beam entered the LHC ring in 2008. However, an incident intervened the LHC plans. It was caused by the mechanical damage of the tunnel equipment due to the release of the helium. Thus, the real data taking period (called LHC Run-1) had started only in 2010, lasted for two years, and 7-8 TeV COM energies were used. The recorded dataset contained enough Higgs bosons to claim a discovery of this rarely produced particle. After this achievement, the LHC was closed for the first long shutdown (LS1) that happened in 2012. During this time necessary upgrades of the main detectors and the LHC were performed. This was an unavoidable and essential step to prepare the LHC for more challenging environment of COM energies increased to 13 TeV.

If we denote the area of  $10^{-28} \text{ m}^2$  as barn (b), with the femtobarn ( $fb$ ) equal to  $10^{-43} \text{ m}^2$ , then in terms of these new units the LHC can theoretically produce 80 –

$120/fb$  (inverse femtobarns) of data a year. In practice numbers were lower, because LHC operated at the revolution frequency below the nominal, used fewer proton bunches in the beam, etc. All this resulted in lower than expected instantaneous luminosity, which is a very important term in collider physics and will be explained in the next section.

The LHC Run-2 has started in 2015 and the CMS collected  $4.2\ fb^{-1}$  of data that year. Over the course of the 2016 data taking, an integrated luminosity of  $35.9\ fb^{-1}$  was recorded. This luminosity is the amount of data that has been collected by the CMS detector and later approved by the CMS physics coordination for the use in the physics analyses. The data set of proton-proton collisions collected in 2016 at 13 TeV COM energy is used in this thesis to analyse double Higgs boson decays. Together with the 2017 and 2018 data taking, almost  $150\ fb^{-1}$  have been delivered and recorded by the CMS detector during the whole Run-2 period of four years.

At the moment of writing this thesis, the LHC has entered the LS2. The next data taking will resume in 2020 and proton-proton collisions will continue for three years with the expected delivered integrated luminosity equal to nearly  $300\ fb^{-1}$ . This will conclude the LHC Phase-1 programme.

The new upgraded LHC, the High-Luminosity LHC (LHC) or the Phase-2, will start operations in 2026 and run until 2035. The COM energy will be increased to 14 TeV and one expects to record an unprecedented dataset of  $3000\ fb^{-1}$ .

#### 1.1.4 Luminosity

The instantaneous luminosity  $\mathcal{L}$  is the coefficient which relates the cross section  $\sigma$  of the process to the number of events  $N_{events}$  produced during the interaction:  $N_{events} = \mathcal{L}\sigma$ . Luminosity is the parameter controlled by the machine and can be written as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev}}{4\pi\sigma_x\sigma_y}$$

where  $N_b$  is the number of particle in the colliding bunch,  $n_b$  is the number of colliding bunches in the beam,  $f_{rev}$  is the revolution frequency of the beam,  $\sigma_x$  and  $\sigma_y$  are the standard deviations of the beam density profile (BDP) in the transverse plane, where it is assumed that the BDP of both beams can be described by a Gaussian distribution.

To maximise the amount of collected data, the luminosity parameter should be as high as possible. It is worth noting that the luminosity is not constant and decays with time due to the degradation of the initial circulating beams. Theoretical decay time (the time to reach  $1/e$  level) is approximately 29 h. In practice, taking into account the decrease of protons in the bunch due to collisions, contributions from the intrabeam scattering, scattering on the residual gas, etc., the real luminosity lifetime is about 15 h.

A useful variation of the luminosity parameter is a total integrated luminosity. This is the number normally quoted for the dataset collected over the period T:

$$L = \int_0^T \mathcal{L} dt.$$

In collider physics the "beam dump" is a process of burning off exhausted low luminosity beams by intentionally directing them towards the target made of concrete and steel. The time from the start of the collisions to the beam dump is usually called the "run".

We can calculate the amount of data delivered by the LHC during a single run period  $O(10)$  h. Performing the integration, we obtain:

$$L = \mathcal{L}_0 \tau_{\mathcal{L}} \left[ 1 - e^{-\frac{\tau_{run}}{\tau_{\mathcal{L}}}} \right],$$

where  $\mathcal{L}_0$  is the initial peak instantaneous luminosity at the start of the run,  $\tau_{run}$  is the total duration of a run, and  $\tau_{\mathcal{L}}$  is the luminosity lifetime. The optimum run time is 12 hours. During the runs, the LHC centre needs to dump the old beams, fill the rings with the new beams, and increase ("ramp") the energy of new beams to 13 TeV.

After that a new run can be started. This restarting process normally takes two to six hours.

### 1.1.5 LHC infrastructure

The equipment of the LHC tunnel serves several purposes with the main objective to keep the colliding beams on the circular orbit. This requires a complex synchronised work of bending dipole magnets, cooling systems, accelerating radio frequency cavities, and vacuum insulation systems.

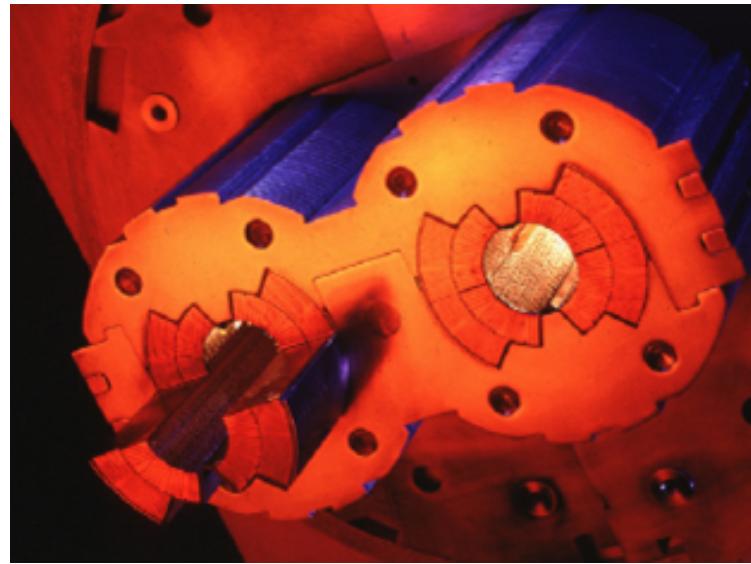
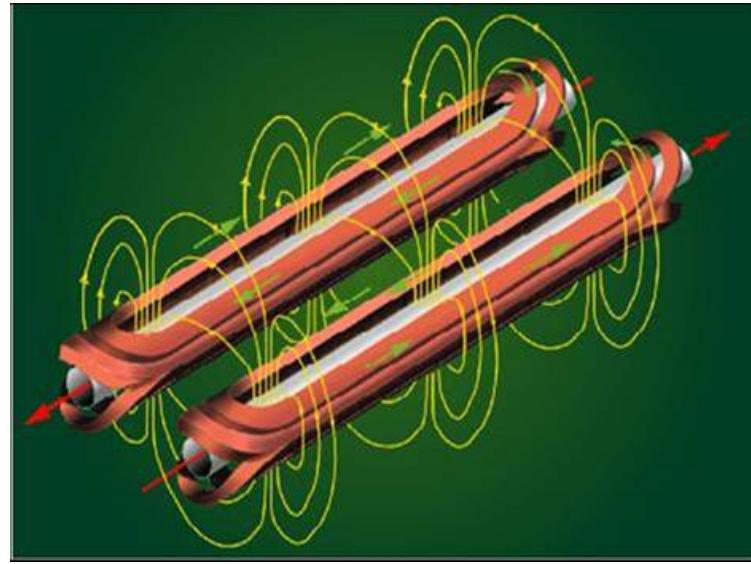
#### 1.1.5.1 Magnets

Most of the LHC circumference is used by 1232 superconducting magnets placed evenly around the tunnel to approximate the circular orbit. These are dipole magnets (see Fig. 1.2) that bend the beam and keep it on the circular orbit, that is why they are commonly called "Main Bends" (MB). The proven technology existed since Tevatron and relied on NbTi superconductors. This technology also satisfied the LHC cost and performance requirements, thus, it was decided to reuse the same choice of the alloy for the LHC superconducting dipole magnets that steer the proton beams.

The dipoles need to produce the magnetic field of 8.3T. Each dipole is 16.5 *m* (with ancillaries) long and 570 *mm* in diameter and is placed inside of the dipole cryostat which is called the "Helium bath".

This cryostat is a long cylindrical tube 914 *mm* in diameter made of low-carbon steel, where the dipole mass is cooled down to 1.9 *K*. Even though the inner structure of such cryostat is very complex and includes two beam pipes, two sets of coils for two beam pipes, vacuum pipes etc., one normally calls this compound object simply a dipole magnet. The name "dipole" is reserved for MBs since for each beam pipe

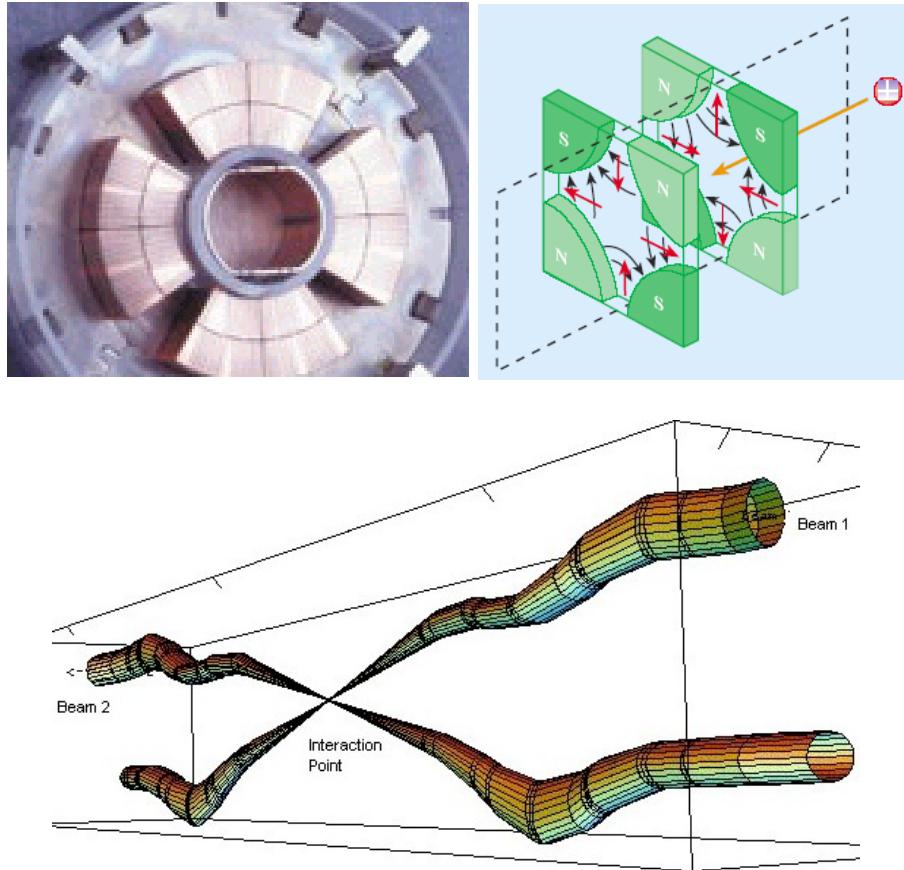
the magnet consist of two "poles" that provide a vertical magnetic field similarly to a simple dipole system of magnets.



**Figure 1.2:** LHC dipole magnets. Top: two dipole coils and magnetic field lines. Bottom: two beam pipes with the coils inside of the dipole magnet.

A dipole magnet must be curved to help a chain of dipoles complete 360 degrees. The curvature is 5.1 *mrad* per dipole, which is equivalent to a sagitta of about 9 mm, corresponding to a radius of curvature of 2812.36 m.

The other important set of magnets is quadrupoles. They are used to ensure the proper beam dynamics. In total 392 quadrupole magnets ranging from 5 to 7 metres in length are used to squeeze the beam in transverse direction and to keep it narrow during the run duration. Additional special quadrupole magnets (SQM) are installed right before the IPs to focus the beams even more. That increases the density of protons in the beam and guarantees the maximum luminosity. In addition, SQMs help to decrease the chance of the parasitic collisions when bunches from the same beam or bunches outside of the IP centre interact (see Fig. 1.3). To further correct the beam path (orbit), about 5000 higher order correcting magnets are used, which are evenly spaced around the circular trajectory of the LHC.



**Figure 1.3:** LHC quadrupoles. Top left: the coil of the quadrupole magnet. Top right: schematic view of the magnetic fields in the quadrupole. Bottom: two beams and the IP.

To power the LHC, 1612 electrical circuits are used. Mostly these circuits are needed to power the dipole and quadrupole magnets, which is done in eight evenly spaced location of the LHC. A total of 3286 current leads are needed to connect all the circuits and power cables. More than a thousand of the leads operate between 600 A and 13 kA (see Fig. 1.4). The other leads operate in the range 60 to 120 A.



**Figure 1.4:** 13 kA high-temperature superconducting current lead.

### 1.1.5.2 Cooling System

To ensure that dipoles are in the superconducting state, they have to be cooled to 1.9 K using superfluid helium-4.

The cooling (cryogen) system is needed to keep superconducting LHC magnets at the appropriate temperature. The choice of the cooling gas depends on the magnet type and location. This dictates the required range of temperatures, which differs from system to system by 75 K. The cryogen system uses layered design with the

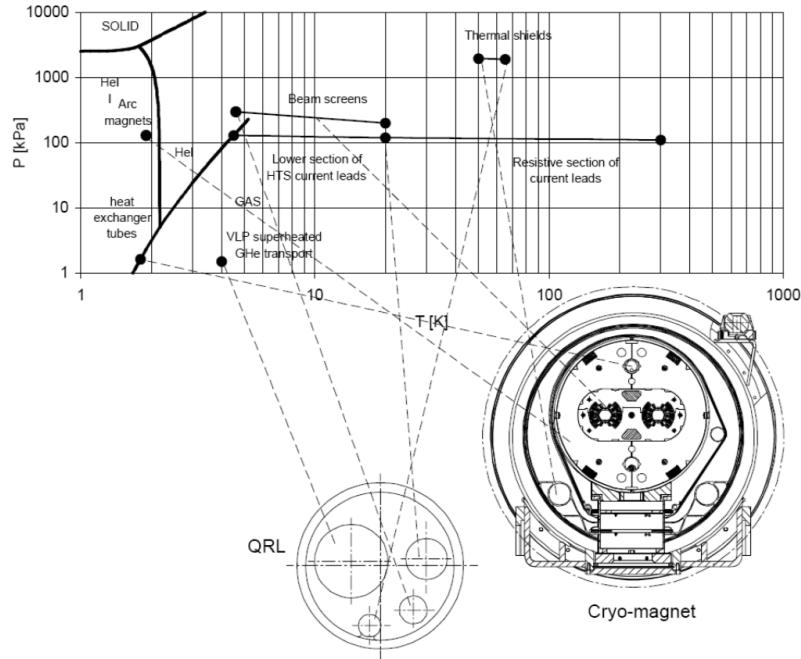
temperature becoming progressively colder going from outside the dipoles closer to the beam pipe.

The "coldest" part of the cryogen system is designed for the inner part of the dipoles. This system (see Fig. 1.5) must cool down 37 Mkg of the LHC magnets within 15 days to the required temperatures, which is done through the system of pipes that transports and directs the flow of the superfluid helium. The cryogen system must also be able to deal with the fast increases of the pressure flow and flow surges, as it is crucial for the LHC operation to keep dipoles constantly cooled and at the superconducting state.

The LHC tunnel is inclined in the horizontal plane by  $1.41^\circ$ . This translates to 120 m difference in the vertical location of two diametrically opposite points of the tunnel with respect to the surface level; and results in the additional hydrostatic pressure that can affect the flow of helium. This has been an important concern during the design of the cryogen system.

Since the cost to cool the LHC equipment to 1.8-1.9 K temperatures is high, several temperature levels are employed (see Fig. 1.5):

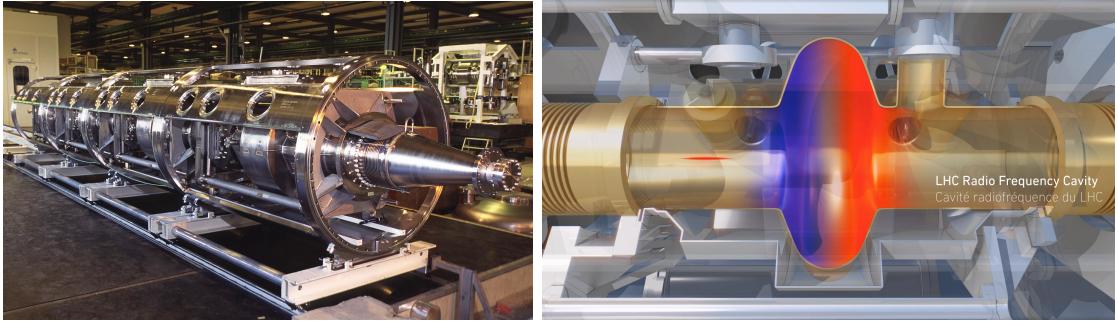
- 50 to 75 K for the thermal shielding used in the dipoles,
- 20 to 300 K for upper ("warm") sections of the high-temperature superconducting current leads,
- 4.6 to 20 K for lower temperature interception,
- 4.5 K for radio frequency cavities and lower ("cold") sections of the high-temperature superconducting current leads,
- 4 K for the transportation system that directs the 1.8 K helium to dipoles,
- 1.9 K for helium in the superfluid state to cool magnet masses.



**Figure 1.5:** LHC cryogenic states and the temperature scale.

### 1.1.5.3 Radio Frequency Cavities

Proton bunches need to be ramped to 7.5 TeV energies. To achieve this 13 TeV COM energy, eight superconducting radio-frequency cavities (RFC) are used per beam. They are located in front of the IPs of four experiments. Electromagnetic waves of 400 MHz with a peak field strength of 5.5 MV/m adjust the speed of protons in bunches. Each RFC (see Fig. 1.6) increases the energy of protons by 60 keV per revolution and it takes  $O(20)$  minutes to reach 6.5 TeV beam energy. The RFC frequencies are increased gradually by 1 kHz to match the speed up of protons in the bunch as they gain more energy. When the ramp is completed, the RFCs are used to compensate for small energy losses due to the synchrotron radiation (7 keV per revolution).



**Figure 1.6:** LHC RF cavities. Left: a cryomodule with four RF cavities. Right: a schematic drawing of a single RF cavity. The colour field is used to denote positive (red) and negative (blue) polarities. A narrow beam traversing the cavity is coming from the left and is shown in red.

#### 1.1.5.4 Vacuum System

The work of the LHC depends on three vacuum systems [1]. Without them, dipoles will not be at the superfluid state, the beams will not be able to circulate, and no stable collisions would be taken. With a total of 104 kilometres of vacuum pipes, the LHC owns the largest vacuum system in the world. The main types of vacuum systems are:

- insulation vacuum for cryomagnets,
- insulation vacuum for the helium distribution line,
- beam vacuum.

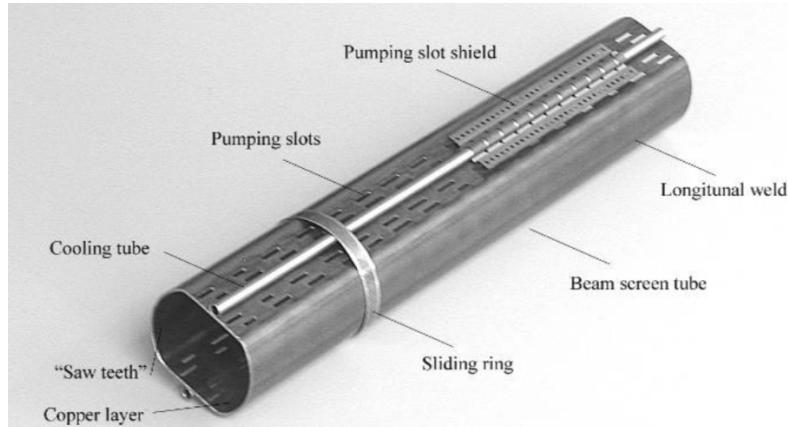
The insulation vacuum is needed to ensure the operations at both low temperatures of the magnets and the room temperatures in the tunnel. The insulation vacuum of  $10^{-6}$  mbar is used for a total of 15000 cubic metres. To build this vacuum system, the LHC used 250,000 welded joints and 18,000 vacuum seals.

The vacuum for the helium distribution lines is needed to protect from the heat the flow of the helium-4. This helium flow is used to cool down the dipole mass.

Cryogenic distribution lines (QRL) of 3.3 km each are connected to eight cryogenic plants that pump the helium-4 into the LHC. The vacuum in these systems is at  $10^{-7} - 10^{-10}$  mbar level.

For the beam pipes the LHC uses ultra-high vacuum of  $10^{-10}$  mbar at cryogenic temperature of 5 K. The vacuum is getting progressively closer to  $10^{-11}$  mbar near the IPs, because in these locations collisions take place and any additional gas is highly undesirable. This vacuum is the emptiest space in the Solar System. This ultra-high vacuum is needed to reduce the beam degradation due to the beam-gas interactions in the pipe and parasitic collisions of bunches with the collimators near the IPs.

Vacuum system are affected by the heat produced from the synchrotron radiation emitted by the proton beams when they are bent. To reduce the amount of this heat and to narrow down the beam size in the transverse direction when the beam widens, the LHC uses "beam screens", which operate between 5 and 20 K.



**Figure 1.7:** Beam screen.

The beam screens are necessary to reduce the number of protons scattering on the residual gas of the beam pipes, which could lead to a magnet quench and even interrupt the machine operation.

The table below summarises the main heat sources that degrade the vacuum quality in the beam pipe, where the vacuum must exist at 1.9 K:

- synchrotron radiation ( $0.2 \text{ W/m}$  per beam),
- energy loss by nuclear scattering ( $30 \text{ mW/m}$  per beam),
- image currents ( $0.2 \text{ W/m}$  per beam),
- electron cloud related effects (vary).

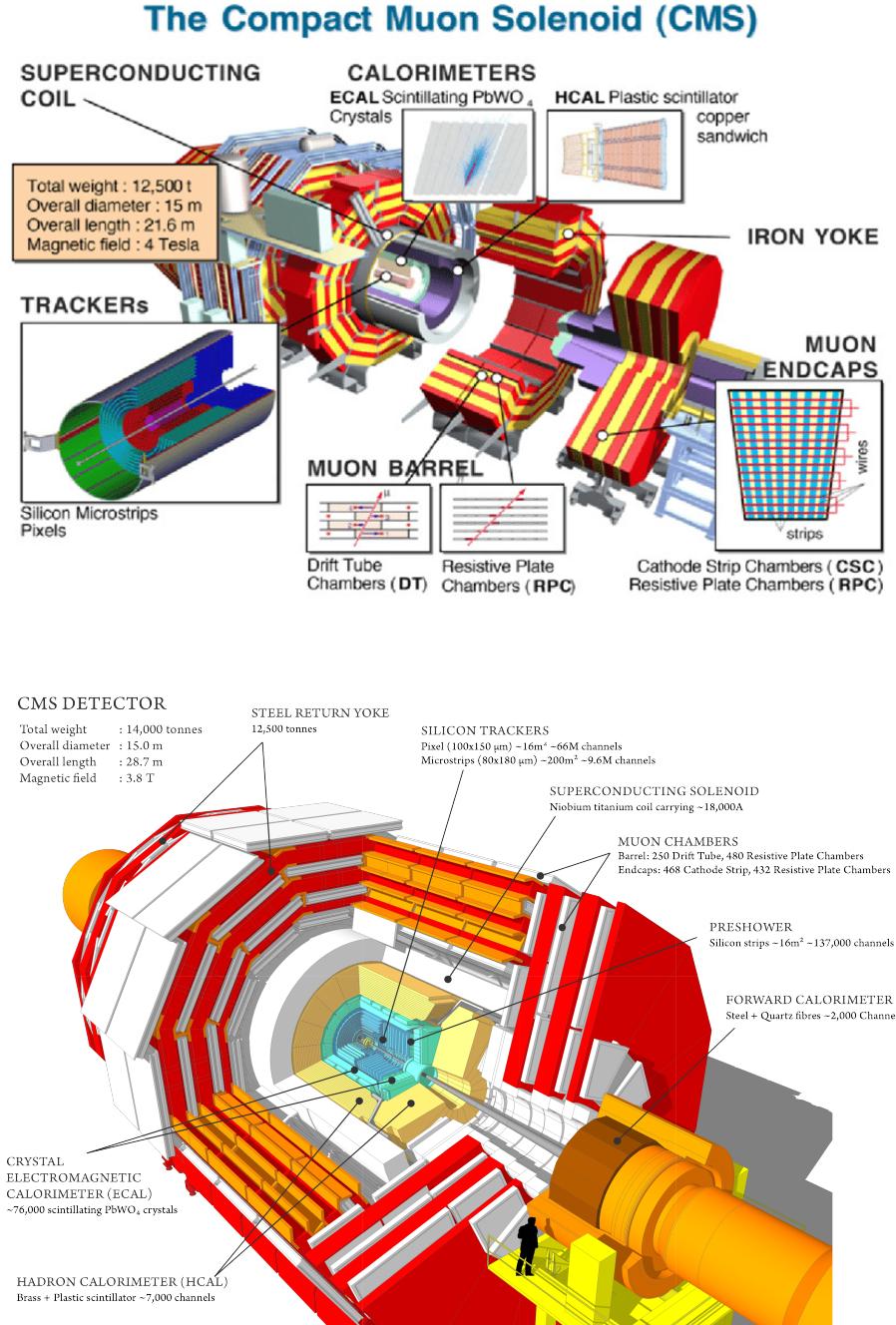
Now, that we discussed the LHC collider, we can continue with one of the main LHC detectors - the CMS detector - the one that was used to collect the data analysed in this thesis.

## 1.2 The CMS experiment

The Compact Muon Solenoid (CMS) is a multi-purpose particle detector built to study a variety of complex particle interactions produced by the LHC. CMS is located in the underground cavern at the "Point 5", which is one of the four main IPs of the LHC. The CMS detector with the additional computing infrastructure is able to detect the produced particles, measure their main physics parameters, and to send the related data to computing data centres for persistent storage.

The CMS detector has a cylindrical shape and consists of a central ("barrel") and two forward ("endcaps") sections (see Fig. 1.8). CMS is the heaviest detector ever built with the mass of nearly 12500 tons. The mass is explained by the amount of the used superconducting metal, which serves as the magnet. The CMS is 21.6 m long and 14.6 m high. The CMS has an onion-like structure of concentric layer of

detectors around the IP. In addition, at the outer part it has a large superconducting solenoid to produce inside the detector a homogeneous magnetic field of 3.8 T.



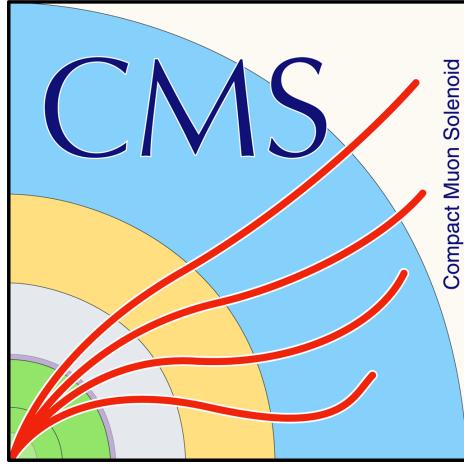
**Figure 1.8:** CMS experiment with the main sub-detectors.

All sub-detectors can be categorised into trackers and calorimeters [2]. As the

particle passes through the material of the tracker, it leaves a "track", which is a path of the emerging particle. Trackers focus on the direction and the track curvature of the charged particles. Tracking information allows the determination of the particle's momentum.

There are two trackers in CMS: an inner tracking system that encloses the IP and the outer tracking system that is located outside of the solenoid magnet. The first system contains the Pixel and the Strip trackers. The second tracking system is dedicated for the muon detection and is usually called a muon tracker or a muon system. This system is embedded within a steel yoke of the magnet.

The magnet yoke is made of five barrel wheels. Such an arrangement saves the CMS some space and also is used for the magnetic flux return. Additionally, it serves as a support for the embedded muon system, which is located outside of the ECAL and HCAL systems. Muons are energetic enough to traverse the ECAL and leave the detector. This muon system-magnet yoke structure provides a return field of the magnet of about 2 T and is used to measure the momentum of muons. This "two-directional" magnetic field with respect to the magnetic yoke, causes the muons trajectories to be bent in opposite directions in the inner tracker in contrast to the outer tracker. This important feature of the CMS detector is depicted in the CMS logo (see Fig. 1.9).



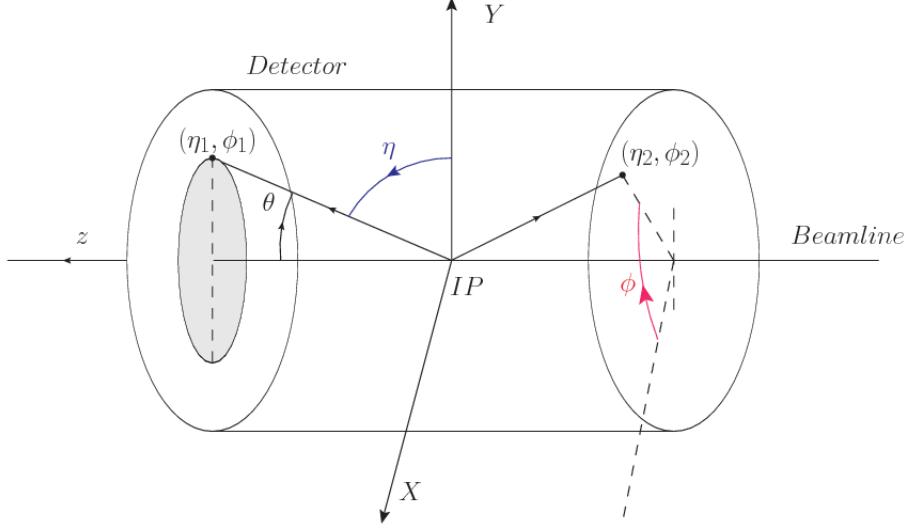
**Figure 1.9:** The logo of the CMS experiment that is showing curved trajectories of the emerging muons.

The CMS has two calorimeters: the electromagnetic and the hadronic calorimeters. They both rely on high density materials either to sample or to contain almost all the energy of the incoming particles with their secondary interaction products. However, these two systems focus on two different sets of particles. As will be discussed later, electromagnetic calorimeter (ECAL) is dedicated to measuring the energy of photons and electrons, while the hadronic calorimeter is targeting the measurement of the energy of hadrons. 1! The rate of the incoming data at the LHC is 40 MHz. This corresponds to almost 70 TB produced every second! It is impossible to store that much data, and, most importantly, most of the information in this data is not interesting for future physics analyses. To reduce the data rate, the CMS uses a highly efficient system of triggers. The first one, the Level-1 (L1) trigger, reduces the nominal collision rate of 40 MHz to 100 kHz. The subsequent High-Level Trigger (HLT) further decreases the rate to 1 kHz. With the help of the trigger system, the original 40 TB per second rate is transformed into manageable 1 GB per second that is stored for offline analysis use.

### 1.2.1 The CMS coordinate system

The CMS uses a right-handed Cartesian coordinate system to define the axes of the colliding beams (see Fig. 1.10). The centre is located at the IP and the x axis points to the centre of the LHC ring. The y axis points upwards, and the z axis points along the proton beam direction. Since the CMS detector has a cylindrical shape, the polar system is used in the x-y plane: a standard set of the azimuthal angle  $\varphi$  and the radial coordinate  $r$ . The polar angle  $\theta$  is defined in the r-z plane and a widely used in this thesis angular variable  $\eta$  (called pseudorapidity) is defined as  $\eta = \ln \tan(\theta/2) = \ln(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z})$ . Additionally, a popular quantity in the collider physics - the rapidity - is given by  $y = 1/2 \ln(\frac{E + p_z}{E - p_z})$ . Rapidity is a function of the energy E and longitudinal momentum  $p_z$  of the particle (the projection of  $\vec{p}$  on the z axis). Note that  $\eta$  converges to  $y$  when the mass is negligible and the particle travels with the speed close to the speed of light. Most angular variables that are used currently in the modern high-energy physics (HEP) are defined in terms of  $\eta$  and  $\varphi$ :  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2}$ , with  $\Delta\eta$  and  $\Delta\varphi$  being the absolute values of the relative differences of  $\eta$ 's and  $\varphi$ 's of two particles.

Another extremely useful quantity is the projection of the momentum of a particle on the transverse plane and is called "transverse momentum"  $p_T$ . This variation of the momentum is independent of the z axis, hence, from the Lorentz boost. Similarly, the "transverse energy" of a particle is defined as  $E_T = \sqrt{m + p_T}$ .



**Figure 1.10:** Coordinate system of the CMS detector ???. Two particles (1 and 2) are shown with the corresponding angular variables  $(\Delta\eta_1, \Delta\varphi_1)$  for the first and  $(\Delta\eta_2, \Delta\varphi_2)$  for the second particle respectively.

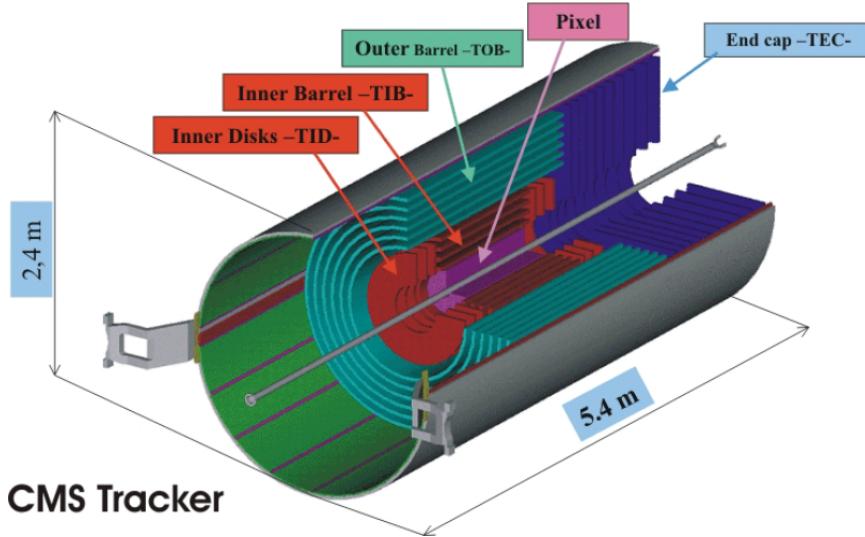
### 1.2.2 The Inner Tracker

The inner tracker [3] (see Fig. 1.11) is the closest subdetector to the IP. Using the tracker the experiment measures the trajectories of charged particles and reconstructs decay vertices. Since this system is constantly under the radiation coming from the interactions with the particle flux of nearly 100 MHz/cm at  $r = 4$  cm, the design of the tracker focused on two main requirements: high granularity for precise determination of the vertices and tracks, and robustness against the radiation-hard environment with the operational time of at least 10 years. As a solution to both challenges, the CMS relies on the silicon technology that provides the tracker with the large surface of thin but highly granular active detectors. The tracking system has a diameter of 2.4 m and a length of 5.4 m covering the detector space of  $|\eta| < 2.5$ .

The inner most part of the tracker - the Pixel detector ("Pixel")- consists of three layers in the barrel at the radii of 4.4 cm, 7.3 cm, and 10.2 cm respectively. The Pixel

also has two detector disks in forward regions. They are positioned 34.5 and 46.6 cm away from the IP. The Pixel is made of 1440 modules which contain 66 million pixel cells. Each cell is 100 by 150  $\mu\text{m}$  with 285  $\mu\text{m}$  thickness, which allows the determination of "hit" positions (the passage of the particle through the Pixel cells) in two directions z- $\varphi$  in the barrel and r- $\varphi$  in the endcaps.

The spatial resolution of each pixel about 10  $\mu\text{m}$  in the r- $\varphi$  plane and 20  $\mu\text{m}$  along the z direction. The spatial information that comes from the tracker is used to determine the main interaction point of the hard scattering ("the primary vertex") and also additional interaction vertices ("pileup"). Tracker also helps to reconstruct the displaced vertices ("the secondary vertices") of the particles that decay relatively fast, e.g., b-jets, which will be discussed later in this chapter.



**Figure 1.11:** The inner tracker. Pixel and Strip detectors are shown.

The outer part of the inner tracker is the strip tracker. It contains several subsystems and is made of almost 9.3 million strips arranged in different configurations in 15148 modules. The first subsystem is the tracker inner barrel (TIB), which consists of the four barrel layers of strip modules. The second subsystem is the tracker inner

disks (TIDs), which is made of three disks of strip modules. Increasing the radius to about 60 cm, the tracker outer barrel (TOB) starts. TOB is made of six layers of strips. Finally, to cover high  $\eta$  regions, the tracker endcaps (TECs) are used, which are made of two sets of nine disks of strips.

Each strip is about  $O(20)$  cm long. Its thickness varies from 320  $\mu\text{m}$  for TIB and TID, to 320  $\mu\text{m}$  - 500  $\mu\text{m}$  for TOB and TEC, respectively. Also width changes from 80  $\mu\text{m}$  - 141  $\mu\text{m}$  for TIB and TID, to 97  $\mu\text{m}$  - 184  $\mu\text{m}$  for TOB and TEC, correspondingly. The resolution on the single point in the radial direction is 20 - 50  $\mu\text{m}$ , and in the z direction it varries from 200 to 500  $\mu\text{m}$ , depending on the value of  $r$ .

All subsystems of the inner tracker have to be cooled down to about  $-20^\circ$ . This requirement is needed to minimise the damage of the tracker caused by the radiation from the collisions and to reduce overheating of the electronics.

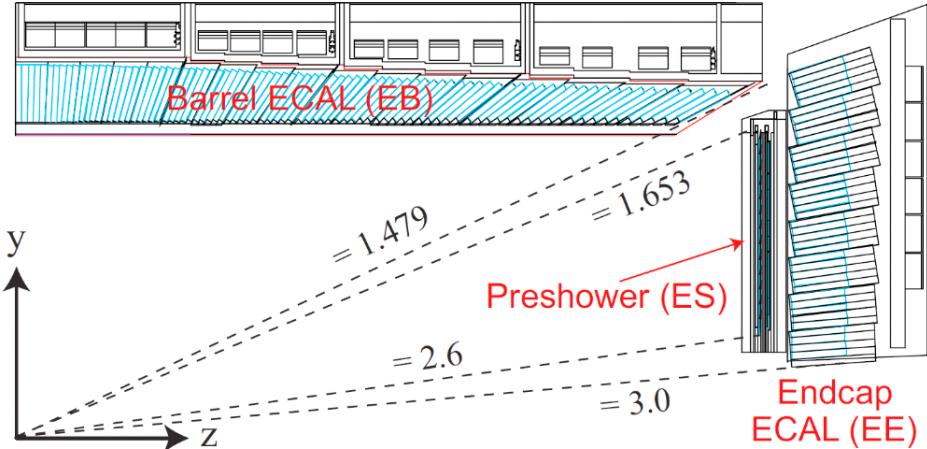
The material of the inner tracker has 0.4 to 1.8 radiation lengths ( $X_0$ ), which corresponds to 0.1 to 0.5 nuclear interaction lengths ( $\lambda_i$ ). Numbers vary with the  $\eta$ .

### 1.2.3 The ECAL

The inner tracker and the ECAL provide the detector with complementary measurements. The tracker focuses on the direction and the momentum of the particle and identifies only charged particles. The ECAL [4] (see Fig. 1.12), on the other hand, determines the energy of the particles and detects all particles that interact electromagnetically, including photons and neutral pions. However, primarily the ECAL is designed to measure precisely the energy of electrons and photons.

The ECAL is a highly granular detector that relies on the lead tungstate crystal ( $\text{PbWO}_4$ ) technology. Electrons and photon passing through the crystal interact with

its material and their energy is converted into the produced electromagnetic "shower". PbWO<sub>4</sub> crystals are known for being a popular choice of the scintillators: interactions with the crystal material produce the scintillation light that is further read out by the electronics. The PbWO<sub>4</sub> crystals have a high density (8.28g/cm<sup>3</sup>), a small radiation length ( $X_0 = 0.89$  cm), a short Moliere radius ( $R = 2.2$  cm), and a fast response (80% of its scintillation light is produced within 25 ns). These characteristics are making PbWO<sub>4</sub> crystals ideal candidates for the ECAL, since they guarantee an excellent containment of the electromagnetic shower within the crystals.



**Figure 1.12:** The ECAL and the Preshower detectors.

The ECAL has a barrel part (EB), covering the  $|\eta| < 1.479$ , and two endcaps (EE) covering  $1.479 < |\eta| < 3.0$ . In the barrel ECAL is made of 61 200 crystals. Each crystal is 22 by 22 mm with a length of 23 cm. In the endcaps ECAL has 7324 crystals. There each crystal is 28.62 by 28.62 mm with a length of 22 cm. The crystals' layout is following a quasi-geometric projection with axes of crystals slightly tilted to ensure particle trajectories are never aligned with the intercrystal cracks. This layout is optimised for the best particle shower containment with respect to the position of the interaction point.

The resolution of the ECAL is a function of energy of the incident particle E and can be decomposed into three terms. The first term is a stochastic term that is inversely proportional to the square root of the number N of scintillation photons produced in the interaction. In the main formula N is replaced by E, since N is proportional to E. The second term is a "noise" term that describes the noise in the detector. The third term is related to detector imperfections and is represented by a constant C. The final dependence of the ECAL energy resolution  $\sigma$  on the particle energy E is given by:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (1.1)$$

From the dedicated calibration studies, the parameters in the formula above are found to be equal to: S = 2.8%, N = 12%, and C = 0.3%. As a "standard" procedure, the CMS often optimises the performance of the subdetectors for 45 GeV electrons, since they correspond to a classical Drell-Yan decay of Z boson to two electrons. In this case, a typical energy resolution for 45 GeV electrons is about 2% in EB and 2-5% for EE. Near the Z peak (91 GeV), the constant terms dominates the resolution.

The ECAL is operated at a temperature of 18 ° C and the "active width" of the ECAL material corresponds to 25 X<sub>0</sub>.

An additional subdetector, called the "Preshower", is installed right in front of the EE and covers  $1.653 < |\eta| < 2.6$ . The Preshower is designed to improve the discrimination of single photons from diphoton decays of neutral pions  $\pi^0 \rightarrow \gamma\gamma$ . This is a sampling calorimeter in which the material that produces the particle shower is distinct from the material that measures the deposited energy. Typically the two materials alternate. The Preshower has two lead layers which launch the electromagnetic showers. This "samples" the energy of the particles traversing the Preshower material.

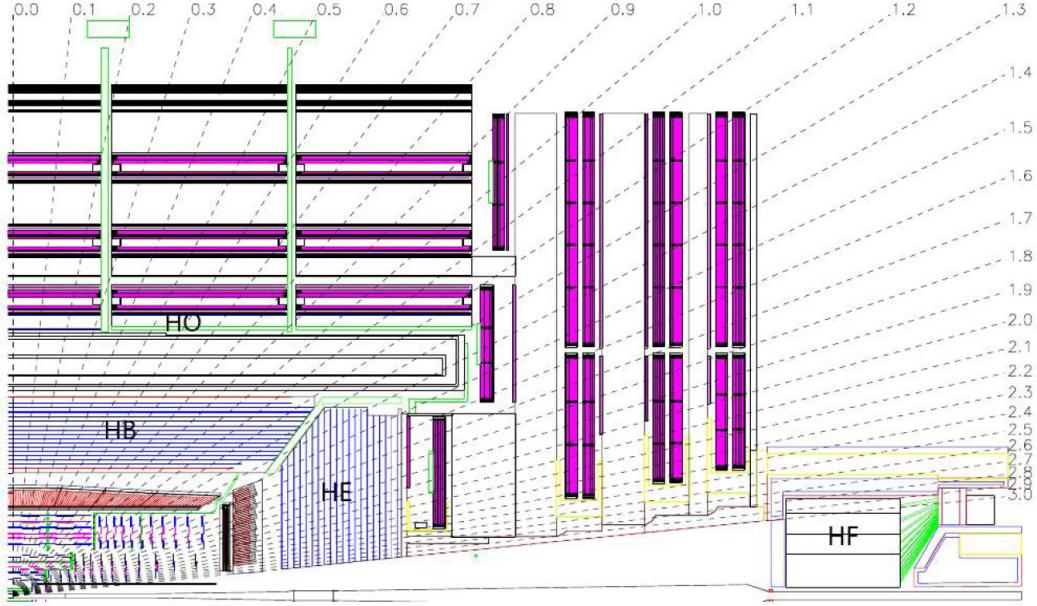
After these layers, 2 mm-wide silicon strips are placed. They measure the deposited energy and transverse profile of the shower shape initiated by the lead layers. The "thickness" of the Preshower material corresponds to  $3 X_0$ .

#### 1.2.4 The HCAL

Hadrons normally go through the ECAL layers without being stopped. To absorb these particles, the HCAL [5] (see Fig. 1.13) is placed around the ECAL. The HCAL focuses on particles that hadronise. This is a process of the formation of hadrons out of quarks and gluons. The HCAL detects with the charged and neutral hadrons such as pions, kaons, protons, and neutrons. Hadrons also produce collimated streams of secondary particles (jets) and these jets are identified by the HCAL. Additionally, the HCAL is used to measure indirectly the transverse energy of neutrinos, by the momentum imbalance technique, which will be discussed later in this chapter.

The HCAL is split into HCAL barrel (HB) and HCAL endcap (HE) sections. They cover  $|\eta| < 1.3$  and  $1.3 < |\eta| < 3.0$  respectively. HB and HE are sampling calorimeters. They are made of a brass absorber and of active plastic scintillating tiles. The brass plates in HB have thickness of 56.5 mm and in HE the thickness is increased to 79 mm. The absorber material corresponds to  $5.82 \lambda_I$  at  $\eta = 0$  to almost  $10 \lambda_I$  at  $|\eta| < 1.3$ .

The gaps in the absorber of the HCAL are filled with an active medium of 70000 plastic scintillator tiles. The scintillation light is guided by wavelength shifting fibres (WLSs) to hybrid photodiodes (HPDs). The scintillator is quite fast with the 68 % of the light been collected within 25 ns.



**Figure 1.13:** The HCAL with the  $\eta$  coverage map.

The CMS also has an outer calorimeter (HO) placed above the HB outside the solenoid. HO is called a tail catcher system and increases the total calorimeter thickness to  $11.8 \lambda_I$  in the barrel, with the magnet coil working as an extra absorption layer. The HO consists of five rings of scintillator tiles. A supplementary iron plate of 19.5 cm in thickness and a second layer of sensitive material are placed around  $\eta = 0$  to enhance the absorber depth there.

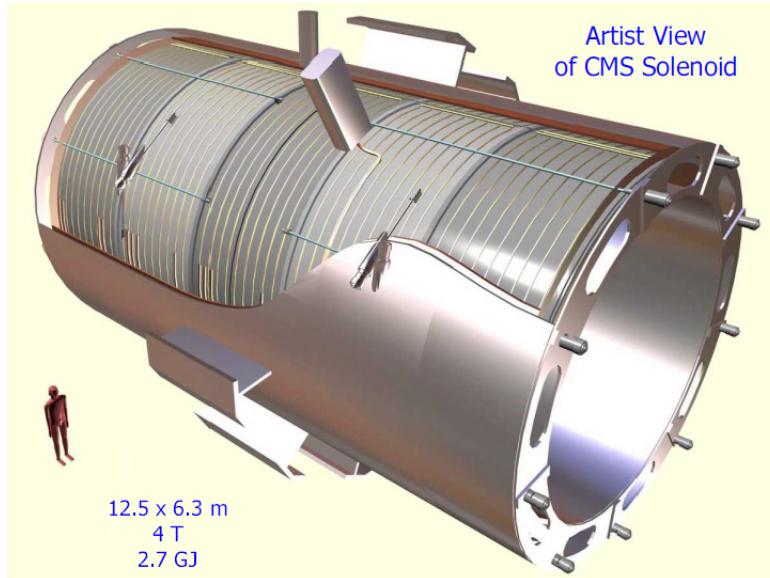
In the forward directions, two forward calorimeters (HF) extend the coverage to  $|\eta| = 5.2$ . The HF is composed of steel absorbers and quartz fibres that produce Cherenkov light when the particle in the material travels faster than the light in that medium. The light is further collected by photomultiplier tubes (PMTs).

Since the HCAL is located between the ECAL and the internal surface of the solenoid, the space allocated for the HCAL is not enough for the HCAL to fully absorb the hadronic showers and this imperfect containment of the hadronic shower limits the performance of the HCAL. Comparing with the formula 1.1, for the single

pions the values are given by: by S = 115 %, N = 52 %, and C = 5.5 % [6].

### 1.2.5 The Superconducting Solenoid

The NbTi superconducting solenoid (see Fig. 1.14) of 6 m in diameter is the core of the CMS experiment. The magnet operates at a temperature of 4.5K. The bulk of the CMS detector weight (90 %) comes from the magnet steel return yoke and structural supports which together weigh 12500 tonnes.



**Figure 1.14:** The CMS superconducting solenoid. The person on the left is shown to emphasise the size of the magnet.

The solenoid is central part in the CMS detector design. The idea was to have a uniform magnetic field capable of bending the trajectories of charged particles as they traverse the detector. When a low energy particle is produced, it has a helical path and will be fully contained within the detector. On the other hand, when a highly energised particle is produced, the trajectory is seen as a "straight" incomplete arc. Both situations lead to imperfect measurement of the momentum. The primary measurements of the tracking system are presumed to be Gaussian distributed; but

the momentum of the particle that the tracker measures is not Gaussian distributed. However, the sagitta is Gaussian distributed, and that is why widely used in the particle physics.

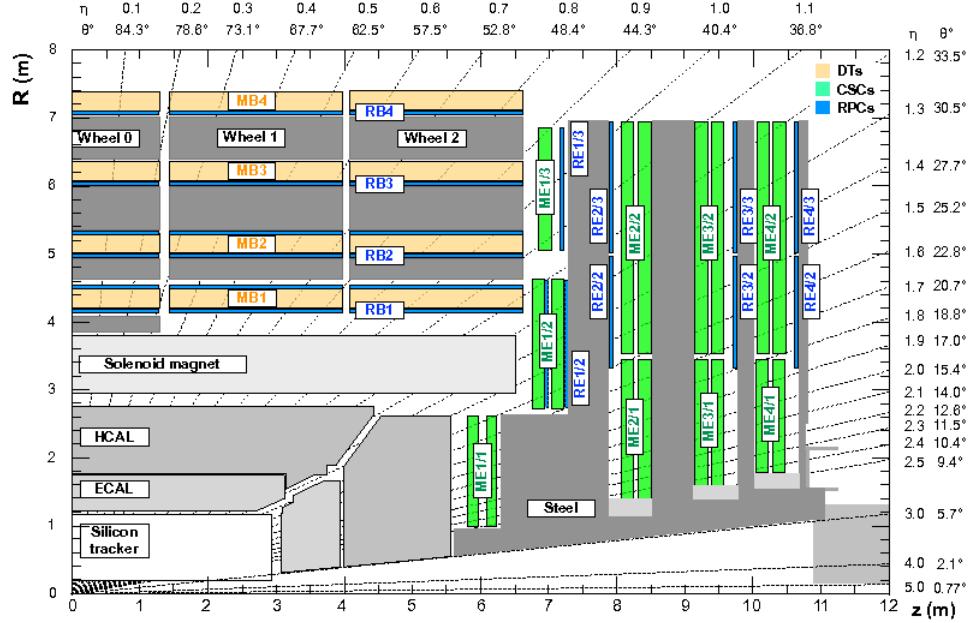
When the particle in the magnetic field passes thorough the material of the detector, the path deviates from the ideal circular line due to random fluctuations and multiple scattering. The sagitta term is used to quantify the depth of the circular arc and is equal to the distance from the centre of the arc to the centre of its base. Since the sagitta is following a Gaussian distribution, it may be approximated by simpler expressions in many calculations of the momentum resolution.

The magnetic field strength  $B$  and the length of the track  $L$  are dictated by the design of the detector. Since the momentum resolution is given by  $\sigma_p/p^2 \approx \sigma_x/BL^2$  (see [2]) and improves linearly with magnetic field  $B$ , this is the reason why the CMS decided to invest much of the detector space and budget in the magnet. For a track of the length of  $O(1)$  m in the magnetic field of  $O(3)$  T, the sagitta is equal to 1 mm, which can be measured very precisely.

### 1.2.6 The Muon Tracker

Many physics analyses in the CMS rely on a precise measurements of the muons in the detector. Although muons are detected by the inner tracker, that information cannot be used by the trigger (will be discussed in the following subsection). Therefore, CMS has an outer tracker or muon tracker [7] (see Fig. 1.15) located outside the calorimeters and the solenoid. Because of the typical muon energies, muons produced in collisions at the LHC traverse the detector material with the minimal energy losses. To measure energy of muons, the CMS uses the muon tracker, which relies on various gaseous detector technologies. The muon tracker is inserted into the gaps of the flux-

return yoke. Tracks in the muon system are used to reconstruct standalone muons, and in combination with the inner tracker, to reconstruct the global muons.



**Figure 1.15:** The CMS muon tracker. DT, CSC, and RPC detectors are shown in yellow, green, and blue respectively.

CMS muon system has three subdetectors: the drift tubes (DTs), the cathode strip chambers detectors (CSCs), and the resistive plate chambers (RPCs). In the barrel region, the CMS is equipped with the DT system, which is 250 drift tubes arranged into five barrel section ("wheels"). Each wheel is made of four concentric rings of DT stations. The working elements of the DT system - cylindrical cells with the rectangular base of 4.2 by 1.3 cm<sup>2</sup> - are tubes with an anode wire in the mix of argon and CO<sub>2</sub> gases. DT cells are 2.4 m long and are organised in three groups of four elements (three "super-layers"). When the muon passes through super-layers, it ionises the gas in the cells and released electrons start moving to anodes. Using the time it takes for electrons to reach the anodes, the muon position and direction can be determined. DT resolution of a single-cell hit positions ranges from 200  $\mu$ m in the

r- $\varphi$  plane to 200-600  $\mu\text{m}$  for forward directions.

CSCs are used in the forward direction to cover the region of  $0.9 < |\eta| < 2.4$ . CSCs chambers are multi-wire chambers made of cells that have a trapezoidal shape. Chambers contain radial copper cathode strips and, perpendicular to those, gold-plated tungsten anode wires. Each cell is filled with the mix of Argon, CO<sub>2</sub>, and CF<sub>4</sub> gases. The strip cells have a single-layer resolution of 300-900  $\mu\text{m}$ . A CSC chamber provides a spatial resolution of 40 -150  $\mu\text{m}$ .

To improve the performance of DTs and CSCs, RPCs are used and are covering the barrel and endcaps in the range of  $|\eta| < 1.9$ . RPCs are double-gap chambers consisting of two resistive 2 mm in thickness Bakelite layers separated by a 2 mm layers filled with a mix of C<sub>2</sub>H<sub>2</sub>F<sub>4</sub>, *i*C<sub>4</sub>H<sub>10</sub>, and SF<sub>6</sub> gases.

RPCs operate in avalanche mode, producing an avalanche when the muon traverses the gas of the cell. RPCs have a spatial resolution of 0.8 - 1.2 cm, which is not as good as the ones provided by other muon subsystems, but RPCs have an advantage in terms of an excellent time resolution - just 3 ns. The barrel and the endcaps contain in total 10 RPC stations.

### 1.2.7 The Triggers and DAQ

The CMS trigger [8] is a system responsible for selecting events of interest and storing them for the offline analysis. The trigger has two stages: the L1 trigger (see Fig. 1.16), which reduces the event rate from 40 MHz to 100 kHz, and the HLT trigger, which further decreases the rate to nearly 1 kHz. The L1 trigger consists of the custom hardware that processes a part of the information from calorimeters and outer tracker systems. The HLT trigger is a part of the detector readout system (DRS) and uses the full detector information for event reconstruction. The HLT is a computing farm

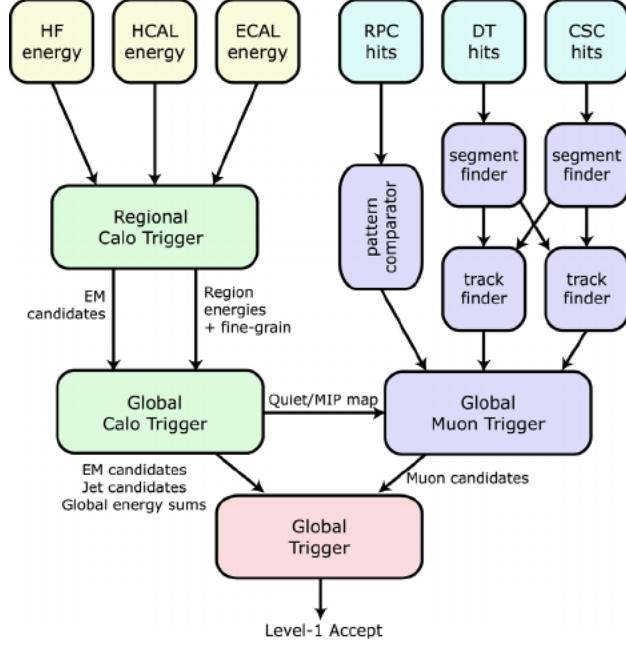
consisting of 22000 CPU cores that produce a decision on whether to save or to skip the event in an average time of about  $220\ \mu s$ . DRS is integrated in the higher level data acquisition (DAQ) system [9]. The selected events are collected and sent by the DAQ to the tapes of the CERN Tier-0 for the persistent storage.

L1 and HLT systems have differences and similarities. They operate at the different time scales and the volumes of data they are processing are completely different. However, the goals of these systems are the same - to identify and reconstruct physics objects and combine their properties to produce an acceptance/rejection decision for each event.

#### 1.2.7.1 The L1 Trigger

L1 system [10] contains a "menu" of 500 algorithms or "seeds" designed to identify useful physics events. These menus include trigger criteria varying from basic single-object identification to complicated selections requiring some topological conditions to be met. Each seed has a set of assigned "Prescale" factors  $f$  that reduce the rate of events accepted by a particular trigger algorithm from 100% to  $100/f\%$ . Prescale factors are necessary since the luminosity level decreases during the run period. They adjust the trigger rate to keep it constant during the data taking time.

Since the processing time of the L1 system is very important for the whole CMS operation, the L1 is built using FPGAs and ASICs custom hardware. L1 produces decisions within  $3.8\ \mu s$ . Data from all the calorimeters is first processed by the L1 regional calorimeter trigger (RCT) and then by a more selective global calorimeter trigger (GCT).



**Figure 1.16:** The CMS L1 trigger layout.

The RCT receives the information about energy deposits from all calorimeters and covers the range  $|\eta| < 5$ . The RCT processes this information in parallel and produces  $e/\gamma$  candidates as an output. The information from the inner tracker is not available, therefore, L1 identifies both electrons or photons, but cannot distinguish them. L1 also detects jets, taus, missing transverse energy (MET or  $\cancel{E}_T$  or  $\cancel{\not{p}_T}$ ), and muons. L1 RCT is responsible for determining the first estimates of the several main parameters of interest:  $p_T$ , isolation (described later in this chapter), etc.

First stage reconstruction uses particle "hits" in the muon detectors and analyses them using track finder algorithms. All three muon detectors of the CMS are used by the L1 muon trigger. Using DT and CSC systems, track segments from the hit information are identified. The pattern recognition algorithms are applied to these segments to reconstruct muon candidates and measure their momenta.

More complex but slower algorithms then re-use hits for a more precise particle identification using a global muon trigger (GMT). The hits from the RPCs are used

directly by the pattern comparator trigger (PACT) that reconstructs muon candidates at the high radii. Then several regional track finder algorithms sort the identified muon candidates and send this information to the GMT. Each candidate contains  $p_T$  and angular information.

The GMT then combines the muon information from different subsystems to avoid duplicating the candidates. The GMT also performs more tight track quality checks and may discard a portion of the input candidates it received.

Finally, the information from the GCT and GMT is combined by a global trigger (GT). The GCT sorts the created  $e/\gamma$  candidates, identifies jets, and calculates  $\cancel{E}_T$ . The final decision of the GT is to store or to skip the event. If the event satisfies the acceptance requirements and is going to be kept, the L1 accept signal (LAS) is generated and propagated by the trigger control and distribution system (TCDS) to all subdetectors.

The GT is the final step of the CMS L1 trigger system and implements a menu of triggers. The output of this system is used as an input to the HLT algorithms.

### 1.2.7.2 The HLT Trigger

The selection done by the HLT mimics the offline analysis - for all reconstructed objects in the event: electrons, muons, and jets - the identification criteria is applied to select only events of interest. What these events are each offline analysis defines in a different way, but to name a few, almost all analyses need true prompt leptons, well reconstructed jets, or some other commonly used objects.

The HLT computing farm has the event filter farm, which consists of filter-builder units (FBU). In the FBU the parts of the events and information from different detector subsystems is combined to produce "complete" events. Then the filter unit unfolds the raw detector data into experiment specific data structure and performs

the event reconstruction and trigger filtering.

The whole event processing procedure of the HLT is centered around the HLT path. The HLT path is a set of algorithmic instructions that in a sequential manner reconstructs physics objects and performs the object selection. The complexity of the steps in the path sequence increases and the quality of the physics objects (the probability to have a correct label) improves too. After this step is completed, selected events are sent to another software processing farm. In this storage manager farm, the data is archived, stored locally on disk, and later sent to the CMS Tier-0 computing center for offline use.

Most data enters the queue for processing and is ready to be sent to Tier-0 very soon. In some cases the special data, the "parked" data, may be collected and kept until the run is finished. In this situations the CMS tape is used and the data has a high-priority for "parking". This may include "hot topic" analyses such as vector boson fusion or parton distribution studies such as Drell-Yan process.

The output of the HLT is limited by capacities of the Tier-0. This includes the bandwidth of the data transfer as well as the amount of available tape. This complicates the work of the DAQ, since in addition to physics data streams, the calibration streams also need to be stored. These streams, though, use information only from few subdetectors.

#### **1.2.7.3 The DAQ system**

The DAQ systems in the modern high energy physics are responsible for any tasks. The challenges are well known: high data rates and volumes, limited tape space and transfer bandwidth. CMS DAQ is based on the homogeneous architecture, scales well with the different beam energy regimes and data rates, and stable performance in a variety of operating conditions.

To illustrate an example of a complex computing task that is elegantly solved by the DAQ system, let us discuss in more details the aforementioned FBUs system of the HLT. FBU relies on a single multi-core machine and the communications with other units are done via shared memory. The data from the full detector is used for the filtering process. Complicated offline-like reconstruction algorithms are then used for the full precision event selection. With more CPU cores available for the Run-2, the per-event time budget is increased to "comfortable" 175 ms per event, which is a long enough time to run most of the CMS reconstruction algorithms.

The current CMS DAQ was developed to address these core requirements:

- The data from one or several data transfer lines is available for other lines,
- the event building is done in parallel profiting from multiple processing units,
- almost real-time process monitoring,

Proper design patterns are used in the software for DAQ, which decouple the user interface from the implementation. The design also allows for the remote control. The software system can be run on a number of different operating systems and hardware platforms. The memory management tools of the underlying system are not linked directly to the applications, it is done using a dedicated abstract addressing scheme.

### 1.2.8 The CMS design

Now that we discussed all CMS subdetectors and DAQ, we can look back and summarise in one list what were the requirements on the design of the CMS detector to successfully complete its physics program. Here we refer to the CMS Technical Design Report [10]:

- good muon momentum resolution over the momentum scale covering almost a TeV range, good dimuon resolution (mostly  $Z \rightarrow \mu\mu$  and  $H \rightarrow \mu\mu$ ) at the O(100) GeV. The capability to determine correctly the charge of the highly energetic muon all the way up to 1 TeV,
- good momentum resolution of all charged particles in the inner tracker,
- good diphoton mass resolution with the focus on the  $H \rightarrow \gamma\gamma$  discovery channel. Also, the ability to reject  $\pi^0 \rightarrow \gamma\gamma$ , which is one of the main background processes to many physics analyses. This requirement mostly concern the performance of the ECAL,
- good resolution of the missing transverse energy (discussed in the section below) and of the mass of the two-jet system. This task depends heavily on the performance of the HCAL.

### 1.3 Physics Object Reconstruction in CMS

Excellent spatial resolution of the CMS trackers, high granularity of the calorimeters, and almost  $4\pi$  coverage of the detector, allowed the CMS to introduce the particle flow (PF) algorithm [11] for a global event reconstruction. PF takes the input from all subdetectors, analyses the redundant information, removes the duplicate one, and forms the physics objects. PF procedure starts with identification of tracks and calorimeter clusters, then the reconstruction of the physics objects is performed, such as electrons, muons, jets, etc. In this section we will discuss the whole PF approach and all key elements of this reconstruction sequence.

## 1.4 Track Reconstruction

The reconstruction starts with the clusters of signals ("hits") in the inner tracker. The information from these clusters in the Pixel and Strip subdetectors is aggregated based on their signal-to-noise ratios. The charge-weighted averaging is performed (for different particle charge hypothesis), as well as other corrections are further applied to identify the real hit positions.

The helix trajectory that the particle follows in the magnetic field inside of the detector is characterised by five parameters: the direction in  $\eta$ , the 3D position with respect to the reference point, which is the centre of the IP, and the curvature of the track with the radius  $R$ . This information is enough to compute estimates of basic physics quantities; however, this task is complicated by the presence of high particle multiplicity in the event (number of charged particles produced in the same event) and also by the physics aspect of the electron propagation in matter: an electron traversing the detector has nearly 85 % probability to emit a bremsstrahlung photon. Hadronic effects also need to be taken into account: a hadron has a 20 % probability to experience multiple scattering on the nuclei of the detector before reaching the HCAL.

To keep the track finding efficiency high, while maintaining low the efficiency of miss-identified tracks, track reconstruction is performed sequentially using the combinatorial track finder (CTF) [12]. First the "purest" tracks are reconstructed, they have high  $p_T$  and the hits point towards the primary vertex (PV). The term PV is used to refer to the vertex which is the actual point of origin of the produced particle, when several other hard scattering vertices are present in the event. Then these pure tracks are removed from the collection of tracks and another round of the track reconstruction starts. This procedure applied several times reduces the combinatorial factor and also simplifies the identification of tracks with the low  $p_T$  or those which do not point to the PV. During each iteration, the reconstruction follows these steps:

- Seed generation. Rough estimates of the particle trajectories ("seeds") are produced

using either three hits or two hits and a PV constraint. Based on which iteration the algorithm is at, some additional constraints are applied, e.g., a minimal  $p_T$  selection, the requirement for the seed to originate close to the beam spot, etc.

- Trajectory building. Initial seeds are projected towards the compatible hits in the next layers. This approach is based on the Kalman Filter (KF) procedure [13]. The extrapolation is done until the outermost layer of the tracker is reached or when a "terminating condition" is satisfied, e.g., when the iteration accumulated the maximum allowed number of invalid hits ("fake hits"). Each obtained trajectory is updated using a KF approach based on the compatibility of hits to form a better track candidate. The procedure is complicated by the fact that the same initial seed can give rise to several track candidates or vice versa, the same track candidate may be compatible with different seeds. Additionally, the trajectory building step should take into account energy loses of the particle due to multiple scattering on the detector material, inhomogeneities of the tracker material, and the effects of the regions of non constant magnetic field.
- Track fitting. After the track candidate has been built, the track parameters are refitted by a KF and by the "smoother". This step uses the full available information about the track and gives optimal estimates of the track parameters. To remove a large number of fake tracks, which are present due to a very complicated nature of the problem and the high track multiplicity in the event, a multivariate (MVA) selection is applied. MVA incorporates variables that discriminate real tracks from the fakes: the signed transverse curvature and impact parameters (with respect to the beam spot), the polar and azimuthal angles, number of missing hits, the fit quality variables, etc.

The CTF procedure runs for 10 iterations. In 2016, the proton-proton collisions had a mean pileup (additional hard scattering vertices) of 24, the CTF efficiency to identify real

tracks varied from 80 to 95% with the mis-identification efficiency of 5 - 10 %.

### 1.4.1 Muon tracking

Muons are detected by the inner tracker and also by the outer (muon) tracker. This greatly improves muon track reconstruction and motivated the development of a dedicated muon reconstruction algorithms. The ninth and the tenth iterations of the CTF are focused on the muon reconstruction. These iterations are using three separate algorithms to identify:

- Standalone muons. This algorithm uses only muon tracker information: DTs, CSCs, RPCs. Hits from the inner chambers are used as seeds and are projected to hits in the outer chambers. Then, a standard KF procedure is used to identify track candidates, which are called standalone muons.
- Tracker muons. Only the inner tracking information is used to form tracks. Tracks are further projected to muon subsystems, where a compatibility with at least one muon hit is required. This algorithm works with low momentum muons: tracks with  $p_T$  above 0.5 GeV and a total momentum greater than 2.5 GeV.
- Global muons. Tracker tracks and standalone tracks are projected to the outermost layer of the muon system, checking the compatibility between two approaches. The resulting combined set of track hits is refitted to produce a global muon track. Mostly high momentum muons with  $p_T > 200$  GeV profit from this algorithm.

### 1.4.2 Electron tracking

Electrons are also detected by the inner tracker, however, their reconstruction is complicated by the fact that they emit bremsstrahlung photons and the trajectory becomes more complex. As a result, the clustering algorithms need also to identify the bremsstrahlung photons and account for the fact that the energy clusters corresponding to these photons may be located outside of the main electron trajectory, when extrapolated to the ECAL.

Since a KF approach assumes that energy losses are Gaussian, and this is not the case for electrons, a dedicated procedure is developed - a modified KF - the Gaussian Sum Filter (GSF) [14]. In this method the radiated energy losses are approximated by the sum of Gaussian distributions.

The electron seeds for the GSF are built using the ECAL information. Two different approaches are developed for the track reconstruction:

- Super Cluster based electrons. Cluster of energy in the ECAL and grouped together to form super clusters (SCs). Using the information of the energy spread among the clusters, the curvature of the electrons is estimated and tracker seeds are formed, with the SC position as a constraint.
- Tracks based electrons. Tracks from the inner tracker are projected to ECAL clusters, checking the compatibility using quality variables, such as  $\chi^2$ , number of missing hits (absent hits along the path of the track), etc.

A typical momentum resolution for electrons in  $Z \rightarrow e^-e^+$  decays is approximately 1.7 - 4.5%.

### 1.4.3 Primary Vertex reconstruction

At the LHC energies, several hard scattering vertices are produced in each collision. The location of all primary vertices (PV) is reconstructed using the tracks. However, normally only one (the main PV) out of these vertices (referred to as additional vertices or pileup) produces the interesting physics interactions. All the vertices are important since they are reused in the feed-back loop of the track reconstruction procedure. Also, a precise identification of primary vertices is important for determining the effect of the pileup on all physics objects in general and b-tagging in particular (will be discussed later in this chapter).

The PV identification consists of three steps: the tracks are selected, then the tracks from the same PV are combined in clusters, and, finally, the position of the PV is determined from the fits to tracks.

Selecting the tracks, only those consistent with the location of the PV are considered. Several variables are used to improve the quality of the track selection procedure: the transverse impact parameter (the relative distance in the vertical plane with respect to the centre of the beam spot), the number of strip and pixel hits associated with a track, and the  $\chi^2$  of the fit .

Clustering of the tracks is based on their z position with respect to the beam spot. A deterministic annealing (DA) algorithm [15] is used to find the global minimum for this problem with many degrees of freedom. The idea of this algorithm is based on the physics example of a thermodynamic system approaching a state of minimal energy through a series of temperature reductions.

Once the clustering is completed and PVs are identified, the candidates with more than two tracks are fitted using an adaptive vertex fitter [16]. The result of this procedure is a set of probabilities assigned to tracks. Each probability can be thought of as a likelihood that the track originated from a given vertex.

The resolution of PVs varies between 10 to 100  $\mu\text{m}$  and depends on track qualities. The main PV is determined as the vertex with the highest sum of the squared transverse momenta ( $\sum p_T^2$ ). Then, the vertices are ordered by the  $\sum p_T^2$  of the corresponding tracks and the vertex with the highest value of the sum is called the main hard scattering vertex, or the PV.

## 1.5 Particle level objects

### 1.5.1 Particle Flow links and blocks

A particle leaves the signal in several CMS subdetectors which is stored as PF [17] elements. To connect these PF elements, the link algorithm (LA) was developed. The LA can test any pair of elements in the event. To speed up the calculations, only pairs of elements that are "neighbours" in the specified  $\eta - \varphi$  plane are considered. When the pair of elements is linked, the LA determines the distance between the elements which is related to the quality of the link. This procedure produces PF blocks of elements, where the blocks are connected either by a direct or an indirect link (through the common elements).

The procedure produces inner tracker-calorimeter, HCAL-ECAL cluster-to-cluster, ECAL-ECAL, and ECAL-Preshower links. In most cases the link distance is defined as the  $\eta - \varphi$  or  $x - y$  distance between the two cluster positions. In case of the ambiguity when, e.g., several HCAL clusters are linked to the same ECAL cluster, only one link is kept - with the smallest distance.

The last stage of the link algorithm is dedicated to formation of the inner tracker-muon system links. Once all links are established and PF blocks are formed, the PF algorithm proceeds reconstructing objects in the following sequence: muon candidates (with corresponding PF tracks and clusters been removed from the PF block), electron candidates (taking into account bremsstrahlung photons) and high momentum isolated photons (with related tracks and clusters also been removed), and, finally, charged hadrons and non-isolated photons.

The elements that are still left in the PF blocks are then re-considered for another round of identification of other objects: charged hadrons and neutral hadrons, photons from parton fragmentation and hadronization, and jet decays. Lastly, when all PF blocks have been sorted out, the global event description is completed and the reconstructed event is re-processed by a post-processing step (PP step), which addresses the possible particle

misidentification and misreconstruction during the previous steps.

### 1.5.2 Muons

Muon reconstruction is the first stage of the PF algorithm. It identifies muons using global and inner tracker muon properties. Global muon candidates are selected at this step and the isolation requirement (explained later in this chapter) is applied. This isolation requirement is efficient enough to reject hadrons mis-identified as muons.

The muons inside jets or secondary muons from hadron decays are complicating the identification of prompt muons, such as those originating from Higgs, W, or Z boson decays. Therefore, additional more stringent selection is further applied.

For non-isolated global muons, in addition to the Tight WP selection, the following extra selection is applied: more than two matching track segments should be present in the muon detectors or the calorimeter energy deposits must be compatible with the muon candidate. This selection discriminates against the high  $p_T$  hadrons.

If muons fail Tight WP, there are still several "recovery" iterative procedures that use the relaxed selection criteria.

Even at this stage in the PF procedure, the muon identification and reconstruction is not finished. Charged hadrons reconstructed during the next PF stages can be reconsidered as muon candidates. Only after the whole PF sequence is completed, including the PP step, the PF algorithm terminates.

All muons considered for this measurement are global muons that satisfy the Tight WP requirements (Tight WP as an initial selection) with extra requirements on the number of hits in the tracker and muon system, on the impact parameter, and the quality of the global track.

The efficiency  $\epsilon$  to successfully identify a prompt isolated lepton (in our case a muon or an electron) can be decomposed as:

$$\epsilon = \epsilon_{\text{tracker}} \cdot \epsilon_{ID|\text{tracker}} \cdot \epsilon_{ISO|ID}$$

where the first term refers to the tracker efficiency, the next term is the Bayesian term which refers to the identification (ID) efficiency given that the lepton already passed the tracker requirements, and the final term refers to the isolation (ISO) efficiency given that the lepton already satisfied identification criteria. All the efficiencies are well optimised in the CMS and are in the range from 85 to more than 99 % depending on the  $p_T$  and  $\eta$  of the lepton (muon in this case).

The muon resolution for 20 to 100 GeV momentum range varies from 1 % in barrel to 5 % in endcaps.

### 1.5.3 Electrons and isolated photons

The reconstruction of electrons is complicated by the fact that they lose energy emitting bremsstrahlung photons and, thus, their trajectory becomes more complex than the one of muons. Additionally, bremsstrahlung photons often convert to  $e^+e^-$  pairs. And this is a recursive process: daughter electrons also emit photons. Due to this complication, it was decided to use almost the same procedure to reconstruct electrons and photons. First a GSF track plays a role of the seed for the electron candidate. For the photon candidate, an ECAL supercluster with no links to the GSF track is used as a seed. For both electron and photon candidates energy deposits in the HCAL must not exceed 10% of the ECAL energy.

Then, all ECAL clusters in the PF block are linked to the SC or to one of the GSF track tangents of the candidate. The total energy of the collected ECAL clusters is corrected for the energy losses during the linking procedure and is assigned to photons. An electron candidate is formed from a combination of the corrected ECAL energy and the electron direction given by the GSF track. Additional MVA discriminator of O(20) variables is applied to improve electron identification efficiency. The MVA approach based on the Boosted Decision Trees (BDT) classifier [18] profits from the following highly discriminating variables: the amount of energy radiated off the GSF track, the distances between the ECAL SC position and the projection from the GSF track, track-cluster linking variables, KF and

GSF track quality variables, etc.

Photon candidates are kept if photons are isolated and the corresponding configuration of ECAL energy deposits is compatible with those expected from a given photon shower.

All identified electron and photon tracks and clusters in the PF block are masked before the algorithm starts processing hadrons. Since some offline physics analyses may apply different selection for electrons and photons, PF selection is relatively loose and the full electron and photon reconstruction information is saved in case a different re-interpretation must be run. This offers the saving of the computing time in the future, since re-running the electron track reconstruction would not require re-running the complete PF algorithm again.

#### 1.5.4 Hadrons and non-isolated photons

After the muons, electrons, and isolated photons are reconstructed, they are removed from the PF blocks. The next PF algorithm iterations proceed with hadrons from jet fragmentation and hadronization. These particles can be seen by the detector as charged pions, kaons or protons, neutral pions and kaons, and non-isolated photons from neutral pion decays. During the reconstruction, the precedence is given in the ECAL to photons over neutral hadrons. This priority does not hold above  $|\eta| > 2.5$ . In that region ECAL clusters linked to a given HCAL cluster are identified as hadrons and, only if ECAL clusters are without such a link, then they are classified as photons.

What is left in the PF block, gets classified in the following manner: if energy deposits are consistent with the energy hypothesis from the tracker, then no neutral hadron is found and each track that remains corresponds to a charged pion candidate.

In situations when the energy deposits in the ECAL do not match well the energy hypothesis from the tracker and this discrepancy is larger than three standard deviations - a new muon reconstruction starts, with the relaxed muon selection. This approach allows to improve the muon identification efficiency without increasing the rate of mis-identified

muons.

At times the track momentum sum may be found to be significantly larger than the calorimetric energy. Usually this excess in momentum is often found to arise from residual misreconstructed tracks with a  $p_T$  larger than 1 GeV. These tracks are sorted in decreasing order of their  $p_T$  and are removed one-by-one from the PF block until no such tracks are left in the initial track collection.

The hadron traversing the material of the tracker interacts with the nuclei of the tracker material and often produces secondary hadrons. These secondary hadrons are evidently produced outside of the PV - at the secondary (intermediate) interaction vertex. When the tracks of the charged particles corresponding to these secondary particles are linked together, the resulting secondary particle candidates can be replaced in the list of reconstructed particles by a single (original) charged hadron.

Estimate of the energy of the primary charged hadron is then given by:

$$E = E_{\text{secondary}} + f \cdot p_{\text{primary}}$$

where  $E_{\text{secondary}}$  is a vectorial sum of the momenta of the secondary charged particles,  $p_{\text{primary}}$  is the momentum of the incoming track, and  $f$  is the factor determined from the simulations.

### 1.5.5 Jets and jet corrections

Jets are collimated streams of particles created during the processes of the fragmentation and hadronization of the original parton, quark or gluon. As jets propagate through the CMS detector, they leave tracks in the tracking system and interaction showers in the calorimeter crystals.

Several jet reconstruction algorithms have been developed. In the Higgs boson group of the CMS, most measurements are using anti- $k_T$  algorithm [19]. If the jet clustering uses PF particles - PF jets are reconstructed. If only the ECAL and HCAL information is used - calorimeter jets are identified ("calo jets"). When all stable particles (in case

of the simulation, at the generator level) excluding neutrinos are used - reference jets are reconstructed ("Ref jets"). In this measurement PF anti- $k_T$  jets are used.

The anti- $k_T$  algorithm is one of the "cone" algorithms that takes as input a collection of PF objects inside of the cone of the radius R around the object (R is usually defined by the jet size). The algorithm defines the distance parameter:  $d_{ij} = \min(\frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2}) \times \frac{R_{ij}^2}{R}$ , where  $p_{T_i}$  and  $p_{T_j}$  refer to the transverse momenta of PF particles  $i$  and  $j$ ,  $R_{ij}$  is the distance in the  $\eta - \varphi$  plane between particles  $i$  and  $j$ .

An additional  $d_{iB}$  parameter is defined as the distance between the particle  $i$  and the beam spot position:  $d_{iB} = \frac{1}{p_{T_i}^2}$ .

The algorithm iteratively finds the minimum distance selecting at each step the minimal value for each  $(d_{ij}, d_{iB})$  pair, using a collection of the PF particles as an input. The algorithm finds the smallest  $d$  for a given input. If the minimum distance is  $d_{ij}$ , then the four-vectors of  $i$  and  $j$  particles are summed to form a new particle. Particles  $i$  and  $j$  are removed from the initial input collection. If the minimum distance is  $d_{iB}$ , then the particle  $i$  is considered a jet. This particle is also removed from the set of particles and the algorithm continues until all initial particles have been combined into jets. The mechanics of the algorithm is such that first soft jet candidates with the hardest particles are clustered, producing a perfect cone-shaped jets, then more "complicated" jets are reconstructed.

PF jets are used in this thesis. They are superior to Calo jets since the former have a better angular resolution. PF algorithm allows the precise determination of the charged hadron direction and momentum, while in calorimeters, the energy deposits of charged hadrons are spread along the  $\varphi$  direction in the presence of the magnetic field, which leads to an extra degradation of the azimuthal angular resolution of jets.

On average, the relative contributions to jet energy are from: 65% from charged hadrons, 25% from photons, and 10% from neutral hadrons. The possibility to identify the contributors to the total jet energy during the jet reconstruction is one of the reasons to use the PF algorithm for jet reconstruction. In practice the identification of particles inside jets is

done comparing the jet energy fractions measured in PF jets to those of the corresponding Ref jets.

To remove the jet energy dependence on  $p_T$  and  $\eta$  (JE map) and, hence, to make the corresponding two dimensional JE map uniform, the jet energy correction (JEC) procedure is introduced. Additionally, the jet energy resolution (JER) correction is necessary. The latter is defined as the Gaussian width of the ratio of the energies of the corrected PF jets to Ref jets. Both corrections improve the angular resolution, energy response, and energy resolution of jets.

JEC scales the four-momentum of jets. The various detector effects are addressed. Depending on how sensitive the physics analysis is to JEC, one can factorise this correction into separate components and apply them individually in a sequence. Most important individual corrections remove: energy contributions due to pileup, effects of the calorimeter response, residual Data-MC discrepancies, effects of the jet flavor, etc.

JER smears the four-momenta of reconstructed jets to match the energy resolution observed in data. The smearing procedure derives the correction factors which scale the reconstructed jet momentum with respect to the the momentum of the same jet, but clustered at the MC generator level.

## 1.6 Other important physics quantities and objects

### 1.6.1 The b tagging and secondary vertices

In this measurement one of the Higgs bosons decays to b quarks. Jets produced during the hadronisation of b quarks are called b jets. A dedicated b tagging is necessary since this decay is a great test of the SM validity - Higgs boson decay to b quarks has the highest branching fraction of almost 58%.

Bottom quarks will produce jets that contain B mesons, which have a relatively long lifetime  $c\tau \approx 500\mu\text{m}$ . This distance, travelled at almost the light speed, would correspond to a dislocation of a few mm from the PV. The positions of B meson decays will be clearly seen in the detector. Each such position with a corresponding displaced vertex, the secondary vertex (SV), is a unique signature of b quark and is used to identify the b quark decay. Sometimes the vertex cannot be unambiguously reconstructed, but even in these cases the properties of tracks within b jets are different from the ones originating from gluons or light quarks (light jets).

After passing the selection criteria, tracks are considered for b tagging. The selection requirements, that include kinematic and impact parameter properties of tracks, are needed to reject fake tracks, tracks coming from pileup vertices, and tracks from the long-lived hadrons.

Two main approaches are used to reconstruct the secondary vertices (SV). One of the methods that pioneered the b tagging is an adaptive vertex reconstruction (AVR) algorithm. AVR is based on the adaptive vertex fitter, it uses the tracks associated with jets and finds PVs and SVs. The other algorithm is the inclusive vertex finder (IVF). IVF uses all the tracks in the event and is implemented with the selection looser than for the AVR.

As for the b tagging itself [20, 21], multiple algorithms (taggers) have been developed and successfully used over the last decades. The most known ones are:

- the jet probability (JP) and the jet b probability (JBP) taggers. Both are based on the probability of a jet candidate to be compatible with the PV using impact parameter significance (IPS) variables.
- The soft electron tagger (SET) and soft muon tagger (SMT). These taggers are based on the presence of soft leptons within jets, focusing on leptonic decays of B hadrons.
- The combined secondary vertex (CSVv2) tagger. This is a more complex tagger based on the MVA technique. It uses displaced tracks and secondary vertices to tag b jets

and takes as input IPS, decay length, SV parameters, number of SVs, etc. This tagger can use both AVR and IVF vertices.

In this physics analysis a new MVA based (cMVAv2) tagger was used. This superior tagger uses the outputs from all the aforementioned "fundamental" taggers: JP and JBP, SET and SMT, CSVv2 using both AVR and IVF vertices. These ensemble learning procedure [22] of combining outputs from fundamental taggers into one complex MVA based tagger that produces the final output - is a popular machine learning technique that has been proven to lead to better results than the ones achieved by individual algorithms separately.

### 1.6.2 Missing transverse momentum

When neutrinos are present in the event, they cannot be directly detected by the CMS, specific neutrino detectors would be needed in this case. However, using the CMS detector one can indirectly estimate the momentum of neutrinos. This procedure relies on the method of the "missing transverse momentum"  $\cancel{p}_T$  (or missing transverse energy  $\cancel{E}_T$  (MET)). MET is constructed using all PF particles in the event and is calculated as:

$$\cancel{p}_T = \vec{p}_T^{miss} = \left| - \sum_i^N \vec{p}_{Ti} \right|.$$

To reconstruct  $\cancel{p}_T$ , the CMS relies on almost  $4\pi$  coverage of the detector and precise measurement of the particle properties using PF algorithm that takes information from all subsystems. Produced  $\cancel{p}_T$  is still considered "raw", since this MET is not yet JEC or JES corrected. After these corrections are applied, one obtains  $\cancel{p}_T$  that is called "Type-1 corrected MET". This is the definition of MET that is recommended by the CMS JetMET particle Object Group (POG) group and is used in this measurement. Additional set of filters and corrections is further applied to reject events with artificially large  $\cancel{p}_T$  due to the presence of several noise sources, such as ECAL dead cells, poor quality muon candidates, HCAL noise.

### 1.6.3 Pileup interactions

In 2016, a typical collision event contained on average 24 interaction vertices. Some events had nearly 40 inelastic proton-proton interactions. All the vertices excluding the PV, can be referred to as the soft vertices. These soft vertices (pileup or PU), even though do not produce interesting physics, contribute considerably to the total number of particles produced in the event. PU creates additional hadrons and photons and this affects the PF reconstruction of jets and  $\cancel{p}_T$ , and also the lepton isolation calculation.

Backtracking the origin of the tracks and checking the compatibility with the PU vertices, charged hadrons originated from PU are identified. These charged hadrons are removed from the collection of particles used to reconstruct physics objects. This procedure is called the charged-hadron subtraction (CHS). Since neutral hadrons and photons leave no tracks in the inner tracker, their presence needs to be identified and addressed differently. Instead, an average  $p_T$  density  $\rho$  of pileup interactions in the given  $\eta - \varphi$  slice is calculated. Assigning the area of the candidate to the value  $A_{eff}$  ("effective area"), the expected pileup initiated contribution, that needs to be subtracted, is calculated as  $\rho \cdot A$ . Several other techniques are available, with the most simple one relying on the calculation of the ratio of the neutral to the charged energy coming from pileup (around a given lepton). From the studies, this value (called  $\Delta\beta$ ) is determined to be very close to 0.5 [23].

### 1.6.4 Lepton isolation

As mentioned before, lepton isolation is a great method to remove clear fakes and select real prompt muons and electrons produced by Higgs boson decay or by the weak decays of Z or W bosons. The isolation quantifies the activity of other particles around the particle of interest. The lepton isolation is defined as the scalar sum of the  $p_T$ 's of all charged and neutral hadrons and photons inside a cone of the radius  $\Delta R < 0.3 - 0.4$ (depending on the working point (WP) and a lepton flavor). The sum is normalised by the  $p_T$  of the lepton of interest:

$$I_{PF} = \frac{1}{p_T} (\sum^\gamma p_T^\gamma + \sum^{h^\pm} p_T^{h^\pm} + \sum^{h^0} p_T^{h^0})$$

There are other physics analysis-specific definitions of isolation and they are applied offline. In this thesis, two isolation requirements listed below are used.

For electrons, the isolation selection relies on the notion of the effective area defined in the previous subsection. The effective areas are proportional to  $\Delta R$  cone size around the electron (0.3 in this case) and for electrons is given by:

$$I_{PF}^{electron} = \sum^{h^\pm} p_T + \max(0, \sum^{h^0} p_T + \sum^\gamma p_T - \rho \cdot A_{eff}).$$

Mentioned previously  $\Delta\beta$  method is used for muon isolation selection. With the cone size of 0.4, the isolation is given by:

$$I_{PF}^{muon} = \sum^{h^\pm} p_T + \max(0, \sum^{h^0} p_T + \sum^\gamma p_T - \Delta\beta \cdot \sum^{h_{PU}^\pm} p_T).$$

The last term is a sum over the charged hadrons originated due to pileup.

Both isolations are constructed from the collection of particles containing charged and neutral hadrons, photons, and charged hadrons from pileup (in case of muons).

### 1.6.5 Datasets and Trigger Paths

The proton-proton collision data recorded by CMS is split into "eras", which are labelled using alphabetic letters A → H. Period A was dedicated to commissioning of the LHC for 2016 data taking. Periods from B to H were used for physics. Each era corresponds to a relatively stable period of the LHC conditions, such as the collision rate, the set of trigger menus, etc. Eras B to G were re-reconstructed at the end of 2016 to take advantage of the updated calibration of subsystems and the detector alignment. The data from the last era (H) was re-reconstructed during data-taking itself.

Dozens of dataset (Primary Datasets or PDs) are recorded and stored by the CMS and its computing centres. The data that is analysed in this thesis belongs to "Dimuon"("DoubleMuon") or "Dielectron"("DoubleEG") datasets. The naming is coming from the fact that the final state contains two prompt muons or electrons. The final state also contains other objects, but the leptons are given the "superiority", since the selection of the "on shell" Z boson

decays to leptons significantly improves the sensitivity of the measurement. “On shell” here refers to Z boson decays when the bosons are near the pole of the Z mass distribution (91.2 GeV).

The name of the trigger paths (L1 in this case) reflects the number of leptons selected by a given trigger, the type of the lepton(s), following by the minimal  $p_T$  requirement(s) on the lepton candidate(s). If two leptons are present, their  $p_T$ ’s are referred to as  $p_T$ ’s of the leading and subleading(trailing) lepton (another common naming is two “legs”). If the suffixes “Iso” or “Id” are present, it indicates that the isolation or identification requirements respectively have been also applied. The label “DZ” or “dz” is an additional requirement on the spatial compatibility along the z axis between the lepton candidates and the PV location. Abbreviations “VVL” and “VL” refer to very very lose and very lose selections respectively. Their exact definition may vary, but the important point is that this selection does reject some clear fakes, while is still loose enough to leave enough statistics for offline analysers.

A simplified version of the PF event reconstruction sequence is performed at the HLT level. This sequence is based only on the regional track finding and fitting, relying on the muon and electron candidates found by L1 trigger. The HLT paths refer to triggers that select events where one or two leptons that pass certain selection criteria are present. HLT muons (abbreviated by “Mu”) are formed propagating the L1 track inwards to the inner tracker, or by starting from an inner track and projecting it to the outer tracker. HLT tracker muons (abbreviated by “TkMu”) are reconstructed using the muon tracking procedure discussed in the Section 1.4.1. The HLT electrons are reconstructed using L1 seeds and checking their compatibility with the energy deposits in the ECAL. If the identification selection (ID) is applied on the track, the candidate name contains the suffix “TrackId”. If the ID is applied on the ECAL energy cluster parameters, the name of the path contains “CaloId”.

### 1.6.6 Data-Monte Carlo corrections

Even though particle interactions in the detector, the detector response, and the work of subsystems are simulated at the high precision level; the disagreement with the data is still present. Used approximations to speed up the subsystem responses, slight detector misalignments, impossibility to know the exact parton distribution function (PDF) of the interacting particles at the interaction vertex, fluctuation of the LHC parameters and other factor contribute to the Data-Monte Carlo discrepancy. To improve this disagreement, corrections factors are introduced. They are measured both in data and in Monte Carlo (MC), and their ratio is applied to MC to make the MC more “data-like”. These corrections, also called scale factors (“SF”), are derived for all physics objects and the most important ones are discussed below.

#### 1.6.6.1 Lepton efficiencies and the Tag-and-Probe method

In the process of selecting a prompt lepton several steps are involved: tracker, identification, isolation, and trigger efficiencies. A very popular technique to measure the efficiency is called the Tag-and-Probe (*T&P* or *TnP*) method. Decays of  $Z \rightarrow e^-e^+$  or  $Z \rightarrow \mu^-\mu^+$  are used in this technique. The procedure first picks one lepton that has to pass a relatively tight ID selection, this lepton is called the “tag”. Tags are often referred to as “golden” electrons or muons since this selection has a very low fake rate of the order of 1% or less. Then the other lepton, called the “probe”, is selected to make a pair with the tag. This step results in total of  $P_{total}$  pairs. This pairing procedure includes some very basic selection: probes should be of the opposite sign and the same lepton flavor (OSSF). Consistency with the Z boson pole mass is further checked. The exact definition of the probe object varies depending on the specifics of the selection of interest or the working point. In this framework, the efficiency is defined as a ratio of the number of probes  $P_{pass}$  that pass a relatively “tight” ID WP to the total number of probes  $P_{total}$  formed by the pairing procedure:

$$\epsilon_{WP} = \frac{P_{pass}^{WP}}{P_{total}}$$

Since the efficiency cannot be flat for all  $p_T$  and  $\eta$  ranges, a set of efficiencies is derived for different  $p_T$  and  $\eta$  slices. The procedure also has an uncertainty associated with the method. Mostly the uncertainty comes from statistical and systematic uncertainties of the fitting procedures and particular fit parameterisation model. The TnP procedure allows for the removal of the combinatoric backgrounds by kinematic fitting or sideband subtraction methods [24].

The TnP method applied independently to data and MC, produces scale factors given by:  $SF_{WP} = \frac{\epsilon_{WP}^{data}}{\epsilon_{WP}^{MC}}$ .

L1 trigger did not have a proper simulation for 2016 data taking settings, so only HLT trigger SFs are measured for the trigger. Therefore, for L1, simulated events were weighted by the efficiency measured directly in data.

The formula to compute the HLT efficiency uses individual efficiencies of both legs:

$$\epsilon_{HLT} = \epsilon_{1T}\epsilon_{2P} + \epsilon_{2T}\epsilon_{1P} - \epsilon_{1T}\epsilon_{2T},$$

where the naming convention is as follows:  $\epsilon_{1T}$  refers to the efficiency of the first lepton passing tag selection,  $\epsilon_{2P}$  refers to the efficiency of the second lepton passing probe selection, etc.

At times the probe can also pass a relatively tight selection of a tag. In this case a double counting occurs. The extra pairs need to be removed. On the Fig. ?? the area that contains double counted pairs is at the top right corner and corresponds to the di-lepton efficiency  $\epsilon_{1T}\epsilon_{2T}$ , this area is subtracted in the calculation of the  $\epsilon_{HLT}$ .

In addition, some HLT paths required the DZ requirement, while others did not. Therefore, one needs to estimate additional efficiency related to the DZ selection and derive the corresponding SFs. The DZ scale factor calculation is very similar to estimation of efficiencies of all other sorts: the numerator contains events that pass the DZ requirement and the denominator is equal to the number of events that pass the selection without DZ selection. Exactly the same procedure is applied to derive tracker, ID and ISO SFs. Analysis specific figures will be shown in the data analysis chapter.

### 1.6.6.2 b tagging efficiency

Scale factors also need to be derived for b jets. Tracker misalignment, impossibility to know the content of the PDF of the interacting particles, and the imperfect knowledge of the hadronization process of the b quark - all these factors lead to data-MC discrepancies. Additionally, Strip tracker had known inefficiencies during 2016 data taking, which resulted in worse b tagging performance: lower efficiency to tag real b jets and higher fake rate (incorrect mistagging of light jets or gluons as b jets. The SFs have been derived by the b tagging CMS POG for the use of the whole collaboration. These correction factors are measured using the “true” or “generator” (gen) flavor of the original quark in the MC. Based on that, the weight is assigned to the particle level jet that is matched to the reconstruction level jet. SFs are provided in  $p_T$  and  $\eta$  slices to make the b tagging efficiency uniform across the whole kinematic phase space. Further, MC events are reweighted using the combined weights from all jets present in the events.

In this measurement, the data analysis is performed with b jets tagged by the cMVAv2 (or just CMVA) algorithm. This algorithm uses the  $t\bar{t}$  process containing top and anti-top quark decays to determine the b tagging weights. CMVA discriminant should be above a certain threshold for b jets to be considered originating from the b quarks. The threshold is chosen to correspond to the medium working point of the algorithm defined such that the misidentification rate for light-quark and gluon jets is about 1%. The b jet tagging efficiency for this WP is about 66%.

In the next chapter we will dive deep into the physics analysis of the data produced by the LHC and collected with the CMS detector that has been used to perform the measurement of the double Higgs boson decays.

---

## References

---

- [1] Cecile Noels. Literature in focus - The Large Hadron Collider: A Marvel of Technology. Literature in focus - The Large Hadron Collider: A Marvel of Technology. (BUL-NA-2009-414. 51/2009):11, Dec 2009.
- [2] John Hauptman. *Particle physics experiments at high energy colliders*. 2011.
- [3] CMS Collaboration. The Phase-2 Upgrade of the CMS Tracker. Technical Report CERN-LHCC-2017-009. CMS-TDR-014, CERN, Geneva, Jun 2017.
- [4] Philippe Bloch, Robert Brown, Paul Lecoq, and Hans Rykaczewski. *Changes to CMS ECAL electronics: addendum to the Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2002.
- [5] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.
- [6] G Baiatian, Albert M Sirunyan, Virgil E Emelianchik, Igor Barnes, Alvin T Laasanen, and Arnold Pompos. Design, Performance, and Calibration of CMS Hadron-Barrel Calorimeter Wedges. Technical Report CMS-NOTE-2006-138. 1, CERN, Geneva, May 2007.

- [7] CMS Collaboration. The Phase-2 Upgrade of the CMS Muon Detectors. Technical Report CERN-LHCC-2017-012. CMS-TDR-016, CERN, Geneva, Sep 2017. This is the final version, approved by the LHCC.
- [8] Vardan Khachatryan et al. The CMS trigger system. *JINST*, 12(01):P01020, 2017.
- [9] G. Bauer et al. The CMS data acquisition system software. *J. Phys. Conf. Ser.*, 219:022011, 2010.
- [10] G L Bayatian, S Chatrchyan, H Hmayakyan, A Pobladuev, M E Zeller, and B S Yuldashev. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items.
- [11] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.
- [12] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.
- [13] Giuseppe Cerati et al. Parallelized and Vectorized Tracking Using Kalman Filters with CMS Detector Geometry and Events. 2018.
- [14] Wolfgang Adam, R Fröhwirth, Are Strandlie, and T Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. Technical Report CMS-NOTE-2005-001, CERN, Geneva, Jan 2005.
- [15] E Chabanat and N Estre. Deterministic Annealing for Vertex Finding at CMS. 2005.

- [16] R Fröhwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Adaptive Vertex Fitting. Technical Report CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.
- [17] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.
- [18] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.
- [19] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- $k_t$  jet clustering algorithm. *JHEP*, 04:063, 2008.
- [20] Identification of b quark jets at the CMS Experiment in the LHC Run 2. Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016.
- [21] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.
- [22] Lior Rokach. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1):1–39, Feb 2010.
- [23] Gregory Soyez. Pileup mitigation at the LHC: a theorist’s view. *Phys. Rept.*, 803:1–158, 2019.
- [24] N. Adam, J. Berryhill, V. Halyo, A. Hunt, K. Mishra. Generic Tag and Probe Tool for Measuring Efficiency at CMS with Early Data . [http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2009\\_111\\_v1.pdf](http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2009_111_v1.pdf).
- [25]

- [26] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.
- [27] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth.*, A434:435, 1999.
- [28] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. [Erratum: Eur. Phys. J.C73,2501(2013)].
- [29] SM Higgs Combination. Technical Report CMS-PAS-HIG-11-011, CERN, Geneva, 2011.
- [30] S. Frixione, P. Nason, and C. Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *JHEP*, 11:070, 2007.
- [31] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.
- [32] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano.  $HW^\pm/HZ + 0$  and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 10:083, 2013.
- [33] Comparison of nuisances for background only case, 350 GeV mass hypothesis.  
[http://rkamalie.web.cern.ch/rkamalie/feb12/Comparison\\_of\\_nuisances\\_expectedSignal0\\_350GeV.html](http://rkamalie.web.cern.ch/rkamalie/feb12/Comparison_of_nuisances_expectedSignal0_350GeV.html)
- [34] Comparison of nuisances for s+b case, 350 GeV mass hypothesis.  
[http://rkamalie.web.cern.ch/rkamalie/feb12/Comparison\\_of\\_nuisances\\_expectedSignal1\\_350GeV.html](http://rkamalie.web.cern.ch/rkamalie/feb12/Comparison_of_nuisances_expectedSignal1_350GeV.html)

- [35] k-factor for DY/Z. [https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns#DY\\_Z](https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns#DY_Z)
- [36] The LHC Higgs cross-section working group. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- [37] Standard Model Cross Sections for CMS at 13 TeV. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat13TeVInclusi>
- [38] SM Higgs production cross sections at  $\sqrt{s} = 13$  TeV. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV#ZHllH>
- [39] NNLO+NNLL top-quark-pair cross sections. [https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO#Top\\_quark\\_pair\\_cross\\_se](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO#Top_quark_pair_cross_se)
- [40] Single Top Cross sections. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopSigma>.
- [41] CMS GEN XSEC Task Force. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/GenXsecTaskF>
- [42] SM Higgs Branching Ratios and Total Decay Widths (update in CERN Report4 2016). [https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR#Higgs\\_2\\_g](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR#Higgs_2_g)
- [43] Ryan Gavin, Ye Li, Frank Petriello, and Seth Quackenbush. W Physics at the LHC with FEWZ 2.1. *Comput. Phys. Commun.*, 184:208, 2013.
- [44] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007.
- [45] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-

- level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.
- [46] Gunter Zech. Upper Limits in Experiments with Background Or Measurement Errors. *Nucl. Instrum. Meth.*, A277:608, 1989.
- [47] A. L. Read. Presentation of search results: the CLs technique. *J. Phys. G: Nucl. Part. Phys.*, 28, 2002.
- [48] Rikkert Frederix, Emanuele Re, and Paolo Torrielli. Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO. *JHEP*, 09:130, 2012.
- [49] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C*, 53:473–500, 2008.
- [50] Serguei Chatrchyan et al. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:P11002, 2011.
- [51] John M. Campbell and R. K. Ellis. MCFM for the Tevatron and the LHC. *Nucl. Phys. Proc. Suppl.*, 205-206:10, 2010.
- [52] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J.*, C71:1547, 2011.
- [53] Gael L. Perrin, Pedro Fernandez Manteca. Muon Identification and Isolation Scale-Factors on 2016 Dataset. [https://indico.cern.ch/event/611558/contributions/2465881/attachments/1407735/2151747/TnP\\_06\\_02\\_2017.pdf](https://indico.cern.ch/event/611558/contributions/2465881/attachments/1407735/2151747/TnP_06_02_2017.pdf).

- [54] CMS Muon POG. Tracking SFs on the full 2016 data. [https://twiki.cern.ch/twiki/bin/view/CMS/MuonWorkInProgressAndPagResults#Results\\_on\\_the\\_full\\_2016\\_data](https://twiki.cern.ch/twiki/bin/view/CMS/MuonWorkInProgressAndPagResults#Results_on_the_full_2016_data).
- [55] Gael L. Perrin. Double Muon trigger efficiency per-leg approach. [https://indico.cern.ch/event/636555/contributions/2577291/attachments/1453162/2241537/TnP\\_DoubleMuSF\\_03\\_05\\_17.pdf](https://indico.cern.ch/event/636555/contributions/2577291/attachments/1453162/2241537/TnP_DoubleMuSF_03_05_17.pdf).
- [56] Grace Dupuis. Collider Constraints and Prospects of a Scalar Singlet Extension to Higgs Portal Dark Matter. *JHEP*, 07:008, 2016.
- [57] Shinya Kanemura, Kunio Kaneta, Naoki Machida, Shinya Odori, and Tetsuo Shindou. Single and double production of the Higgs boson at hadron and lepton colliders in minimal composite Higgs models. *Phys. Rev.*, D94(1):015028, 2016.
- [58] Alexandra Oliveira. Gravity particles from Warped Extra Dimensions, predictions for LHC. 2014.
- [59] Oleg Antipin, David Atwood, and Amarjit Soni. Search for RS gravitons via W(L)W(L) decays. *Phys. Lett.*, B666:155–161, 2008.
- [60] A. Liam Fitzpatrick, Jared Kaplan, Lisa Randall, and Lian-Tao Wang. Searching for the Kaluza-Klein Graviton in Bulk RS Models. *JHEP*, 09:013, 2007.
- [61] Kaustubh Agashe, Hooman Davoudiasl, Gilad Perez, and Amarjit Soni. Warped Gravitons at the LHC and Beyond. *Phys. Rev.*, D76:036006, 2007.
- [62] CMS Higgs PAG. List of question for the preapproval checks. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/HiggsWG/HiggsPAGPreapprovalChecks>.

- [63] CMS BTV POG. Supported Algorithms and Operating Points. [https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco#Supported\\_Algorithms\\_and\\_Operati](https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco#Supported_Algorithms_and_Operati).
- [64] Albert M Sirunyan et al. Search for resonant and nonresonant Higgs boson pair production in the bblnulnu final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV. 2017.
- [65] Abdus Salam and John Clive Ward. On a gauge theory of elementary interactions. *Nuovo Cim.*, 19:165–170, 1961.
- [66] Steven Weinberg. A model of leptons. *Phys.Rev.Lett.*, 19:1264–1266, 1967.
- [67] Albert M Sirunyan et al. Evidence for the Higgs boson decay to a bottom quark-antiquark pair. 2017.
- [68] Albert M Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.
- [69] Aruna Kumar Nayak. Reconstruction of physics objects in the CMS detector. *PoS*, CHARGED2012:010, 2012.
- [70] Lisa Randall and Raman Sundrum. Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, Oct 1999.
- [71] Alexandra Oliveira. Gravity particles from Warped Extra Dimensions, predictions for LHC. 2014.
- [72] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

- [73] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [74] Giuliano Panico. Prospects for double Higgs production. *Frascati Phys. Ser.*, 61:102, 2016.
- [75] I. Belotelov, I. Golutvin, D. Bourilkov, A. Lanyov, E. Rogalev, M. Savina, and S. Shmatov. Search for ADD extra dimensional gravity in di-muon channel with the CMS detector. CMS Note 2006/076, 2006.
- [76] M. Aldaya, P. Arce, J. Caballero, B. de la Cruz, P. Garcia-Abia, J. M. Hernandez, M. I. Josa, and E. Ruiz. Discovery potential and search strategy for the standard model Higgs boson in the  $H \rightarrow ZZ^* \rightarrow 4\mu$  decay channel using a mass-independent analysis. CMS Note 2006/106, 2006.
- [77] A. Brandt et al. Measurements of single diffraction at  $\sqrt{s} = 630$  GeV: Evidence for a non-linear  $\alpha(t)$  of the pomeron. *Nucl. Phys. B*, 514:3, 1998.
- [78] W. Buchmüller and D. Wyler. Constraints on SU(5)-type leptoquarks. *Phys. Lett. B*, 177:377, 1986.
- [79] CMS Collaboration. CMS technical design report, volume II: Physics performance. *J. Phys. G*, 34:995, 2007.
- [80] CMS Collaboration. Jet performance in pp collisions at  $\sqrt{s}=7$  TeV. CMS Physics Analysis Summary CMS-PAS-JME-10-003, 2010.
- [81] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

- [82] Particle Data Group, J. Beringer, et al. Review of Particle Physics. *Phys. Rev. D*, 86:010001, 2012.
- [83] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group, and the SLD Electroweak and Heavy Flavour Groups. Precision electroweak measurements and constraints on the Standard Model. 2010.
- [84] I. Bertram, G. Landsberg, J. Linnemann, R. Partridge, M. Paterno, and H. B. Prosper. A recipe for the construction of confidence limits. Technical Report TM-2104, Fermilab, 2000.
- [85] L. Moneta, K. Belasco, K. S. Cranmer, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf. The RooStats Project. In *13<sup>th</sup> International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. SISSA, 2010. PoS(ACAT2010)057.
- [86] Vardan Khachatryan et al. Search for the standard model Higgs boson produced through vector boson fusion and decaying to  $b\bar{b}$ . *Phys. Rev.*, D92(3):032008, 2015.
- [87] Performance of muon identification in pp collisions at  $s^{**0.5} = 7$  TeV. Technical Report CMS-PAS-MUO-10-002, CERN, Geneva, 2010.
- [88] Serguei Chatrchyan et al. Performance of CMS muon reconstruction in  $pp$  collision events at  $\sqrt{s} = 7$  TeV. *JINST*, 7:P10002, 2012.
- [89] CMS COLLABORATION. Particle-flow event reconstruction in CMS and performance for jets, taus, and  $E_T^{miss}$ . CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.

- [90] CMS COLLABORATION. Commissioning of the particle-flow event reconstruction with the first lhc collisions recorded in the cms detector. CMS Physics Analysis Summary CMS-PAS-PFT-10-001, CERN, 2010.
- [91] Vardan Khachatryan et al. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV. *JINST*, 10(06):P06005, 2015.
- [92] CMS Collaboration. Search for pair production of first-generation scalar leptons in pp collisions at  $\sqrt{s} = 7$  TeV. Submitted to *Phys. Rev. Lett.*, 2010.
- [93] CMS Collaboration. Performance of cms muon reconstruction in pp collision events at  $\sqrt{s} = 7$  TeV. Submitted to *J. Inst.*, 2012.
- [94] ATLAS Collaboration. Search for the Higgs boson in the  $H \rightarrow WW(*) \rightarrow \ell^+ \nu \ell^- \bar{\nu}$  decay channel in pp collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector. Submitted to *Phys. Rev. Lett.*, 2011.
- [95] Matteo Cacciari and Gavin P. Salam. Dispelling the  $N^3$  myth for the  $k_t$  jet-finder. *Phys. Lett. B*, 641:57, 2006.
- [96] CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.
- [97] CMS Muon POG. Reference muon id, isolation and trigger efficiencies for Run-II. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffsRun2>.
- [98] John M. Campbell, R. Keith Ellis, Paolo Nason, and Emanuele Re. Top-pair production and decay at NLO matched with parton showers. *JHEP*, 04:114, 2015.

- [99] Ryan Gavin, Ye Li, Frank Petriello, and Seth Quackenbush. FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order. *Comput. Phys. Commun.*, 182:2388, 2011.
- [100] Ye Li and Frank Petriello. Combining QCD and electroweak corrections to dilepton production in FEWZ. *Phys. Rev. D*, 86:094034, 2012.
- [101] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76(3):155, 2016.
- [102] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [103] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012.
- [104] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 09:111, 2009. [Erratum: JHEP02,011(2010)].
- [105] Michele de Gruttola, Caterina Vernieri, Pierluigi Bortignon, David Curry, Ivan Furic, Jacobo Konigsberg, Sean-Jiun Wang, Paolo Azzurri, Tommaso Boccali, Andrea Rizzi, Silvio Donato, Stephane Brunet Cooperstein , James Olsen, Christopher Palmer, Lorenzo Bianchini, Christoph Grab, Gael Ludovic Perrin, and Luca Perrozzi. Search for the Standard Model Higgs Boson Produced in Association with W and Z and Decaying to Bottom Quarks. [http://cms.cern.ch/iCMS/jsp/db\\_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2015/168](http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2015/168).

- [106] Michele de Gruttola, Rami Kamalieddin, Ilya Kravchenko, Lesya Shchutska. Search for resonant diHiggs production with bbZZ decays with the 2b2l2nu signature using 35.9/fb data of 2016 pp collisions at the LHC. [http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2017\\_198\\_v17.pdf](http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2017_198_v17.pdf).
- [107] Chris Palmer. VHbb Electron Trigger and ID+ISO SFs for 2016 data. [https://indico.cern.ch/event/604949/contributions/2543520/attachments/1439974/2216426/VHbb\\_TnP\\_SF\\_egamma\\_april.pdf#search=vhbb%20AND%20cerntaxonomy%3A%22Indico%2FExperiments%2FCMS%20meetings%2FPH%20%2D%20Physics%2FEgamma%22](https://indico.cern.ch/event/604949/contributions/2543520/attachments/1439974/2216426/VHbb_TnP_SF_egamma_april.pdf#search=vhbb%20AND%20cerntaxonomy%3A%22Indico%2FExperiments%2FCMS%20meetings%2FPH%20%2D%20Physics%2FEgamma%22).
- [108] CMS JetMET group. Jet Energy Resolution. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution>.
- [109] CMS MET group. MET Corrections and Uncertainties for Run-II. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETRun2Corrections>.
- [110] CMS MET group. MET Filter Recommendations for Run II. <https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2>.
- [111] CMS EGM POG. Multivariate Electron Identification for Run2. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MultivariateElectronIdentificationRun2>.
- [112] Helge Voss, Andreas Höcker, Jörg Stelzer, and Frerik Tegenfeldt. TMVA, the toolkit for multivariate data analysis with ROOT. In *XIth International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)*, page 40, 2007.

- [113] CMS Higgs WG. Documentation of the RooStats -based statistics tools for Higgs PAG. <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit>.
- [114] CMS Higgs WG. Binned shape analysis with the Higgs Combination Tool. [https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit#Binned\\_shape\\_analysis](https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit#Binned_shape_analysis).
- [115] bbbb team. Search for resonant pair production of Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at 13 TeV. <http://cms.cern.ch/iCMS/analysisadmin/get?analysis=HIG-17-009-pas-v5.pdf>.
- [116] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- $k_t$  jet clustering algorithm. *JHEP*, 04:063, 2008.
- [117] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.
- [118] C. Giunti and M. Laveder. Neutrino mixing. In F. Columbus and V. Krasnoholovets, editors, *Developments in Quantum Physics*. Nova Science Publishers, Inc., 2004.
- [119] Richard Phillips Feynman, Robert Benjamin Leighton, and Matthew Sands. *The Feynman lectures on physics; New millennium ed.* Basic Books, New York, NY, 2010. Originally published 1963-1965.
- [120] Savas Dimopoulos, Stuart Raby, and Frank Wilczek. Proton decay in supersymmetric models. *Physics Letters B*, 112(2):133 – 136, 1982.
- [121] David J Griffiths. *Introduction to elementary particles; 2nd rev. version.* Physics textbook. Wiley, New York, NY, 2008.

- [122] M. Della Negra, P. Jenni, and T. S. Virdee. Journey in the search for the higgs boson: The atlas and cms experiments at the large hadron collider. *Science*, 338(6114):1560–1568, 2012.
- [123] Jennifer Ouellette. Einstein’s quest for a unified theory. *APS*, 2015.
- [124] E A Davis and Isabel Falconer. *J.J. Thompson and the discovery of the electron*. Taylor and Francis, Hoboken, NJ, 2002.
- [125] Oreste Piccioni. *The Discovery of the Muon*, pages 143–162. Springer US, Boston, MA, 1996.
- [126] G. Danby, J-M. Gaillard, K. Goulian, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Phys. Rev. Lett.*, 9:36–44, Jul 1962.
- [127] M. L. Perl, G. S. Abrams, and et al Boyarski. Evidence for anomalous lepton production in  $e^+ - e^-$  annihilation. *Phys. Rev. Lett.*, 35:1489–1492, Dec 1975.
- [128] K. Kodama et al. Observation of tau neutrino interactions. *Phys. Lett.*, B504:218–224, 2001.
- [129] S. M. Bilenky. Neutrino in Standard Model and beyond. *Phys. Part. Nucl.*, 46(4):475–496, 2015.
- [130] S Chandrasekhar. *Newton’s principia for the common reader*. Oxford Univ., Oxford, 2003. The book can be consulted by contacting: PH-AID: Wallet, Lionel.
- [131] Hanoch Gutfreund and Jurgen Renn. *The road to relativity: the history and meaning of Einstein’s "The foundation of general relativity" : featuring the orig*

- inal manuscript of Einstein's masterpiece.* Princeton University Press, Princeton, NJ, Apr 2015.
- [132] J. Butterworth. *Smashing Physics*. Headline Publishing Group, 2014.
- [133] W N Cottingham and D A Greenwood. *An Introduction to the Standard Model of Particle Physics; 2nd ed.* Cambridge Univ. Press, Cambridge, 2007.
- [134] Eric W. Weisstein. Fundamental forces.
- [135] Carl Bender. Mathematical physics.
- [136] Andrew Wayne. QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga by Silvan S. Schweber. *The British Journal for the Philosophy of Science*, 46(4):624–627, 1995.
- [137] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.
- [138] Michelangelo L Mangano. Introduction to QCD. (CERN-OPEN-2000-255), 1999.
- [139] Matt Strassler. Of particular significance: Conversations about science with theoretical physicist matt strassler.
- [140] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.
- [141] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [142] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

- [143] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.
- [144] Pauline Gagnon. *Who cares about particle physics? : making sense of the Higgs boson, the Large Hadron Collider and CERN*. Oxford University Press, 2016.
- [145] Matthias U. Mozer. Electroweak Physics at the LHC. *Springer Tracts Mod. Phys.*, 267:1–115, 2016.
- [146] Roger Wolf. *The Higgs Boson Discovery at the Large Hadron Collider*, volume 264. Springer, 2015.
- [147] Steven Weinberg. The Making of the Standard Model. *Eur. Phys. J. C*, 34(hep-ph/0401010):5–13. 21 p. ; streaming video, 2003.
- [148] Jose Andres Monroy Montanez, Kenneth Bloom, and Aaron Dominguez. Search for production of a Higgs boson and a single Top quark in multilepton final states in pp collisions at  $\sqrt{s} = 13$  TeV, Jul 2018. Presented 23 Jul 2018.
- [149] Albert M Sirunyan et al. Search for Higgs boson pair production in the  $\gamma\gamma b\bar{b}$  final state in pp collisions at  $\sqrt{s} = 13$  TeV. 2018.
- [150] Matthew J. Dolan, Christoph Englert, and Michael Spannowsky. New Physics in LHC Higgs boson pair production. *Phys. Rev.*, D87(5):055002, 2013.
- [151] Peisi Huang, Aniket Joglekar, Min Li, and Carlos E. M. Wagner. Corrections to di-Higgs boson production with light stops and modified Higgs couplings. *Phys. Rev.*, D97(7):075001, 2018.
- [152] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.

- [153] Kunihito Uzawa, Yoshiyuki Morisawa, and Shinji Mukohyama. Excitation of Kaluza-Klein gravitational mode. *Phys. Rev.*, D62:064011, 2000.
- [154] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo. Phenomenology of the Randall-Sundrum Gauge Hierarchy Model. *Phys. Rev. Lett.*, 84:2080, 2000.
- [155] Michael Forger and Hartmann Romer. Currents and the energy momentum tensor in classical field theory: A Fresh look at an old problem. *Annals Phys.*, 309:306–389, 2004.
- [156] Chuan-Ren Chen and Ian Low. Double take on new physics in double Higgs boson production. *Phys. Rev.*, D90(1):013018, 2014.
- [157] Roberto Contino, Margherita Ghezzi, Mauro Moretti, Giuliano Panico, Fulvio Piccinini, and Andrea Wulzer. Anomalous Couplings in Double Higgs Production. *JHEP*, 08:154, 2012.
- [158] Gennadi Sardanashvily. *Noether’s theorems: applications in mechanics and field theory*. Atlantis studies in variational geometry. Springer, Paris, 2016.
- [159] Thomas Schörner-Sadenius. *The Large Hadron Collider: harvest of run 1*. Springer, Cham, 2015.
- [160] CERN. *Large Hadron Collider in the LEP Tunnel*, Geneva, 1984. CERN.
- [161] Lyndon R Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001. 164 p, 2008. This report is an abridged version of the LHC Design Report (CERN-2004-003).
- [162] Karsten Eggert, K Honkavaara, and Andreas Morsch. Luminosity considerations for the LHC. Technical Report CERN-AT-94-04-DI. CERN-LHC-Note-263. LHC-NOTE-263, CERN, Geneva, Feb 1994.

- [163] Oswald Gröbner. The LHC Vacuum System. (LHC-Project-Report-181. CERN-LHC-Project-Report-181):5 p, May 1998.
- [164] Thomas Lenzi. Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors. Master's thesis, U. Brussels (main), 2013.
- [165] Prospects for HH measurements at the HL-LHC. Technical Report CMS-PAS-FTR-18-019, CERN, Geneva, 2018.
- [166] Combination of searches for Higgs boson pair production in proton-proton collisions at  $\sqrt{s} = 13$  TeV. Technical Report CMS-PAS-HIG-17-030, CERN, Geneva, 2018.
- [167] The CMS collaboration. Missing transverse energy performance of the cms detector. *Journal of Instrumentation*, 6(09):P09001, 2011.
- [168] Erwin Schrödinger. *Statistical thermodynamics; 2nd ed.* Cambridge Univ. Press, Cambridge, 1952.
- [169] Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model. Technical Report CMS-PAS-HIG-14-009, CERN, Geneva, 2014.
- [170] Search for resonant double Higgs production with  $b\bar{b}ZZ$  decays in the  $b\bar{b}\ell\ell\nu\nu$  final state. Technical Report CMS-PAS-HIG-17-032, CERN, Geneva, 2018.
- [171] R. P. Feynman. The theory of positrons. *Phys. Rev.*, 76:749–759, Sep 1949.
- [172] Werner Herr and B Muratori. Concept of luminosity. 2006.

- [173] Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation / Charles W. Misner, Kip S. Thorne, John Archibald Wheeler.* W. H. Freeman San Francisco, 1973.
- [174] Charles W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation.* W. H. Freeman, San Francisco, 1973.
- [175] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.
- [176] A Zee. *Quantum Field Theory in a Nutshell.* Nutshell handbook. Princeton Univ. Press, Princeton, NJ, 2003.
- [177] Francis Halzen and Alan Douglas Martin. *Quarks and leptons: an introductory course in modern particle physics.* Wiley, New York, NY, 1984.