Top quark physics at the LHC with the CMS detector

Sergey Senkin

School of Physics

University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Science

Autumn 2013

Contents

1. The LHC and the CMS detector

1.1 The Large Hadron Collider

The LHC [1] is currently the largest and the most powerful particle accelerator ever built. It is installed in the 26.7 km tunnel that was originally constructed for the LEP accelerator in the 1980s. The tunnel lies at a depth of 45 m to 170 m underground between the Jura mountain and Lake Geneva, being the main part of the CERN accelerator complex.

The machine is designed to accelerate proton beams and provide collisions at a centre of mass energy of $\sqrt{s} = 14$ TeV. Unlike particle-antiparticle colliders, the LHC requires two rings with opposite magnetic dipole fields in order to maintain and collide two counter-rotating proton beams. Since the tunnel was originally designed for the electron-positron LEP, it has an internal diameter of 3.7 m which is not enough to install two separate independent rings. Therefore, a twin-bore magnet design was adopted [2], which resulted in substantial cost savings.

A schematic view of the LHC accelerator chain is shown in Figure 1.1. Initially, the protons are obtained by stripping orbiting electrons from hydrogen atoms. Then they are injected into the linear accelerator LINAC2 to reach the energy of 50 MeV and enter the Proton Synchrotron Booster (PSB). The booster accelerates them to 1.4 GeV and passes the beam to the Proton Synchrotron (PS) where the energy rises to 25 GeV. In the next step, protons enter the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Finally, the beam is transferred to the LHC in both clockwise and anti-clockwise directions where it takes about 20 minutes to reach the design 7 TeV energy (per beam).

[add the number of magnets, total energy stored, number of bunches, bunch spacing etc.?]

The LHC has four interaction points, providing collisions to four major experiments. Two of them, CMS and ATLAS, are multi-purpose high-luminosity experiments with a peak luminosity of $L = 10^{34}$ cm^{-2} s^{-1}. The other two experiments operate at low luminosities and have more specific physics goals: LHCb studies b-meson decays, and Alice is a dedicated heavy ion experiment.

The instantaneous luminosity of a collider can be calculated as

$$L = \frac{n_1 n_2 f}{4\pi\sigma_x\sigma_y}, \tag{1.1}$$

where n_1 and n_2 are the numbers of particles in each of the colliding bunches, f is the

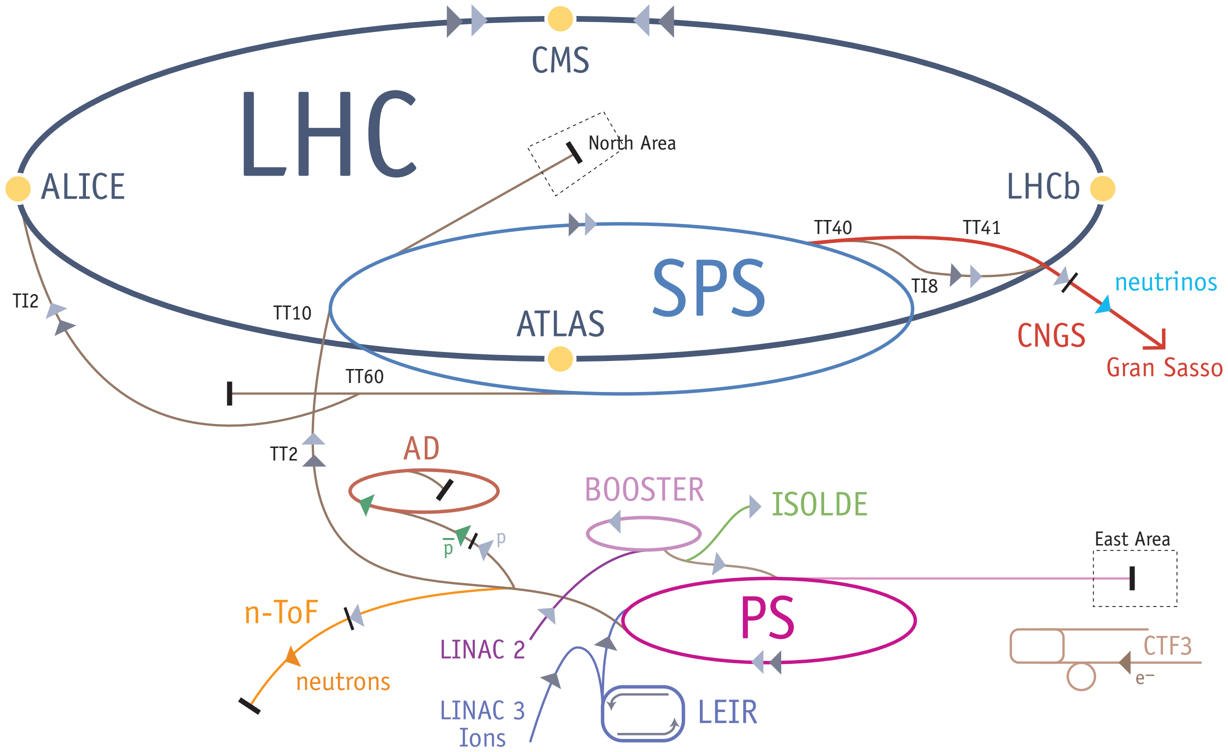

Figure 1.1: CERN accelerator complex.

revolution frequency, σ_x and σ_y are the horizontal and vertical beam sizes, assuming the two beams have the same size.

The number of events generated in the collisions per second is given by

$$N_{events} = L \times \sigma, \tag{1.2}$$

where σ is the cross section of the process under study.

[add the plots with cross sections and production rates?]

The LHC started operating on the 10th of September 2008, with the first beams fully circulating in both rings. However, only 9 days later a magnet quench occurred in two sectors of the tunnel, which was caused by an electrical fault due to a bad connection between two magnets. A consequent liquid helium explosion damaged a total of 53 superconducting magnets. Over a year was spent on repairs and tests, and the first collisions were recorded on the 23rd of November 2009 at a centre of mass energy of 0.9 TeV. The following few months showed the continuous ramp up of the beam energies up to 3.5 TeV per beam which was achieved on the 30rd of March 2010 when the LHC physics programme started.

Throughout the rest of 2010, the two general-purpose LHC experiments (CMS and ATLAS) recorded approximately 40 pb^{-1} of data, which resulted in the first measurements of various physics processes at the LHC. The following year became the main 7 TeV data-taking period, with about 5 fb^{-1} of data recorded by ATLAS and CMS. On the 5th of April 2012 the centre of mass energy was increased to 8 TeV, and July of 2012 marked the first major discovery of a new boson which was later shown to be consistent with the Standard Model Higgs boson, according to approximately 21.8 fb^{-1} of data recorded until early 2013. A long shut-down is planned for the following two years with various upgrades scheduled. The next physics run is expected in 2015 with the beam energy increased up to 6 or 7 TeV.

[add any upgrade details and distant future plans, like SLHC?]

1.2 The CMS Detector

The Compact Muon Solenoid [3] is a general-purpose detector designed to carry out precise measurements of the Standard Model and searches for physics beyond it. The primary design requirement was the ability to discover the nature of electroweak symmetry breaking, and the first observation of a Higgs boson was obtained in the Summer of 2012 [4].

The detector is installed at one of the LHC interaction points (Point 5) at about 100 m underground near the French village of Cessy, between the Jura mountains and Lake Geneva. The overall dimensions of the CMS detector are a length of 21.6 m, a diameter of 14.6 m and a total weight of 12 500 t.

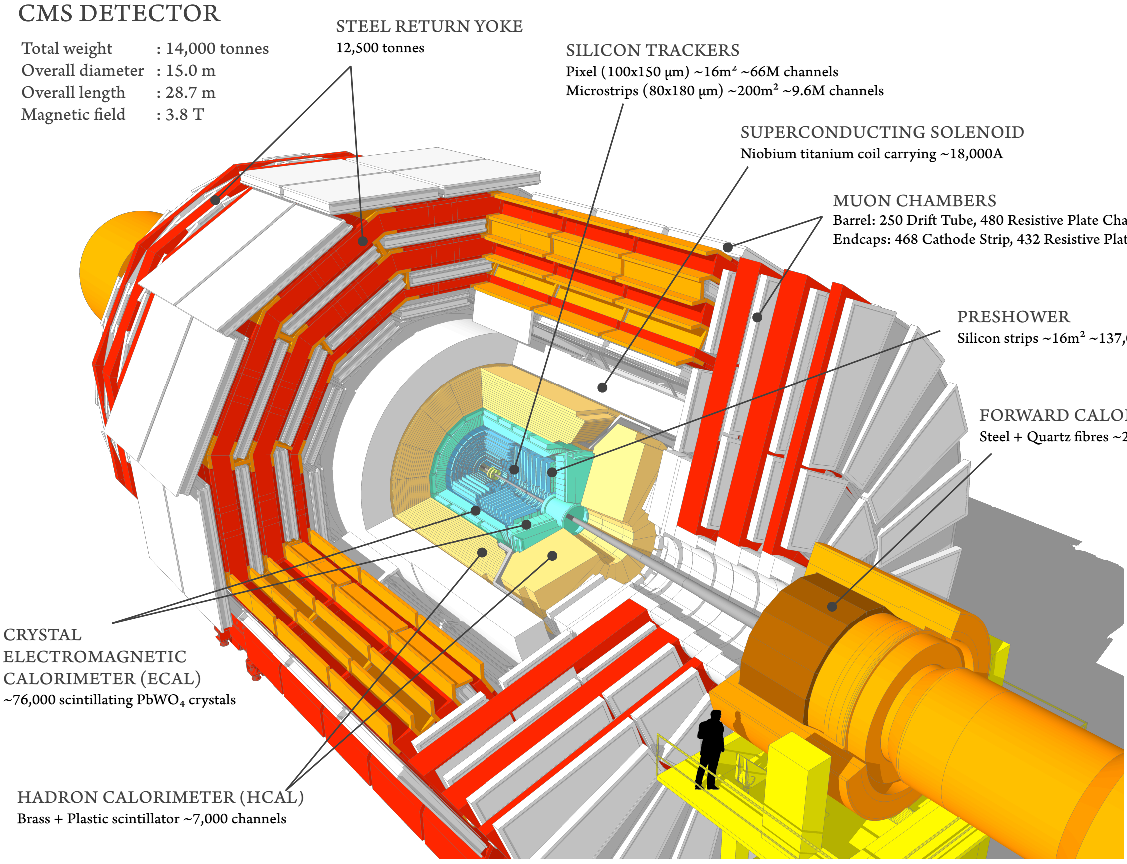

Figure 1.2: Sectional view of the CMS detector.

The sectional view of CMS is shown in Figure 1.2. In the centre of the detector, tracking and calorimetry systems are surrounded by the superconducting solenoid. On the outermost part of it the magnetic flux is returned through the iron yoke in which the muon system is also integrated. All the sub-systems are discussed in the following sections in more detail.

The cylindrical shape of the CMS detector dictates using a cylindrical coordinate system, with the origin centred at the interaction point, the x-axis pointing towards the centre of the LHC ring, the y-axis pointing upwards and the z-axis pointing along the beamline in the anti-clockwise direction. The azimuthal angle ϕ is measured from the x-axis in the transverse $(x-y)$ plane and the polar angle θ is measured from the z-axis. The radial distance to the beamline is denoted by r. Pseudorapidity is defined as:

$$\eta = -\ln \tan \frac{\theta}{2}. \tag{1.3}$$

This implies that the particles moving in the transverse plane (perpendicular to the beamline) have a pseudorapidity of 0, whereas the beam direction has an infinite pseudorapidity. Considering the cylindrical shape of the detector, it has barrel and endcap regions, with the transition occurring at $\eta \sim 1.4$. The momentum and energy transverse to the beamline are denoted by p_T and E_T respectively; the imbalance of the energy measured in the transverse plane, called missing transverse energy, is denoted by $E_{\mathrm{T}}^{\mathrm{miss}}$.

1.2.1 Inner Tracking System

The tracking system lies in the heart of the CMS detector and is the closest to the interaction point where the particle flux has the highest value. This imposes demanding requirements on the configuration of the system. At design luminosity of $L = 10^{34}$ cm^{-2} s^{-1} with the bunch spacing of 25 ns, an average of 1000 particles from about 25 proton-proton interactions (pile-up vertices) is expected to traverse the tracker for each bunch crossing. However, up until the long shutdown a bunch spacing of 50 ns was used, which meant a higher number of protons in each bunch leading to approximately twice the number of pile-up vertices. Therefore, in order for the particle tracks to be identified reliably and separately for each bunch crossing, the tracker requires very fine granularity and fast response parameters. Another complication caused by the intense particle flux is the severe radiation damage, so the tracker has to be highly resilient in operating in the harsh environment for a reasonable lifetime.

To meet these requirements on granularity, response time and radiation resilience, the tracker design was chosen to be based on silicon detector technology. Although capable of meeting such conditions, this technology has a disadvantage of a high power density of on-detector electronics. This implies the necessity of an efficient cooling system. Moreover, a large amount of dense material interacting with the particles leads to higher multiple scattering, bremsstrahlung, photon conversions and nuclear interactions. Therefore, there are complications in the reconstruction of the tracks, meaning some loss of efficiency and precision. This will be discussed in detail later on in the object reconstruction section.

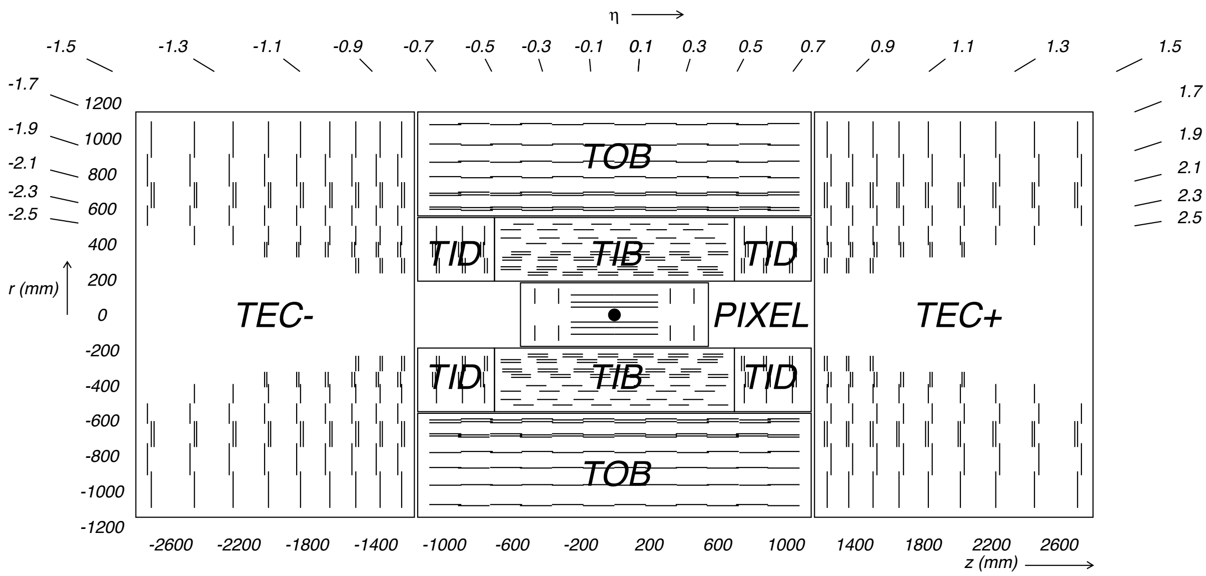

Figure 1.3: Cross-section of the CMS tracker system [3].

Figure 1.3 shows the overall layout of the tracking system. It consists of the inner pixel detector, located in the vicinity of the interaction point, and silicon strip tracker detectors: inner barrel and disks (TIB and TID), outer barrel (TOB) and endcaps (TEC). The geometrical acceptance of the tracker system goes up to $|\eta|< 2.5$. The outer radius of the CMS tracker reaches approximately 110 cm, and its total length is about 540 cm.

The pixel detector consists of three layers of pixel sensors at radii of 4.4 cm, 7.3 cm and 10.2 cm from the beamline in the barrel region. In addition there are two endcap disks on each side at $|z| = 34.5$ cm and 46.5 cm. The pixel size equals 100×150 μm^2 in $r\phi \times z$ coordinates. The pixel detector has 66 million pixels and the total area of about 1 m^2.

The silicon strip tracker consists of several layers of silicon microstrip detectors. It covers the region between 20 cm to 110 cm in radius and extends up to ± 280 cm in the z direction. The Tracker Inner Barrel (TIB) is made out of 4 layers and the Tracker Outer Barrel (TOB) has 6 layers in it. The tracker endcaps (TEC) comprise 9 disks, and there are also the tracker inner disks (TID) that consist of 3 disks filling the gap between TIB and TEC as shown in

Figure 1.3. There are 9.3 million silicon strips covering the area of about 200 m^2. The silicon sensors' thickness varies between 320 and 500 μm and the strip pitch varies from 80 μm in the TIB to 180 μm in TOB and TEC.

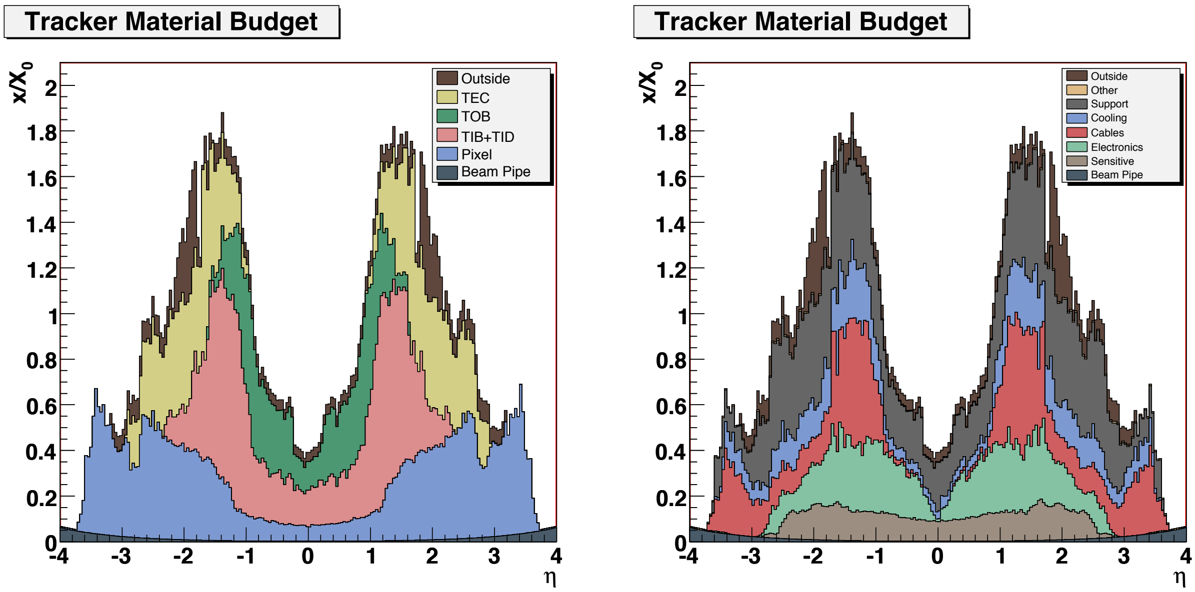

Figure 1.4: Material budget as a function of pseudorapidity η for the different sub-detectors of the tracker (left) and broken down into the functional contributions (right), in units of radiation length [3].

The silicon detectors of the tracker, the readout electronics and support structure form a considerable amount of material for the particles traversing from the interaction point. Figure 1.4 [3] shows the material budget of the CMS tracker in units of radiation lengths[1] (X_0). It grows from about 0.4 X_0 to 1.8 X_0 in the barrel region, and then decreases to about 1 X_0 in the endcaps. This causes a substantial conversion rate for photons and electrons in the tracker material; it also will be discussed in more detail in the electron reconstruction section.

[perhaps need to add the p_T resolution plots]

1.2.2 Electromagnetic Calorimeter

The next detector subsystem which is surrounding the tracker is the electromagnetic calorimeter, or ECAL. It is of a primary importance for the analyses described in this thesis,

[1]A material's radiation length is the mean distance over which a high-energy electron loses all but $1/e$ of its energy by bremsstrahlung; this is equal to 7/9 of the mean free path for pair production by a high-energy photon.

as it provides information for the electron and positron reconstruction. Combination of this information with that from the tracking system must ensure a precise measurement of electron position and momentum, and also sufficient background removal. It has to effectively distinguish the energy deposit shape of an electromagnetic particle from the one of a hadronic particle, which requires good segmentation and high resolution.

ECAL is a hermetic, high-granularity, high-resolution scintillating crystal calorimeter consisting of 61 200 lead tungstate ($PbWO_4$) crystals located in the central barrel region ($|\eta|< 1.479$), and 7324 crystals in each of the two endcaps ($1.479 <|\eta|< 3.0$). All crystals are followed by photodetectors reading and amplifying their scintillation: avalanche photodiodes (APD) are used in the barrel, and vacuum phototriodes (VPTs) are used in the endcaps. These different choices were caused by the configuration of the magnetic field and the expected level of radiation.

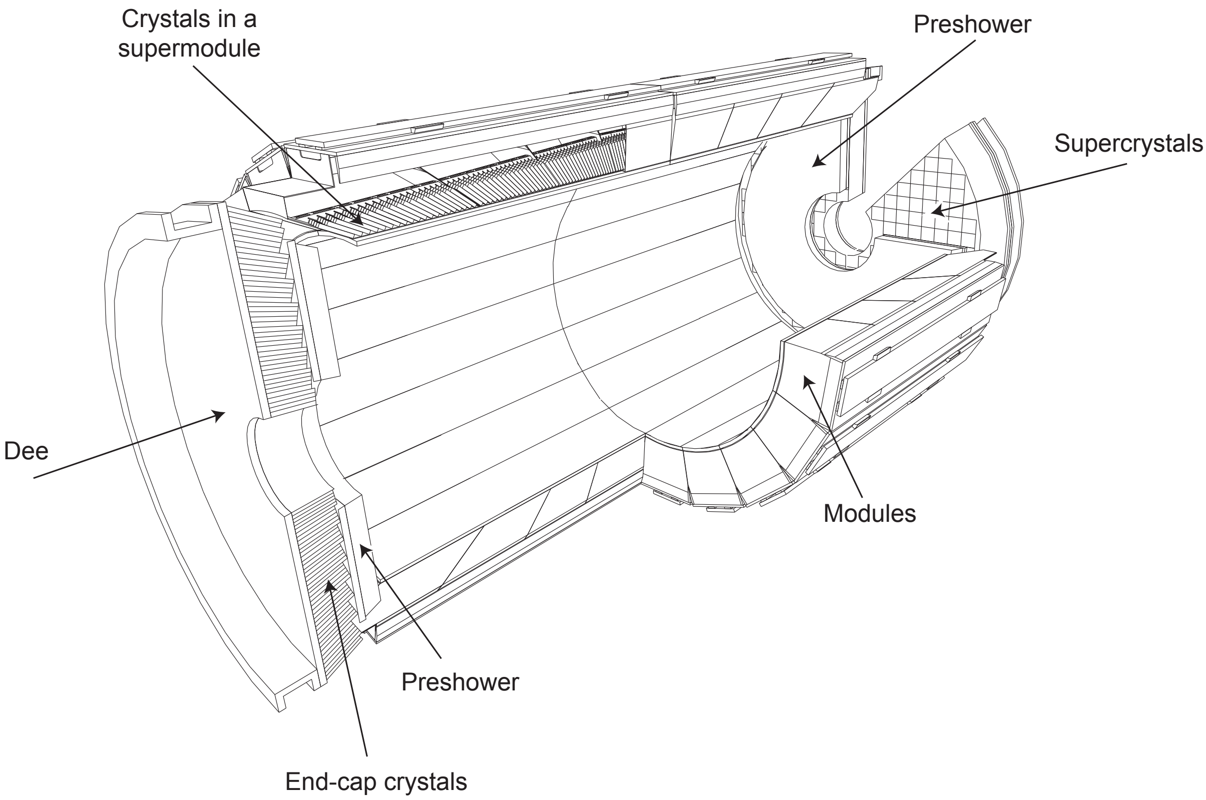

Figure 1.5: Layout of the CMS electromagnetic calorimeter [3].

The layout of the ECAL sub-detector is shown in Figure 1.5. An additional preshower detector is used in the endcap region to lower the required detector depth. Its principal aim is to identify neutral pions in the endcaps, but it also helps to distinguish neutral pions and electrons from minimum ionising particles and improves the position determination of electrons and photons with high granularity.

Table 1.1: ECAL crystal characteristics

	Barrel	Endcaps
number of crystals	61 200	14 648
crystal cross-section in (η, ϕ)	0.0174×0.0174	varies
crystal cross-section at the front	22×22 mm^2	28.62×28.62 mm^2
crystal cross-section at the rear	26×26 mm^2	30×30 mm^2
crystal length	230 mm ($25.8X_0$)	220 mm ($24.7X_0$)

The main geometrical characteristics of the ECAL crystals are shown in Table 1.1. The choice of lead tungstate was driven by the constraints of the CMS design. It is a very dense material (8.28 g/cm^3) with a short radiation length of $X_0 = 0.89$ cm, which allows the calorimeter to fit inside the compact magnet. Lead tungstate also has a small Molière radius[1] of 2.2 cm, which allows a calorimeter with fine granularity. Finally, the crystals emit 80 % of their scintillation light in just 25 ns, however the light yield is relatively low. At 18 °C, about 4.5 photoelectrons per MeV are collected. The dependence of the light yield on temperature requires a cooling system capable of keeping the crystal temperature stable within ± 0.05 °C to preserve energy resolution [5].

The energy-dependent resolution of the calorimeter can be parameterised as follows [3]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2. \tag{1.4}$$

where S is the stochastic term, N is the noise term, and C is the constant term. Figure 1.6 shows the energy resolution measured using incident electrons, during the beam tests in 2004.

1.2.3 Hadron Calorimeter

The hadron calorimeter (HCAL) is the next sub-detector located mostly inside the solenoid and completing the CMS calorimetry system. It is essential for the measurement of hadron jets and missing transverse energy.

[1]The Molière radius R_μ is a characteristic constant of a material giving the scale of the transverse dimension of the fully contained electromagnetic showers initiated by an incident high energy electron or

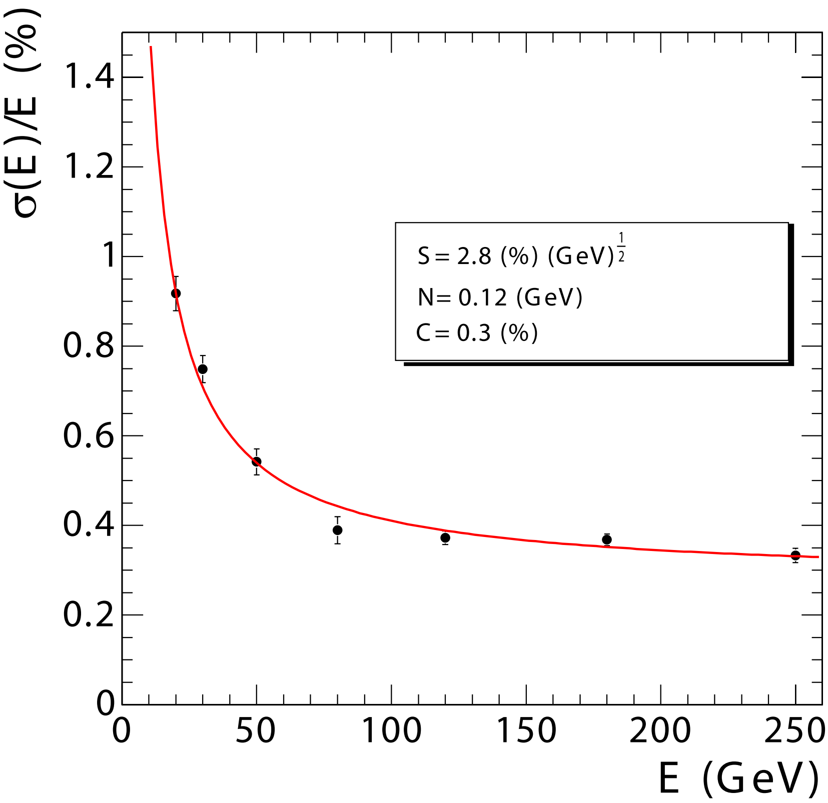

Figure 1.6: ECAL energy resolution derived from the test beam measurements as a function of deposited energy. The stochastic, noise, and constant contributions are shown [3].

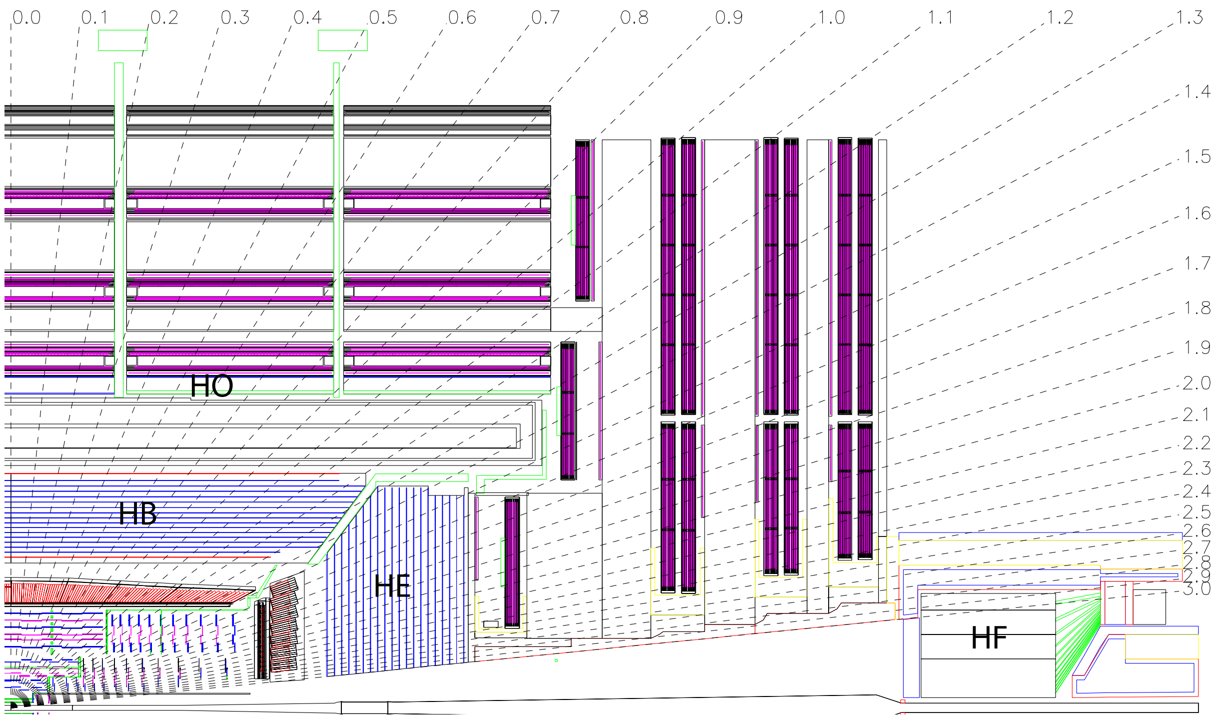

Figure 1.7: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [3].

As shown in Figure 1.7, HCAL consists of four subsystems: the hadron barrel calorimeter (HB), the hadron endcap calorimeter (HE), the hadron outer calorimeter (HO) and the hadron forward calorimeter (HF). The barrel and endcap parts (HB, HE) cover the pseudorapidity range up to $|\eta|< 3.0$, and the forward part (HF) extends it to a total coverage of $|\eta|< 5.0$. HCAL surrounds ECAL from its outer limit of 1.77 m from the beamline, to the inner limit of the magnet coil at 2.95 m from the beamline. However, due to space limitations the barrel calorimeters do not contain complete hadronic showers, therefore an outer calorimeter (HO) was designed to measure the energy leakage. It is placed in the muon system just outside of the solenoid in the barrel region.

HCAL is a sampling calorimeter consisting of alternating layers of brass and stainless steel absorbers, and plastic scintillators as active elements. The choice of the absorber material was caused by its short hadronic interaction length and its property of being non-magnetic, which is crucial in the strong magnetic field of the CMS magnet. The scintillation light is guided by embedded wavelength-shifting (WLS) fibres. The light from the WLS is then transmitted via a network of clear fibres, arranged in read-out towers, to hybrid photodiodes (HPDs) [3].

Both HB and HE scintillators have a granularity of $\Delta\eta\times\Delta\phi = 0.087 \times 0.087$ for $|\eta|< 1.6$, and $\Delta\eta\times\Delta\phi = 0.17 \times 0.17$ for $|\eta|\geq 1.6$. The tower segmentation of the forward calorimeter (HF) varies from $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$ at $|\eta|= 3.0$ to $\Delta\eta \times \Delta\phi = 0.3 \times 0.35$ at at $|\eta|= 5.0$. The HF is placed at about 11 m from the interaction point, and is essential to reconstruct very forward hadron jets. Together with HO, it provides the hermeticity of the calorimetry system, making it possible to measure the transverse missing energy to a reasonable precision.

[perhaps need to add the energy resolution plots]

1.2.4 Superconducting Magnet

The superconducting solenoid is a central feature of the CMS apparatus, essentially giving it its name. The magnet has a length of 12.5 m, diameter of 6.3 m and mass of 220 t. Although it was initially designed to sustain a uniform magnetic field of 4 T within the 5.9 m diameter free bore, operation at 3.8 T was chosen in order to increase the lifetime. The magnetic field is returned by a massive iron yoke. The main parameters of the CMS magnet are shown in Table 1.2.

The large bending power of the solenoid is required to bend the tracks of high energy charged particles to an extent where good momentum resolution is achieved. The design

photon. It is defined as the radius of a cylinder containing an average of 90 % of the shower's energy deposition.

Table 1.2: Parameters of the CMS superconducting solenoid [5] [6].

Field	3.8 T
Inner Bore	5.9 m
Length	12.5 m
Number of Turns	2168
Current	18 160 kA
Stored energy	2.3 GJ

requirement for the strength of the magnetic field was the ability to unambiguously determine the sign of the electric charge for muons with a momentum of $\approx$ 1 TeV/c [5].

The solenoid coil is constructed from four layers of superconducting high-purity niobium-titanium cable co-extruded with pure aluminium, which acts as a thermal stabiliser. The cold mass is cooled down to 4.5 K by liquid helium. If a fast discharge happens (e.g. caused by a magnet quench), about 3 days are necessary to re-cool the coil.

1.2.5 Muon System

The last sub-detector placed on the outermost part of CMS is the muon system. Since the muons are the most penetrating particles detectable by CMS, they have the cleanest signature and play an important role in many physics analyses. Due to their ability to travel through the many layers of the calorimeters, muons are relatively easy to identify and separate from the background.

The layout of the CMS muon system is shown in Figure 1.8. It consists of the drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The entire system surrounds the solenoid and covers the pseudorapidity region of $|\eta| < 2.4$.

The drift tubes are located in the barrel region ($|\eta| < 1.2$). Consisting of four stations, they form concentric cylinders around the beam line; there are 250 drift chambers with about 172 000 sensitive wires in total. When a muon passes through the volume, it knocks electrons off the atoms of the gas, which then follow the electric field and reach the positively-charged wires, providing information on the muon's position. The chambers are filled with the gas mixture of 85 % Ar and 15 % CO_2, where the muon drift time does not exceed 380 ns. Although this value is bigger than the typical bunch crossing time (25 or 50 ns), it is sufficient because of the small muon rate in this region.

In the endcaps, the cathode strip chambers cover the pseudorapidity region of $0.9 < |\eta| < 2.4$. Each of 468 CSCs is a trapezoidal multi-wire proportional chamber consisting of 6 gas gaps with a plane of radial cathode strips and a plane of anode wires which are roughly per-

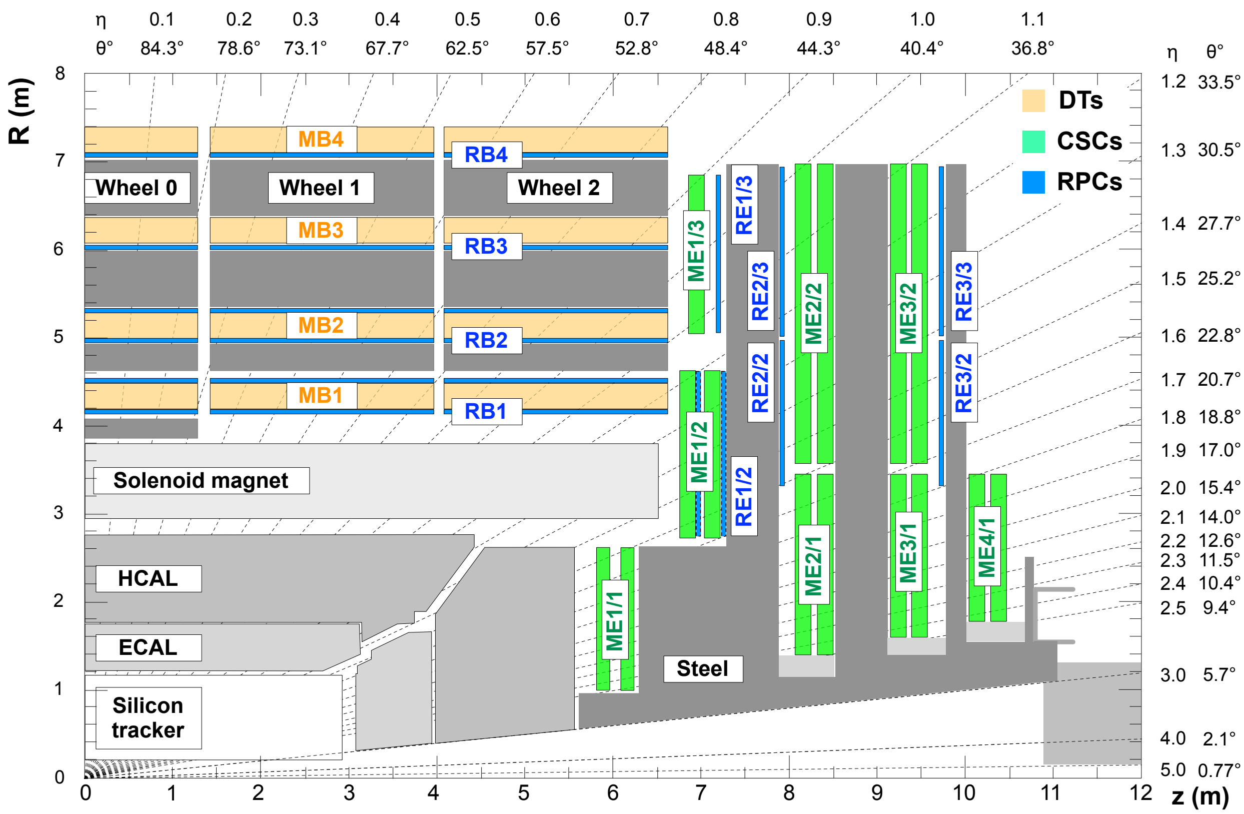

Figure 1.8: Layout of one quarter of the CMS muon system. Four drift tube (DT, in light orange) stations are labeled MB (muon barrel) and the cathode strip chambers (CSC, in green) are labeled ME (muon endcap). Resistive plate chambers (RPC, in blue) are in both the barrel and the endcaps of CMS, where they are labeled RB and RE, respectively.

pendicular. A charged muon traversing each plane of a chamber causes gas ionisation and a subsequent electron avalanche which produces a charge on the anode wire and an image charge on the cathode strips. The gas used in CSCs is a mixture of Ar, CO_2 and CF_4.

The resistive plate chambers system is complementary to both DT and CSC systems, and is located in both barrel and endcap regions ($|\eta| < 2.1$). RPCs also operate in avalanche mode with a gas mixture of $C_2H_2F_4$, C_4H_{10} and SF_6, and due to an excellent time resolution of about 1 ns they provide fast information for triggering. The spacial resolution is, however, quite limited (≈ 1 cm, compared to ≈ 100 µm for DTs and CSCs).

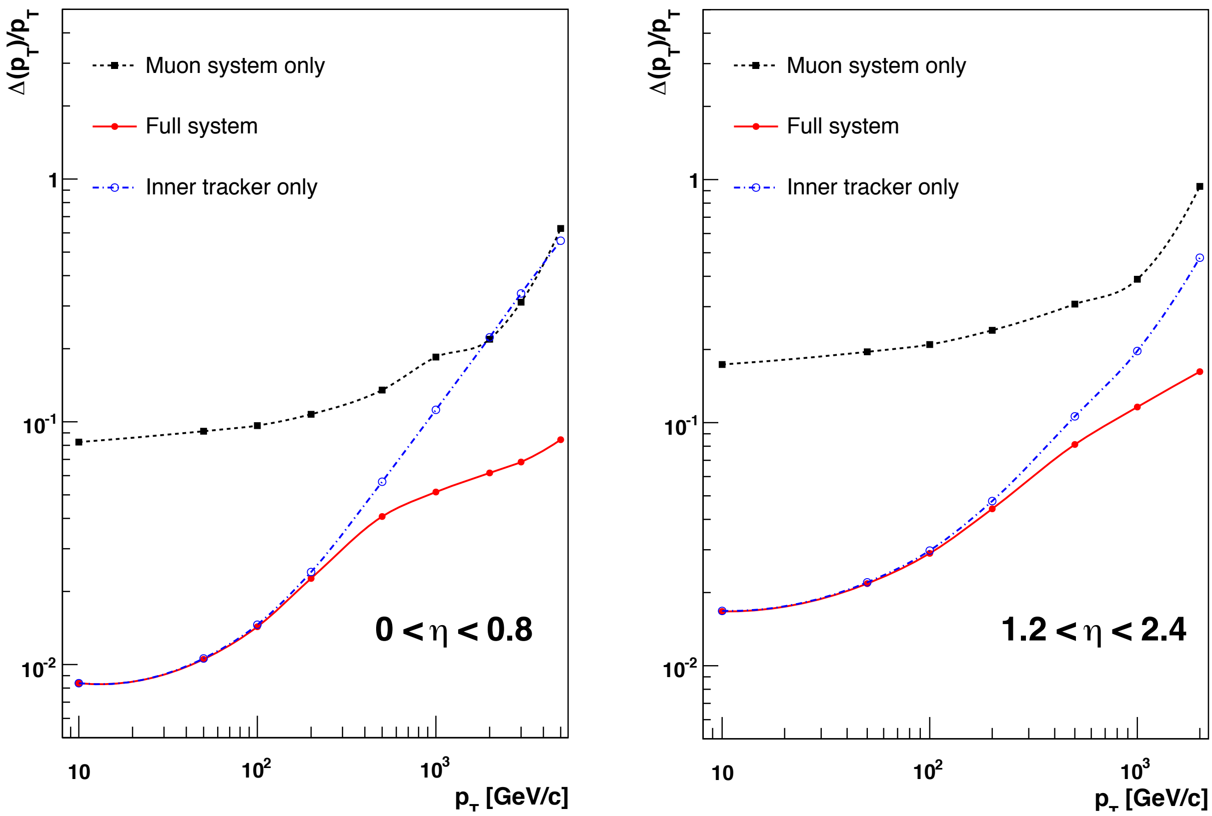

Figure 1.9: The muon transverse momentum resolution as a function of the transverse momentum (p_T) using the muon system only (black), the inner tracking only (blue), and both (red), in regions of $|\eta| < 0.8$ (left) and $1.2 < |\eta| < 2.4$ (right) [3].

The muon momentum is measured in both the tracker and the muon system. As it can be seen on Figure 1.9, both sub-systems contribute to the momentum resolution at different p_T values. This happens due to the difference in the magnetic field and detector technology. For low-p_T muons, the best momentum resolution is obtained in the tracker, whereas in the high-p_T region the muon system provides a significant improvement. Therefore, by using information from both the silicon tracker and the muon chambers (i.e. reconstructing the "global muon"), the momentum resolution is improved in the whole p_T region up to a ≈ 1 TeV/c level.

1.2.6 Trigger and Data Acquisition

At design LHC luminosity of $L = 10^{34}$ cm^{-2} s^{-1}, approximately 25 collisions are expected to occur at each crossing of the proton bunches. The bunch spacing of 25 ns corresponds to a crossing rate of 40 MHz. Since every event produces ~ 1 MB of raw data, it corresponds to a total data production of 40 TB s^{-1}. Attempting to store all of this data is clearly beyond the available technology. Moreover, only a fraction of events contain hard scattering processes that are of interest, therefore an effective trigger system had to be implemented.

The CMS trigger is a two-level system, consisting of two independent parts: the Level-1 (L1) trigger and the High-Level Trigger (HLT). The L1 trigger is a hardware system implemented in programmable electronics residing partly on detector, and partly in the underground control room located at approximately 90 m from the experimental cavern. The maximum latency between the collision and the L1 accept decision received by front-end electronics is 3.2 µs. During this amount of time, the complete event information is buffered in pipelined memories on the detector. The only information used for the L1 trigger decision is that from the muon system and the calorimetry. Since the reconstruction of tracks exceeds the time scale required for the L1 decision, the tracker information can't be used. The L1 trigger reduces the event rate from ~ 40 MHz to ~ 100 kHz, corresponding to a data flow of about 100 GB s^{-1}. These events are fed into the HLT system.

The High-Level Trigger is a software system implemented in a single CPU farm, sometimes referred to as the "Event Filter Farm". Having access to the full event information, customised algorithms of increasing complexity are used which results in a highly flexible trigger system. The event rate is reduced down to ~ 300 Hz, with the final data rate of approximately 300 MB s^{-1} being stored on a large disk cache at the experimental site (the Storage Manager) and later on transferred to CERN Tier 0 for further processing (see Section 1.3).

Since the start of the LHC running, the operating conditions have been changing drastically. During the start-up year of 2010, the instantaneous luminosity went up from about 10^{27} cm^{-2} s^{-1} to approximately 0.2×10^{32} cm^{-2} s^{-1}. In 2011 the luminosity ramped up to a factor of 20 above that of 2010, reaching approximately 4×10^{33} cm^{-2} s^{-1}. This required a lot of continuous effort to control the trigger rates at reasonable level, whilst also keeping its efficiency acceptable. In 2012 the luminosity was more stable, peaking at $\approx 7.6 \times 10^{33}$ cm^{-2} s^{-1} which is just a factor of 2 above the 2011 values. However, it still came as a challenge because of the impact of pile-up. At a bunch spacing of 50 ns and increased centre-of-mass energy of 8 TeV, the average number of pile-up vertices nearly doubled comparing to that in 2011, which required a major CPU extension and implementation of sophisticated PU mitigation techniques at the HLT level. The author's contribution to the HLT development of the trigger paths important for top physics is described in Chapter 2.

1.3 Computing

The vast amounts of data delivered by the CMS detector impose high requirements on the offline computing system. During 2010–2012 operation, CMS collected ~ 10 PB of raw data per year. Including Monte Carlo simulations, reconstructed data and analysis skims, the total annual amount of data essentially doubles. To handle the distributed storage and processing of this data, not just for CMS but for the entire high energy physics community using the LHC, a worldwide LHC computing grid (WLCG) has been put in place.

WLCG is a global collaboration of more than 150 computing centres in about 40 countries. The grid has a tiered architecture, comprising 4 tiers with different resources and services. The first one, Tier 0, is based at CERN and is responsible for data-taking. It accepts raw data from the data acquisition system and repacks it into primary datasets according to the trigger information. The raw data is archived to tape, and is also prompt-reconstructed (within 48 hours) before being distributed to the Tier 1 (T1) centres around the world. There are 8 T1 sites based at large national laboratories in collaborating countries (e.g. RAL in the UK and FNAL in the US). Each of the T1 centres is used for large-scale centrally organised data-processing activities. The data is then distributed in the reduced format (see Section 1.3.1) to a more numerous set of Tier 2 centres, typically located at collaborating universities. Each of these centres is used for the grid-based analysis and Monte Carlo simulation for the whole experiment, as well as local services for groups maintaining them. The last stage of computing system, Tier 3, is meant solely for the local institution's user analysis.

1.3.1 Event Data Model

In the basis of the CMS Event Data Model lies the concept of an event, which is physically a result of a single collision in the LHC. From a software point of view, the event is a C++ object container storing raw data from a single readout of detector electronics (e.g. hits in various sub-detectors), as well as reconstructed data which is based on this information, such as tracks, clusters and physics objects. All these C++ objects are stored in ROOT format [7].

The EDM makes use of three main data formats, based on different levels of detail and precision:

- RAW format, containing full information from the detector as well as L1 and HLT trigger decisions, with the event size of ~ 1.5 MB.
- RECO (reconstructed data) format, which is obtained from raw data by application of pattern recognition and compression algorithms. This data includes reconstructed detector hits, clusters and physics objects (electrons, muons, etc.). The typical event size is ~ 250 kB.
- AOD (Analysis Oriented Data) format, produced by filtering the RECO data from the reconstructed detector objects, leaving just the high-level physics objects required for analysis. The event size is reduced down to ~ 50 kB.

The RECO and AOD data are analysis-ready data formats, produced centrally and used by many physics analysis groups. However, further simplification of the data is also a common practice. By transforming the C++ objects produced by CMS software into plain basic types or vectors of them, only including the analysis-specific content, the event size can be reduced down to ~ 3 kB level depending on the needs of a particular analysis. This data format is often referred to as private "ntuples", and it requires specific analysis software capable of restructuring the data into user-defined classes. By following this approach, the analysis can be run locally and generally much faster than processing the RECO or AOD data. However, it requires "ntuplising" this data every time when new centrally-recommended physics objects or corrections are produced.

1.3.2 Analysis Software

Both of the analyses described in this thesis use the CMS software framework[1] (CMSSW), as well as Bristol Analysis Tools[2] (BAT). The differential cross section analysis also uses an additional level of python scripts for post-processing[3].

CMSSW is the key CMS software framework built around the Event Data Model (see Section 1.3.1). The framework is essential for purposes of Monte Carlo simulation, detector calibration and alignment, as well as data reconstruction and analysis. CMSSW has a modular architecture, consisting of one configurable executable (cmsRun) and a large set of plug-in modules that contain all the code needed for event processing (reconstruction algorithms, calibration, etc.). Different versions of CMSSW were used for different analyses:

[1] http://cms-sw.github.io/cmssw

[2] https://github.com/BristolTopGroup/AnalysisSoftware

[3] https://github.com/BristolTopGroup/DailyPythonScripts

- `CMSSW_4_2_8` for the top mass analysis on 2011 data;
- `CMSSW_4_4_4` for the missing transverse energy analysis on 2011 data;
- `CMSSW_5_3_9` for the top cross pair cross section analysis on 2012 data.

Corresponding versions were used to produce ntuples for processing by BAT, which was used to read the data, apply selections, calculate high-level variables and to create various histograms of distributions. BAT was originally started in 2010 by Dr. Lukasz Kreczko for the needs of the Bristol top group, later on also developed by the author and other researchers from Bristol and affiliated top groups. Like CMSSW, this framework has a modular structure, with its classes falling in four main categories:

- readers, for translating plain data types from ROOT files into C++ objects;
- RECO objects, i.e. output of the readers (physical objects like leptons, jets and its collections);
- selections, for application of event selections;
- analysers for creating histograms, applying selections, algorithms, and filling histograms.

All analysers are independent from each other, making the analysis chain stable and reliable. The final set of python scripts is used to prepare the histograms, perform fitting and unfolding procedures (in case of cross section analysis), and producing final tables and plots. Rootpy[1] package was used to access ROOT libraries in python interface, and matplotlib[2] was used to create plots.

1.4 Object Reconstruction

Most CMS analyses, including the ones described in this thesis, adopt a reconstruction technique called Particle Flow (PF) [8]. This algorithm is used to obtain a global event description at level of individually reconstructed particles by means of combining information coming from all sub-detector systems. The ultimate goal is to determine type, energy and momentum of all the particles in the event with highest possible precision and in the most optimal way. The types of these particles include electrons, muons, charged hadrons, neutral hadrons and photons. All these particles are then used to reconstruct jets (Section 1.4.3), missing transverse energy (Section 1.4.1) and tau leptons from their decay products.

[1]http://rootpy.org/
[2]http://matplotlib.org/

1.4.1 Electron Reconstruction

The reconstruction of the $t\bar{t}$ pair with an electron in the final state imposes high requirements on the electron identification and its energy-momentum measurement, precision of which is of major importance for both top mass and $t\bar{t}$ cross section measurements.

Although the CMS detector is equipped with highly accurate ECAL and tracker systems, electron identification and reconstruction is still a challenging task due to the large amount of tracker material (see Section 1.2.1). This results in a significant Bremsstrahlung photon emission, which often causes an ECAL energy deposit to be widely spread in azimuthal direction because of the high magnetic field. Therefore, dedicated algorithms were developed in order to collect all Bremsstrahlung energy deposits in the calorimeter (Bremsstrahlung recovery), and also to take into account the kinks in the electron trajectory caused by photon emissions.

Electron reconstruction in CMS has following distinct stages: seeding, track finding, pre-identification, Bremsstrahlung recovery, track-cluster linking and final identification. Historically, the original seeding algorithm was designed and optimised for isolated high-p_T electrons. This approach starts from ECAL clusters, and therefore is called the 'ECAL-driven' seeding. It is based on the property of the ECAL energy deposits to have narrow width in the η coordinate, and to be widely spread in ϕ (azimuthal direction) like it was mentioned above. The electron and all the associated Bremsstrahlung energy deposits form a single "super-cluster", and the ability to correctly determine it affects the overall performance of this method. Only the super-clusters with transverse energy above 4 GeV are taken into account. Super-clusters are then matched to pairs or triplets of hits in the inner tracker layers, forming the track seeds on which the electron tracks are built upon.

The performance of the ECAL-driven method is not very well suited for non-isolated and low-p_T electrons. This occurs mainly due to the fact that the super-cluster position and energy can be highly biased by the impact of overlapping particles, especially if the electron happens to be within a jet and therefore non-isolated. Also, high track multiplicity complicates the backward propagation from the super-cluster, because it can be consistent with a number of track seeds corresponding to other particles. To minimise the number of these fake seeds, the ratio between the HCAL and ECAL energy deposits (H/E) is required to be smaller than 0.15. The HCAL towers used in the calculation of this ratio are taken within a cone of $\Delta R = 0.3$ behind the super-cluster position. Although this helps to keep the fake seed rate under control, the efficiency for non-isolated electrons is rather limited. As for the low-p_T electrons, the wider azimuthal spread of the Bremsstrahlung photons leads to poorer reconstruction of the super-cluster, biasing its position and therefore preventing the efficient matching with a track seed.

Within the particle flow method, efficient reconstruction of non-isolated and low-p_{T} electrons is particularly important since it affects the reconstruction of jets and missing transverse energy. Therefore, a different ('tracker-driven') seeding algorithm is used, which starts from reconstruction of the tracks. The baseline of the CMS track reconstruction is the Kalman filter (KF) [9], which is a linear least-squares estimator based solely on Gaussian probability density functions. It is particularly suitable for muon reconstruction, since it is dominated by multiple Coulomb scattering and its impact is well modelled by Gaussian fluctuations. However, this approach usually fails for electrons, because Bremsstrahlung photon emission is highly non-Gaussian. To accommodate for the resulting kinks in the electron trajectory, the Gaussian-Sum Filter (GSF) [10] is used, which is essentially a non-linear generalisation of the Kalman Filter. In this method, Bremsstrahlung energy loss is modelled by a Gaussian mixture, therefore GSF track fit provides a better estimate for the inner and outer track momentum, comparing to the KF algorithm. The downside of this approach is its high CPU usage, which means it can be run on a limited number of seeds.

The GSF tracks are reconstructed upon all the ECAL-driven seeds. In case of the 'tracker-driven' seeds, a pre-identification based on high-purity KF tracks has been adopted. This procedure starts with the tracks reconstructed with very tight criteria, thus decreasing the fake rate yet compromising on tracking efficiency. Then an iterative-tracking strategy is carried out by means of removing hits unambiguously assigned to the tracks from the previous iteration, and progressively relaxing track seeding criteria. This approach leads to both high efficiency and low fake rate, which is crucial for low-p_{T} and non-isolated electrons.

In the next step, track-cluster matching has to be performed. In case if electron has negligible Bremsstrahlung emission, the track is well reconstructed with the KF algorithm all the way to the ECAL internal surface, where the closest cluster is matched to the track. The corresponding cluster energy is compared with the track momentum, and if the ratio (E/p) is close to unity, the track is selected. On the contrary, if the electron experiences a significant Bremsstrahlung emission, other track characteristics have to be exploited. In this case a selection based on the number of hits and the χ^2_{KF} of the KF fit is applied before running a GSF refit. Finally, the number of hits, the GSF refit χ^2_{GSF}, $\chi^2_{\mathrm{KF}}/\chi^2_{\mathrm{GSF}}$ ratio, the energy loss measured by the track and the quality of the ECAL cluster-track matching are fed into a multivariate analysis using a Boosted Decision Trees (BDT) estimator.

Both tracker-driven and ECAL-driven seeds are used in order to obtain the GSF track collection of electron candidates.

1.4.2 Muon Reconstruction

1.4.3 Jet Reconstruction

1.4.4 Missing Transverse Energy

1.5 Summary

2. High level trigger development for Top Physics

References

[1] L. Evans and P. Bryant. LHC Machine. *JINST*, 3(08):S08001, 2008. 1

[2] J.P. Blewett. 200 GeV intersecting storage accelerators. In *Proceedings of The 8th International Conference on High-Energy Accelerators*, page 501, Geneva, Switzerland, 1971. CERN. 1

[3] The CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3(08):S08004, Aug 2008. 3, 6, 7, 8, 9, 10, 11, 12, 15

[4] The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. 3

[5] The CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006. 9, 13

[6] The CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *Journal of Instrumentation*, 5(03), 2010. 13

[7] Rene Brun and Fons Rademakers. ROOT – An object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(12):81 – 86, 1997. New Computing Techniques in Physics Research V. 17

[8] The CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009. 19

[9] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(23):444 – 450, 1987. 21

[10] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9, 2005. 21