

# Chapter 1

## The Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) detector sits at point 5 of the LHC ring, diametrically opposite the ATLAS detector at point 1. It is a  $4\pi$  hermetic general purpose detector, meaning that it has the capability to detect charged and neutral hadrons, photons, electrons, muons, taus, neutrinos, and non-Standard-Model particles predicted to escape the detector with good efficiency over a large range of rapidity. Its main distinguishing feature is a superconducting solenoid that provides a 3.8T magnetic field parallel to the beam line. This strong magnetic field allows precise determination of the momentum and charge of muons and electrons up to a momentum of  $\sim 1$  TeV.

The origin of the CMS coordinate system is at the nominal interaction point. The  $y$ -axis points skyward, the  $x$ -axis points towards the center of the LHC ring, and the  $z$ -axis points counterclockwise along the LHC ring.  $r$  denotes radial distances from the beam line,  $\phi$  is the azimuthal angle measured with respect to the positive  $x$ -axis, and  $\theta$  is the polar angle measured with respect to the positive  $z$ -axis. The *pseudorapidity*  $\eta$  is defined as  $\eta = -\ln \tan(\theta/2)$ , and is a good approximation to

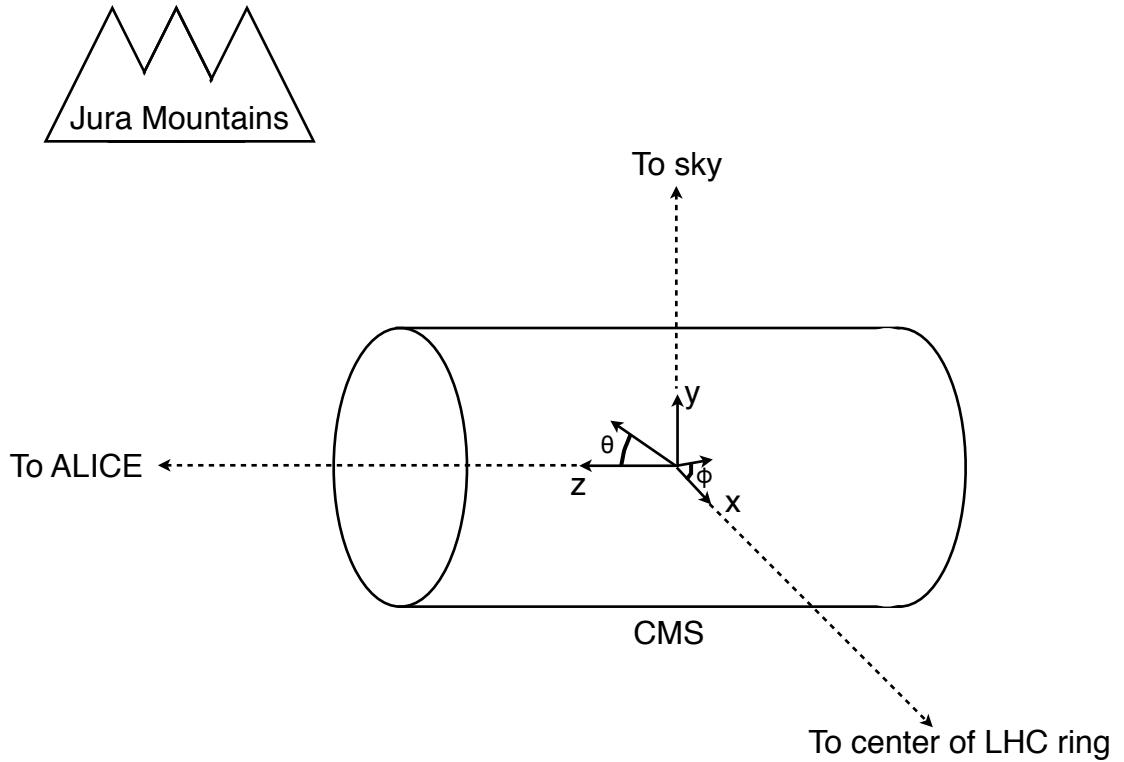


Figure 1.1: CMS coordinate system.

rapidity  $y = (1/2) \ln((E + p_z c)/(E - p_z c))$  for relativistic particles. The transverse momentum and energy ( $p_T$  and  $E_T$ ) of a particle are defined as  $p_T = p \cos \phi$  and  $E_T = E \cos \phi$ , where  $p$  and  $E$  are the magnitude of the particle's momentum vector and the particle's total energy, respectively. A depiction of the CMS coordinate system is shown in Figure 1.1.

The CMS sub-detectors are arranged in concentric cylindrical layers, plus “end-caps,” around the beam line, as shown in Figure 1.2. Closest to the beam line are three layers of silicon pixel detectors, with the innermost at radius 4.4 cm and outermost at radius 10.2 cm [?]. Including the pixel endcaps, the total pixel coverage

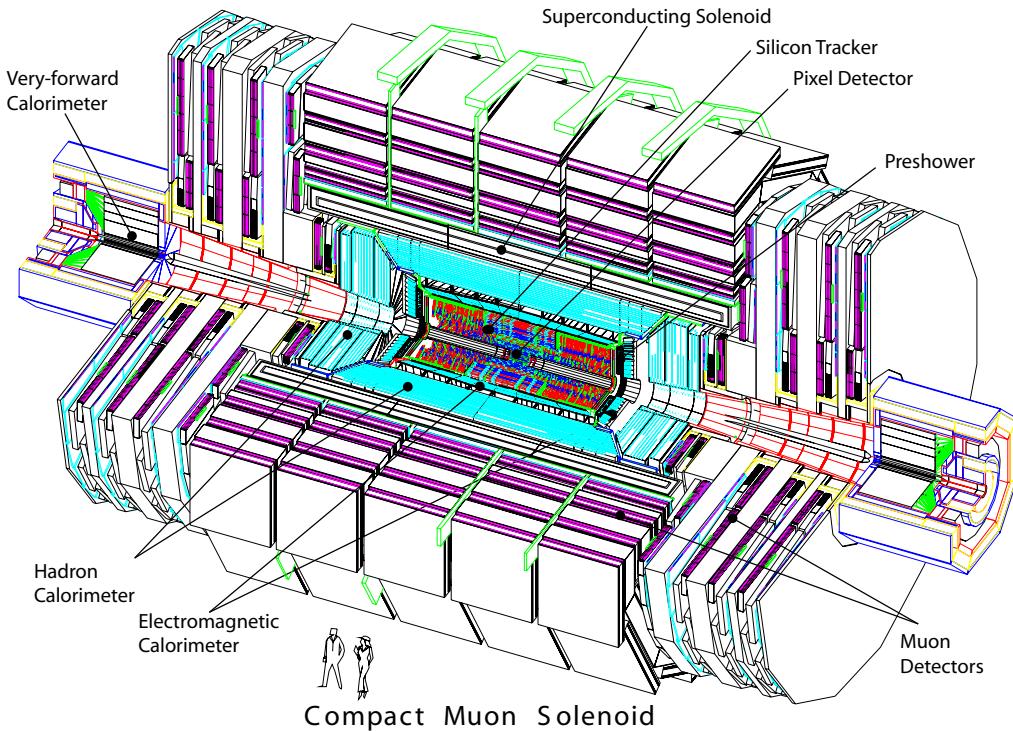


Figure 1.2: Cutaway view of CMS. Reprinted from Fig. 1.1 of ref. [?].

extends to  $\eta = 2.5$  [?]. The pixel detector plays an important role in determining the proton-proton interaction position (*beam spot*) and the impact parameters of charged particle trajectories, and is critical for the measurement of decay positions some distance from the beam spot (*displaced vertices*), such as those due to the showering and hadronization of a  $b$  quark.

The 10 next layers of CMS are comprised of silicon microstrip detectors, with the outermost layer at a radius of 1.3 m from the beam line [?]. As for the pixel detectors, the silicon strip endcaps extend tracking coverage to  $\eta = 2.5$ . The silicon microstrip layers are the workhorse of the CMS tracking system, and provide excellent charged particle momentum resolution and track finding efficiency.

Outside the tracking detectors are the calorimeters, starting with the single-layer lead tungstate crystal electromagnetic calorimeter at a radius of 1.3 m from the beam line (location of crystal front faces) [?]. Each crystal is 23 cm long, corresponding

to 25.8 radiation lengths ( $X_0$ ) [?]. The crystal dimensions are such that most of one electromagnetic shower, and no more, can be contained in a single crystal, leading to excellent energy resolution for photons and electrons. The electromagnetic calorimeter radial and endcap layers cover a pseudorapidity range up to 3.0. A lead/silicon sampling calorimeter sits in front of the crystal endcaps to provide better rejection of neutral pions.

The last layer of calorimetry inside the solenoid is the brass/scintillator sampling hadronic calorimeter, which has a radial extent from 1.77-2.95 m [?]. The hadronic barrel and endcap calorimeters cover up to  $|\eta| = 3.0$ , while the iron/quartz-fiber forward hadronic calorimeter covers the region  $3.0 \leq |\eta| \leq 5.2$ .<sup>1</sup> There is one more layer of hadronic calorimetry outside the solenoid in  $|\eta| < 1.3$  which, together with the layers inside the solenoid, provides approximately 12 hadronic interaction lengths of instrumented absorber. Because of its large  $|\eta|$  coverage and depth, the hadronic calorimeter provides good missing transverse energy resolution and accurate measurements of high energy jets.

The iron return yoke of the solenoidal magnetic field is interleaved with muon detectors from 4.1-7.4 m in  $r$  and 6.6-10.6 m in  $z$ , providing muon detection up to  $|\eta| = 2.4$  [?]. In the barrel region of  $|\eta| < 1.2$ , drift tubes are used to read out the muon tracks, while in the endcaps cathode strip chambers are used. Due to their speed, resistive plate chambers are used throughout the muon system to provide an independent trigger and timing measurement. Combining the tracker and muon system hits, the momenta and charge of muons up to  $p_T = 1$  TeV can be precisely reconstructed.

A longitudinal quarter cross-sectional view of CMS is shown in Figure 1.3. The remainder of this chapter is devoted to explaining the CMS subdetectors and readout

---

<sup>1</sup>The Centauro and Strange Object Research (CASTOR) and Zero Degree Calorimeter (ZDC) detectors provide additional calorimetry beyond  $|\eta| = 5.2$ . However, they are mainly used in the heavy ion and diffractive physics programs of CMS, and play no role in the detection of heavy SUSY particles. Therefore, they will not be discussed here.

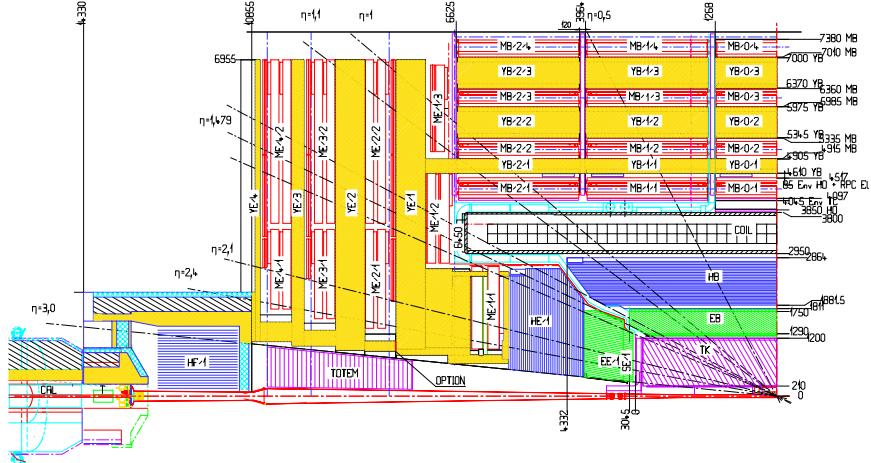


Figure 1.3: Longitudinal quarter cross-sectional view of CMS. The nominal interaction point is at the lower right-hand corner of the drawing. The tracker is shown in purple diagonal hashing, the electromagnetic calorimeter in green, the hadronic calorimeter in blue, and the muon stations in red. The solenoid is shown in black and white and labeled COIL, and the magnet return yoke is shown in yellow. Radial and longitudinal distances are measured in millimeters. Reprinted from Fig. CP 1 of ref. [?].

systems. Section 1.1 describes the subdetector technologies and performance benchmarks, while Section 1.2 details the CMS trigger and data acquisition systems and framework for promptly reconstructing and transferring data worldwide. For a thorough description of CMS, see ref. [?]. Unless otherwise noted, all information in this chapter comes from ref. [?].

## 1.1 The Detectors and Their Operating Principles

### 1.1.1 Tracking System

Given the LHC design instantaneous luminosity, efficient reconstruction of charged particle tracks from transverse momenta of 1 GeV up to 1 TeV can only be achieved with a low occupancy tracker. For  $r < 10$  cm, the hit rate density is highest, leading to the choice of  $100 \mu\text{m} \times 150 \mu\text{m}$  silicon pixel sensors for hit detection. For  $20 \text{ cm} < r < 110 \text{ cm}$ , the lower hit rate allows the use of silicon strips, with length along  $z$  of

order centimeters and length along the  $r \cdot \phi$  curve of order hundreds of microns. This design leads to a pixel hit occupancy of  $\sim 10^{-4}$ /pixel/BX and a strip hit occupancy of  $\sim 10^{-2}$ /pixel/BX, where BX refers to 1 LHC bunch crossing.

As radiation dose from hadrons accumulates over the lifetime of the tracker, silicon leakage current through the semiconductor junctions increases, heating up the sensors. Since the leakage current itself depends on temperature, this can lead to *thermal runaway* that damages the detector. To avoid this, the tracker must be cooled to approximately  $-10^{\circ}\text{C}$ . Operating at this temperature, the signal:noise ratio in the silicon sensors is 10:1, and should remain at that level over the 10-year lifetime of the tracker.

At its thickest ( $|\eta| \sim 1.5$ ), the tracker depth (including services) is  $\sim 1.8X_0$ , and the depth falls off to  $\sim 1X_0$  in thinner areas. Unfortunately, the large mass of the tracker degrades somewhat the performance of the electromagnetic calorimeter behind it, as  $\sim 50\%$  of photons will convert to  $e^+e^-$  pairs in the tracker.

## Pixel Detector

A longitudinal quarter view of the three barrel pixel (BPix) layers and two forward pixel (FPix) disks is shown in Figure 1.4. There are 768 BPix modules in total. Each BPix layer is divided into 32  $\phi$ -wedges, with eight modules per wedge arranged end-to-end in  $z$ . The  $\phi$ -wedges operate nearly independently in terms of clock and readout. Each FPix disk consists of 24  $\phi$ -wedges, with pie-shaped modules attached to the front and back of the disk, for a total of 192 modules. The front- and back-side modules of the FPix disks are constructed of different sized *plaquettes*, or multi-pixel sensor chips, such that the gaps in the front-side module are covered by plaquette area in the back-side module and vice versa. An illustration of the BPix and FPix mechanical layouts is given in Figure 1.5.

Since the electric field in the depletion region of the BPix sensors is perpendicular

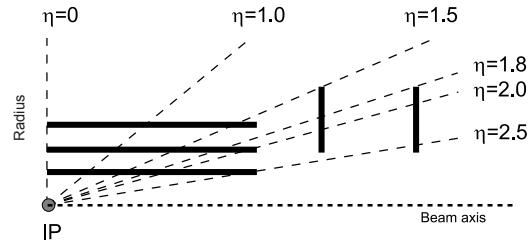


Figure 1.4: Longitudinal quarter view of the pixel detector. Reprinted from Fig. 3.6 of ref. [?].

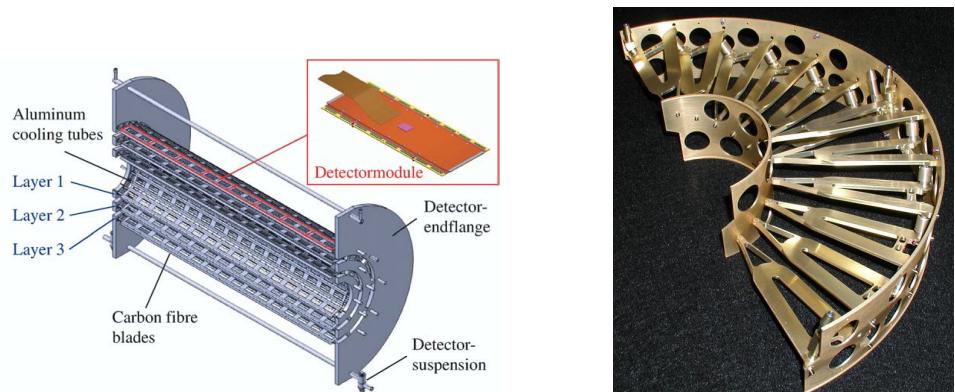


Figure 1.5: BPix and FPix mechanical structures.

(i.e. pointing along  $r$ ) to the magnetic field of CMS (i.e. pointing along  $z$ ), the charge carriers in the silicon experience a Lorentz drift along  $\phi$ . The multi-pixel sensor pitch is such that this causes the charge from one particle hit to be shared among multiple pixels. Particle hits are reconstructed reading out the analog pixel signal and interpolating between signals in multiple pixels. This method achieves a  $15\text{-}20 \mu\text{m}$  spatial resolution, which is comparable to the sensor pitch. To induce this effect in FPix, the sensor wedges are tilted by the approximate BPix Lorentz angle of  $20^\circ$  [?] with respect to the  $y$ -axis.

Each multi-pixel sensor consists of an array of  $52 \times 80$  n-type pixels implanted onto an n-type substrate with  $320 \mu\text{m}$  thickness. The other face of the substrate is covered with a thin layer of p-type semiconductor. Except for the outer edges, which are held at ground potential to prevent sparking between the sensor edges and the connected readout chip [?], the p-side is reverse biased at 150 V (BPix) or 300 V (FPix). The pixels are held at ground potential. A particle entering through the p-side will cause a burst of current to flow across the p-n junction. The charge will be collected by the pixels, which are bump-bonded to the readout. The BPix and FPix sensors employ slightly different technologies for electrically isolating the individual pixels, but both rely on the idea of surrounding the pixels with a p-type material to provide a p-n junction that acts as a barrier to current flow.

Each  $52 \times 80$  pixel sensor is bump bonded to a readout chip (ROC). The ROCs provide zero suppression and amplify, buffer, and communicate the signals from the sensors. A single token bit manager (TBM) controls  $\sim 16$  ROCs in the barrel or  $\sim 24$  ROCs in the endcaps. Its purpose is to distribute the clock and trigger to the ROCs (the latter initiates a transmission of the signal further upstream to be assembled into the full event readout of CMS). The clock and trigger are supplied by the pixel front end controller (pFEC), which interfaces to the central clock and data acquisition systems. Analog signals that are collected from the pixel front ends are digitized by

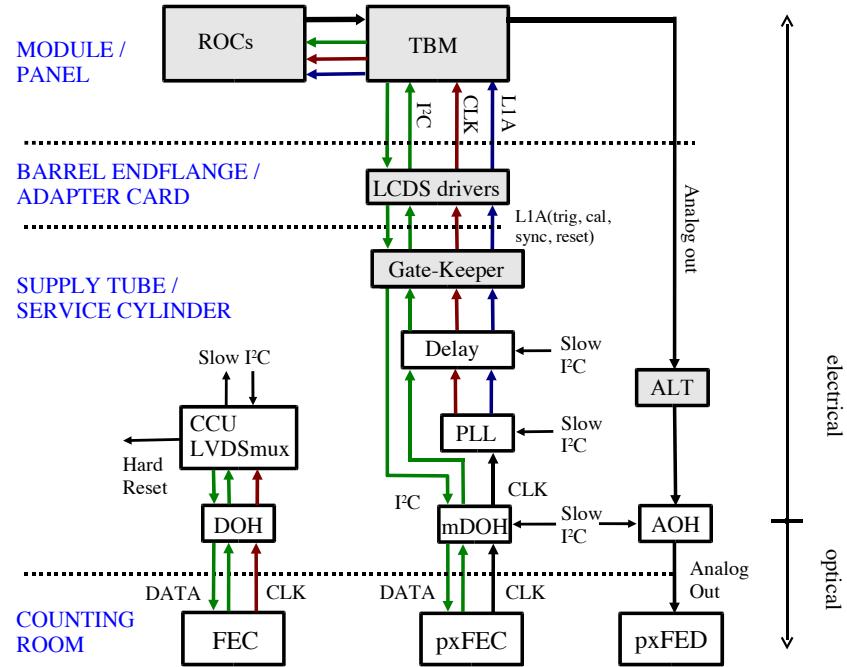


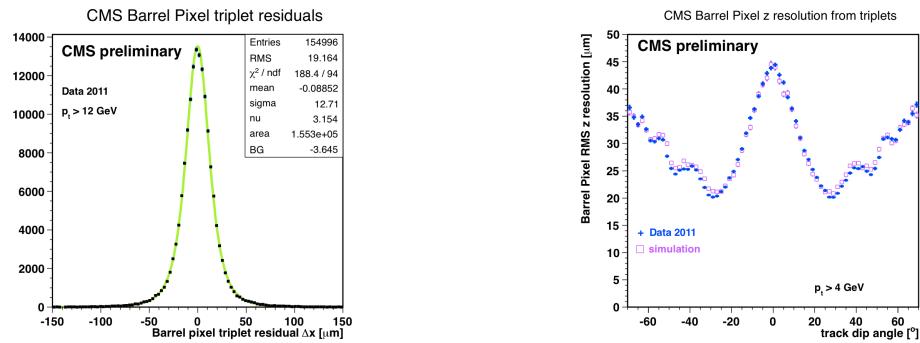
Figure 1.6: Pixel control and readout system. Reprinted from Fig. 3.9 of ref. [?].

the pixel front end digitizer (pxFED). A diagram of the readout system is shown in Figure 1.6.

Figure 1.7 shows some results highlighting the performance of the pixel detector.

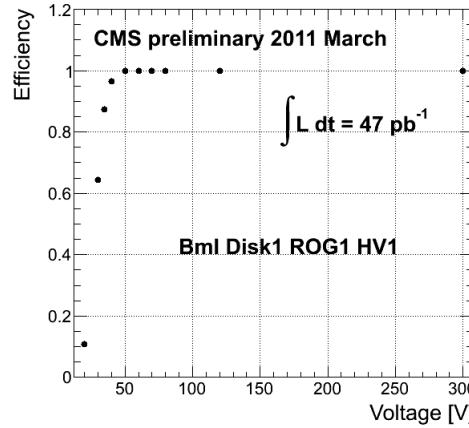
### Silicon Strip Tracker

The silicon strip tracker is divided into four parts: the inner barrel (TIB) and inner disks (TID), covering the radial extent  $20 \text{ cm} < r < 55 \text{ cm}$  and  $z$  extent  $80 \text{ cm} < |z| < 90 \text{ cm}$ ; and the outer barrel (TOB) and endcap (TEC), covering the radial extent  $61 \text{ cm} < r < 108 \text{ cm}$  and  $z$  extent  $124 \text{ cm} < |z| < 282 \text{ cm}$ . A number of the tracker layers and endcaps hold double-sided strip modules (shown as double lines in Figure 1.8), with the rear module tilted at an angle of 100 mrad with respect to the front module, to provide a measurement in two coordinates. There are a total of 15,148 modules in the tracker, arranged as shown in the longitudinal cross-sectional view of Fig. 1.8. For the TIB and TOB, the modules are arranged in straight rows



(a) BPix hit resolution in the  $r \cdot \phi$  coordinate [?].

(b) BPix hit resolution in the  $z$  coordinate vs. track dip angle, showing the effect of charge sharing on resolution [?].



(c) Pixel reconstruction efficiency vs. bias voltage for a group of three wedges in FPix [?].

Figure 1.7: Pixel detector performance highlights.

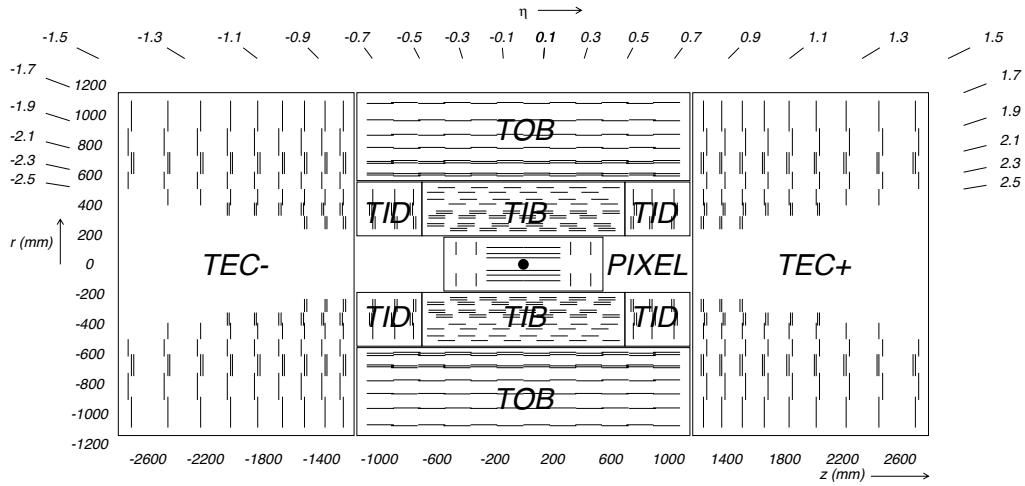


Figure 1.8: Longitudinal cross section of the silicon strip detector. Reprinted from Fig. 3.1 of ref. [?].

end-to-end along  $z$ , with repeating rows covering the full  $2\pi$  extent in  $\phi$ . In each of the TID disks, the modules are arranged into three concentric circular rings of increasing  $r$ . In the TEC, the modules are affixed to  $\phi$ -wedges called *petals*. One side of the TEC and its petal structure is shown in Figure 1.9.

Like the pixels, the strip sensors generate a signal when current flows across a p-n junction in response to interaction with a charged particle. Whereas the pixels are n-type implants on an n-type substrate, with a solid p-type rear layer to which the high voltage is connected, the strips are p-type implants on an n-type substrate, with a solid n-type rear layer connecting to the high voltage. The p-n junction in the strip sensors is at the strip-substrate boundary, whereas in the pixel sensors it is at the boundary between the rear layer and the substrate. Each sensor has either 512 or 768 electrically isolated strips, with pitch varying from 80-205  $\mu\text{m}$  depending on location. Strip lengths in  $z$  range from  $\sim 10$  to  $\sim 25$  cm. Thin (320  $\mu\text{m}$ ) sensors are used in the TIB, TID, and inner four rings of the TEC, while thick (500  $\mu\text{m}$ ) sensors are used in the TOB and the outer rings of the TEC. The thicker sensors compensate for the increased strip capacitance (and hence electronics noise) of the longer strips in the outer layers/disk of the tracker such that strip signal:noise is maintained above

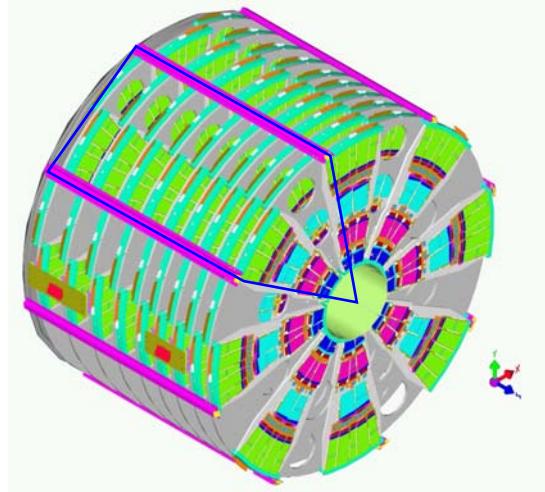


Figure 1.9: View of one tracker endcap, with the outline of a petal shown in blue. There are nine petals per wedge-shaped sector (one per TEC disk). Reprinted from Fig. 3.30 of ref. [?].

10 everywhere.

The strips are wire bonded to a front end readout chip called the APV25. The APV25 amplifies and shapes the strip signals before sending the full analog pulse information to an APVMUX, which multiplexes the output of two APV25s. Then, the electrical signal from the APVMUX is sent differentially a few centimeters to an optical driver, where it is converted to an optical signal and sent to one of the 450 front end drivers (FEDs). The FEDs convert the signal back to an electrical pulse and digitize it for use in the global event assembly. As for the pixels, analog readout is used on detector so that hit reconstruction may benefit from charge sharing.

Clock, trigger, and control signals are sent from the front end controllers (FECs) to phase locked loop (PLL) chips on the front ends. The FECs interface to the global clock and trigger system. Four or six APV25s, an APVMUX, and a PLL chip all sit on a *hybrid*, two thin or two thick sensors are also affixed. The sensor-hybrid combination and its frame form a module. Figure 1.10 shows a diagram of a module, while Figure 1.11 shows a block diagram of the strip readout architecture.

As an example of the strip capabilities, strip hit resolution and signal:noise mea-

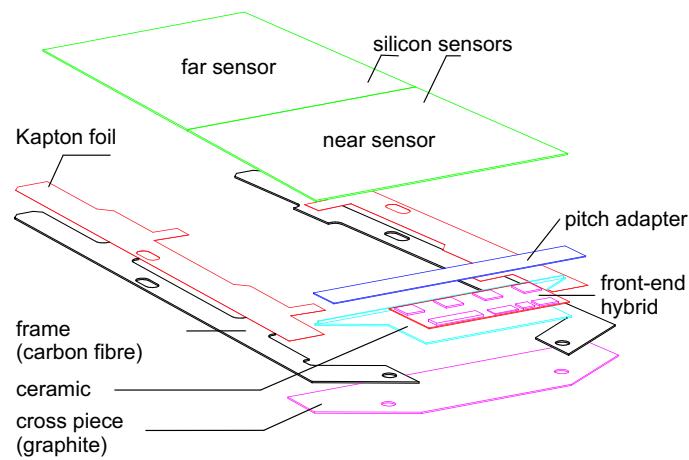


Figure 1.10: Exploded view of a strip module with two sensors. Reprinted from Fig. 3.22 of ref. [?].

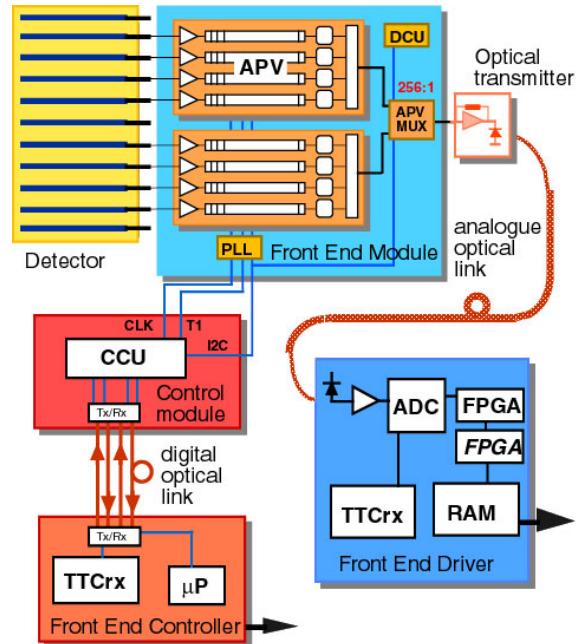


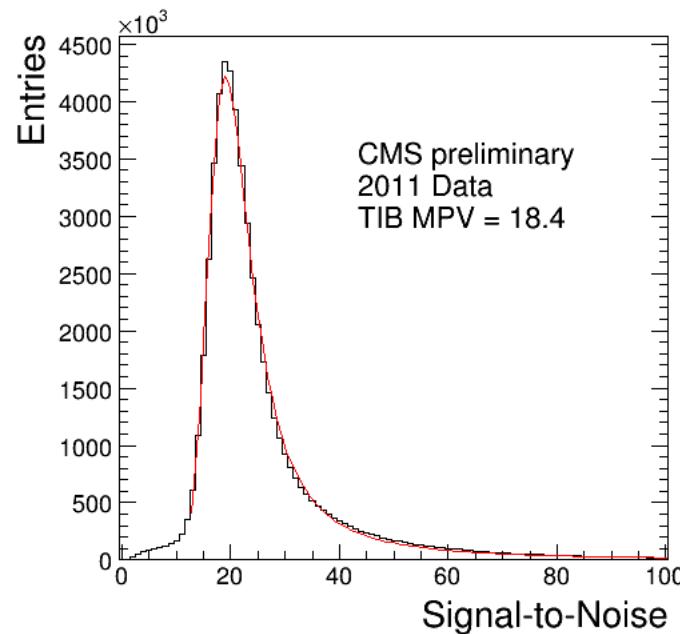
Figure 1.11: Block diagram of the strip readout architecture. Reprinted from Fig. 3.20 of ref. [?].

surements are shown in Figure 1.12. The entire pixel + strip tracker has been used successfully in the reconstruction of primary and secondary vertices, electrons, muons, tau decays, and charm and bottom hadron decays. In addition, the superior performance of the tracker over the hadronic calorimeter for low energy charged hadrons has been exploited in the the particle flow jet and  $\cancel{E}_T$  reconstruction technique (see Sec. ??). The CMS silicon strips, as well as the pixels, are well aligned and operating at close to design performance.

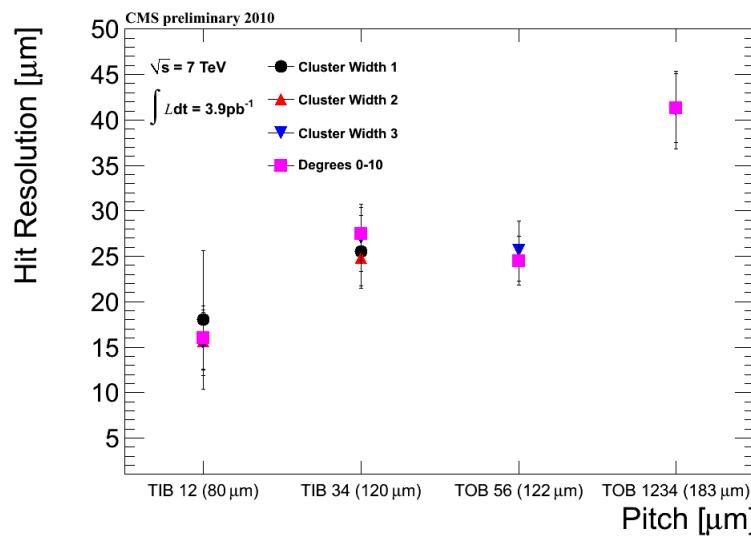
### 1.1.2 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is composed of 68,524 lead tungstate ( $\text{PbWO}_4$ ) crystals, divided into one barrel (EB) layer and two endcap (EE) disks. In EB, there are 1700 crystals per *supermodule* (SM), arranged in a  $20 \times 85$  grid in  $\phi \times \eta$ . Two SMs are laid out end-to-end to form one row at fixed  $\phi$ , with a total of 18 rows needed to cover the entire  $2\pi$  extent in  $\phi$ . The SMs may be operated independently. In EE, the independent unit is a wedge-shaped sector, with nine sectors covering each endcap side. The 7,324 EE crystals are divided approximately evenly between the 18 EE sectors. A two-layer preshower detector is placed in front of the EE disks, each layer consisting of a lead absorber followed by 1.9 mm pitch silicon strip detectors (the strips in the first layer are rotated 90° with respect to the second layer). The ECAL layout is shown in Figure 1.13.

The electromagnetic energy resolution can be parametrized as  $(\sigma/E)^2 = (S/\sqrt{E})^2 + (N/E)^2 + (C)^2$ , where  $S$  characterizes the size of photostatistical fluctuations,  $N$  characterizes the contribution from electronics noise, and  $C$  is a constant accounting for imperfect intercalibration between crystals, non-uniformity of crystal performance, and incomplete shower containment within one crystal. The design goal of the ECAL is to achieve  $C = 0.5\%$ . Therefore, fast, dense, and relatively radiation hard  $\text{PbWO}_4$  was chosen as the crystal material. When a photon or electron strikes the crystal, it



(a) TIB signal:noise [?].



(b) TIB and TOB hit resolution as a function of strip pitch [?].

Figure 1.12: Strip detector performance highlights.

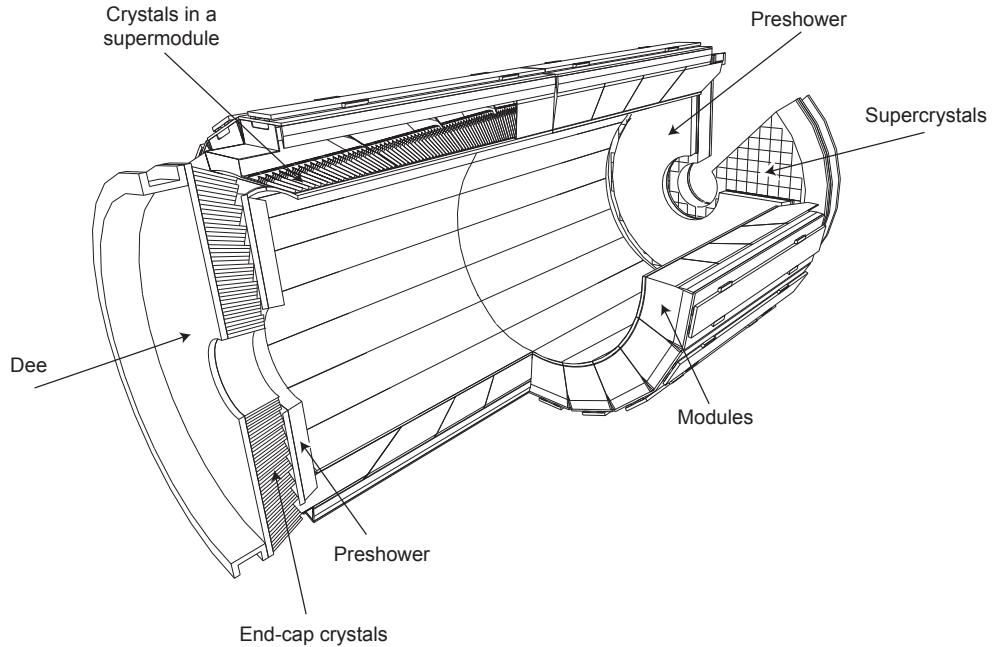


Figure 1.13: Layout of the ECAL detector. Reprinted from Fig. 4.5 of ref. [?].

initiates an electromagnetic (EM) shower. Due to the density, short radiation length, and small Molière radius of  $\text{PbWO}_4$ , nearly the entirety of an EM shower can be contained in a single 23-cm long crystal with front face dimensions  $2.2 \text{ cm} \times 2.2 \text{ cm}$ . The crystals scintillate in the blue-green part of the spectrum, emitting  $\sim 80\%$  of the scintillation light within 25 ns. Light is transmitted along the length of the crystals and collected at the rear with avalanche photodiodes (semiconductor diodes) in EB or vacuum phototriodes (conventional photomultipliers) in EE. Since the light output is low and varies with temperature, the crystals must be kept precisely at  $18^\circ\text{C}$ . The EB and EE crystals, which are slightly tapered to match the lateral shower development, are shown in Figure 1.14.

For each trigger, 10 samples, each separated by 25 ns, are read out. The 10-sample pulse is amplified and shaped by a multi-gain preamplifier (MGPA) residing on a very front end (VFE) card serving five crystals. The MGPA can switch between gains 1, 6, and 12 to avoid saturation of the electronics, and affords a dynamic range up to

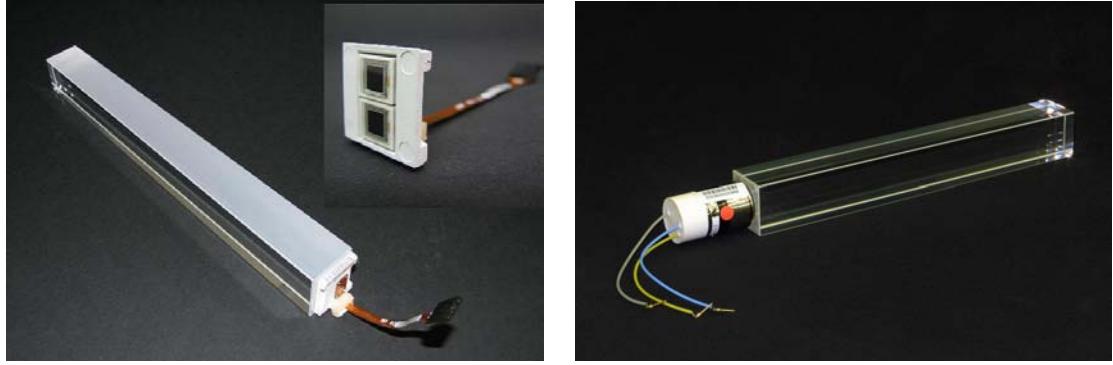


Figure 1.14: Left: EB crystal with attached APD. Right: EE crystal with attached VPT. Reprinted from Fig. 4.2 of ref. [?].

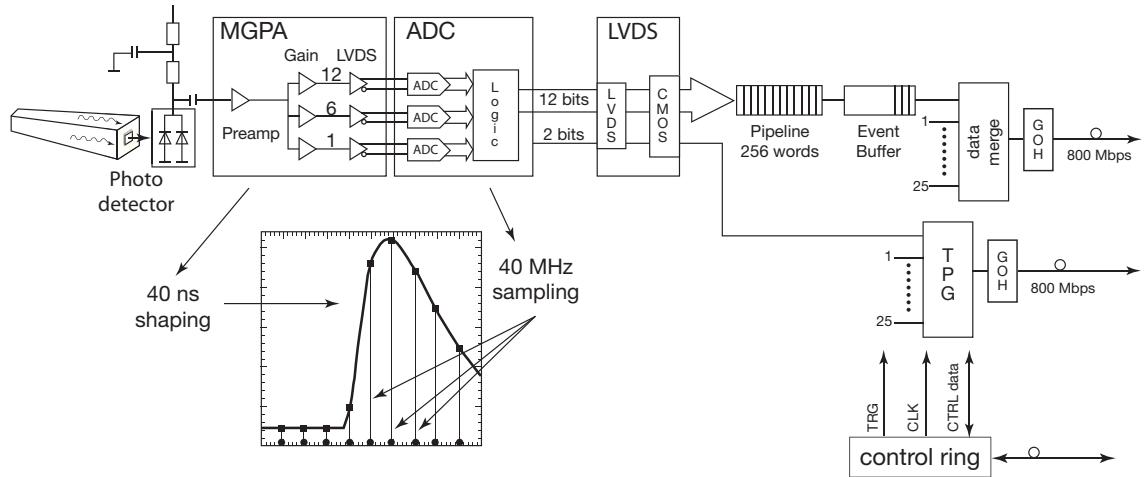


Figure 1.15: Flow chart of the crystal readout, showing the 10-sample pulse shape. Reprinted from Fig. 4.9 of ref. [?].

3 TeV. The samples are digitized on the VFE card, then sent to the front end (FE) card serving five VFEs. Digitized samples are buffered in the FE card until receipt of a trigger, when they are sent over an optical link to the data concentrator card (DCC) that interfaces to the global DAQ. The DCC interfaces to the *selective readout* processor, which decides whether a crystal should be read out with or without zero suppression based on its proximity to a high-energy hit. The clock is transmitted to the FE cards from the Clock and Control System (CCS) boards. A flow chart of the crystal readout is given in Figure 1.15.

At each bunch crossing, the trigger concentrator cards (TCC) of the ECAL compute *trigger primitives* from  $5 \times 5$  non-overlapping transverse energy sums (in the endcaps the geometry is not always  $5 \times 5$ ). This information, along with a special bit in EB only characterizing the transverse shower profile that is used for rejection of anomalous APD hits (see Sec. ??), is transmitted from the TCCs to the synchronization and link boards (SLBs), and then on to the global trigger system. The trigger decision is communicated to the DCCs, which request the buffered data from the front ends if the decision is affirmative.

Despite the radiation hardness of lead tungstate relative to other types of crystals, it still suffers from transparency loss due to radiation-induced lattice damage, as shown in Figure 1.16. In addition, any unforeseen change in the gains of the MGPAs and VPTs, or in the pedestal levels, will degrade the energy resolution. For this reason, a continuously running calibration system is installed with the ECAL. The system makes use of the LHC abort gaps to read out the pedestal levels, test pulses fired into the MGPAs, and laser (EB and EE) or LED (EE only) pulses fired into the crystals at regular intervals. Laser and LED events are used to compute corrections to the crystal gains for transparency loss, while the other types of calibration events serve to monitor changes in the electronics performance due to magnetic field or high voltage cycling. The mean time between transparency measurements is  $\sim 40$  minutes.

Figure 1.17 shows the architecture of the laser monitoring system.

The current ECAL energy resolution is somewhat worse than the design goal of 0.5%. An incomplete understanding of (a) the transparency loss and (b) the photon conversion and electron bremsstrahlung processes in the  $\sim 1X_0$  of tracker material in front of the ECAL are the main limiting factors in improving the resolution. However, as more data accumulate, more refined models of transparency loss and EM interactions in the tracker can be built, leading to better resolution. Energy resolution vs.  $|\eta|$  can be seen in Figure 1.18.

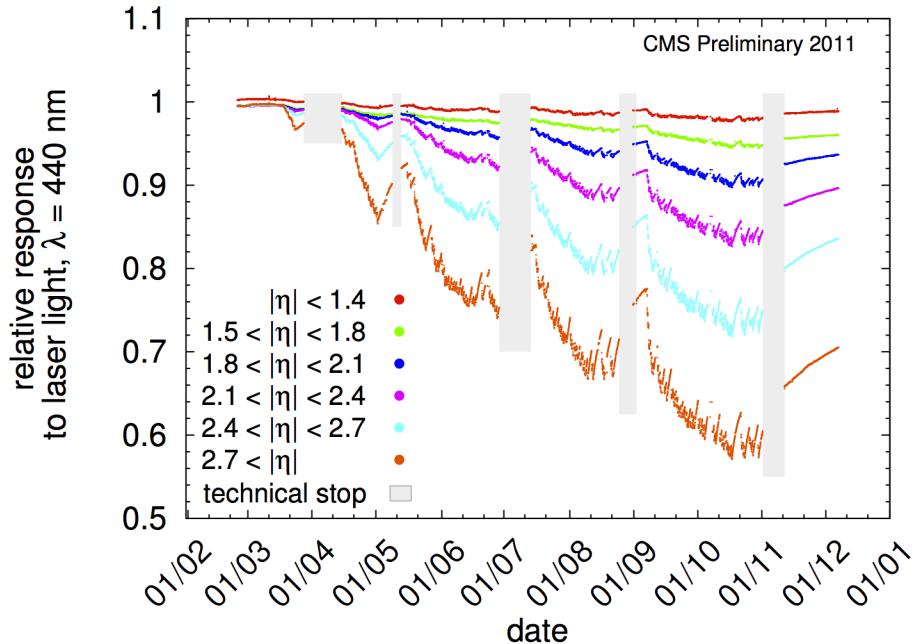


Figure 1.16: Relative response of the crystals to blue laser pulses from February 1, 2011 to January 1, 2012 [?]. Technical stops, during which the LHC is turned off for maintenance and development, are shown in gray. These periods of inactivity correspond to growth in the crystal response, as radiation damage recovery occurs.

The 10-sample readout coupled with the fast scintillation time of lead tungstate allows for a very precise reconstruction of the time of ECAL hits. ECAL timing is used for searches for long-lived particles that decay to photons or jets, such as long-lived neutralinos in GMSB [?]. Figure 1.19 shows the timing resolution in EE.

### 1.1.3 Hadronic Calorimeter

The CMS hadronic calorimeter (HCAL) has four parts: HCAL barrel (HB), HCAL endcap (HE), and HCAL outer (HO), which all utilize the same brass absorber / plastic scintillator sandwich technology; and HCAL forward (HF), which is a Čerenkov detector made of quartz fibers. A quarter longitudinal cross-sectional view of HCAL is shown in Figure 1.20. Like EB, HB is formed of 36  $\phi$ -wedges (18 cover  $2\pi$  in positive  $\eta$ , 18 cover  $2\pi$  in negative  $\eta$ ). Each wedge is divided into 16 along  $\eta$  and four along  $\phi$ , for a total of 64 readout towers per wedge (compare 1700 individually read out

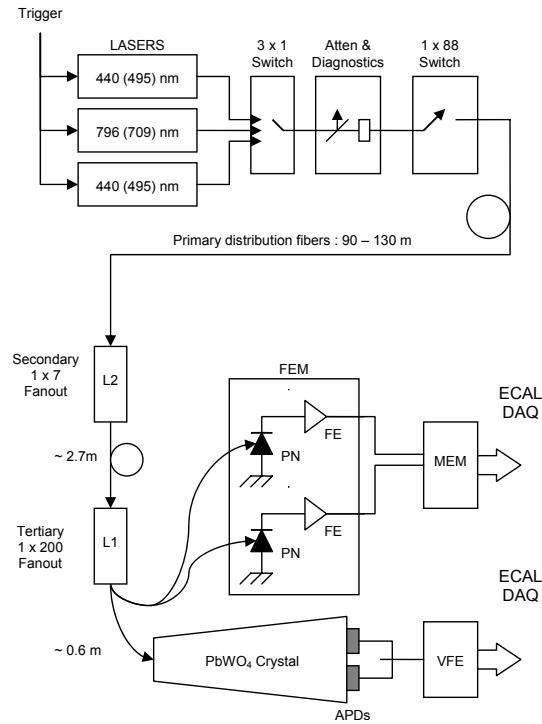


Figure 1.17: Architecture of the laser monitoring system. Reprinted from Fig. 4.16 of ref. [?].

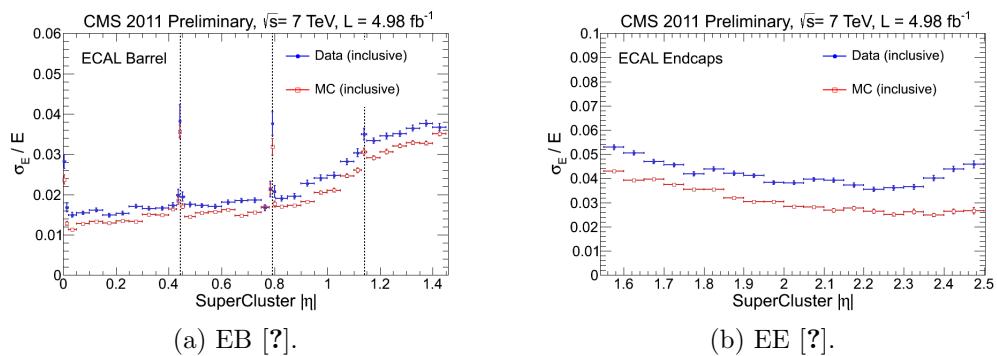


Figure 1.18: Energy resolution vs.  $|\eta|$  for  $Z$  decay electrons for data (filled blue circles) and MC (empty red squares). The dotted lines show the locations of module gaps (three per SM).

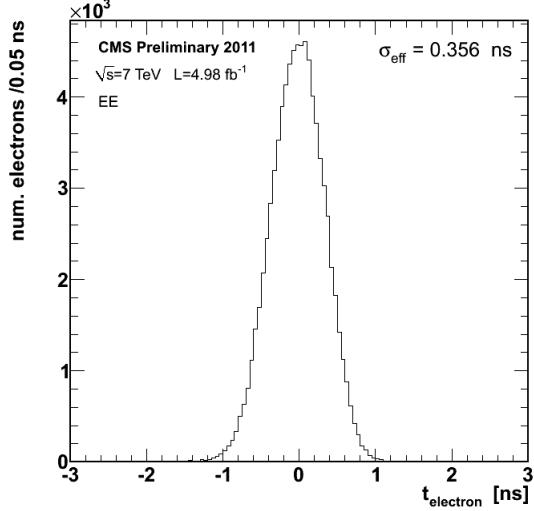


Figure 1.19: Distribution of reconstructed times of  $Z$  decay electrons in EE [?].

crystals per EB wedge). HE is divided into 36  $\phi$ -wedges containing 38 readout towers each. HO consists of five rings around HB and HE distributed symmetrically along  $z$ . There are 72  $\phi$ -slices per ring, with each  $\phi$ -slice further divided into 5, 6, or 8 along  $z$  depending on ring. The HF fibers are distributed within the steel absorber. HF is divided into 18  $\phi$ -wedges per endcap side, each containing 24 readout towers. All HB towers have a single readout channel except for the two in each wedge at highest  $|\eta|$ , which are segmented into two longitudinal layers for readout. In HE, all towers have two longitudinal readout layers, except for the three rings of towers closest to the beam line, which have three. There are also two longitudinal depths of HF fibers.

HB, HE, and HO are all sampling calorimeters consisting of alternating layers of brass absorber and plastic scintillator. The absorber initiates the hadronic shower, and as shower particles travel through the scintillator the scintillation light is read out by wavelength-shifting (WLS) fibers connected to the scintillator tiles.<sup>2</sup> The full development of the shower is sampled by the layers of instrumented scintillator. The scintillator tiles are staggered so that there are no cracks in coverage along the direc-

---

<sup>2</sup>By contrast, in the ECAL, the crystal material acts as both absorber and scintillator, greatly reducing the contribution to energy resolution from sampling fluctuations.

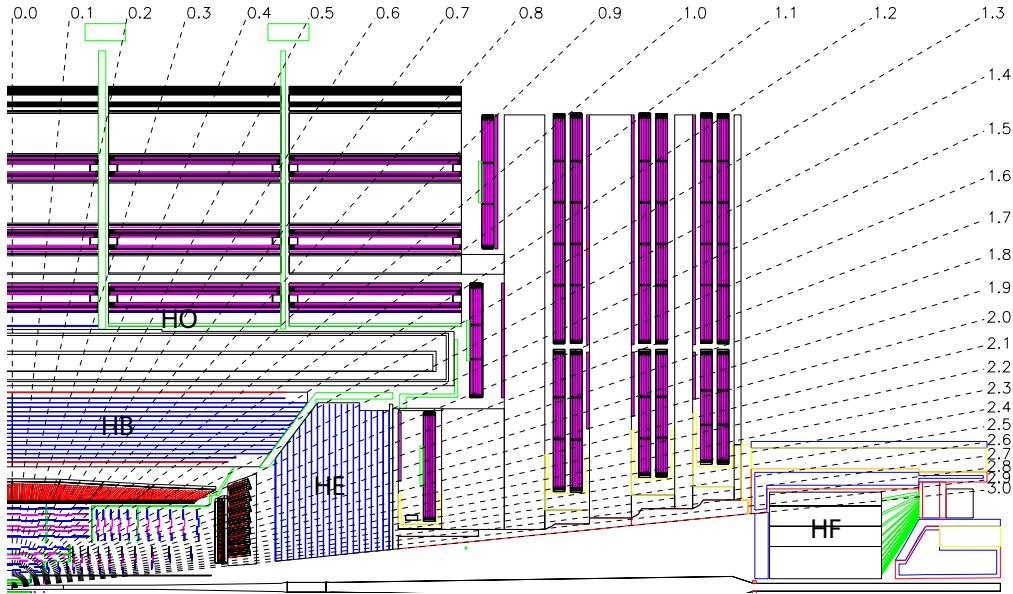


Figure 1.20: Quarter longitudinal cross-sectional view of HCAL (and muon stations in purple). Reprinted from Fig. 5.1 of ref. [?].

tion projected back to the beam spot. Light output from all tiles in a single readout tower is collected via the WLS fibers and merged into a single signal that is amplified by a hybrid photodiode (HPD). A diagram of the optical readout of HB (similar for HE and HO) is shown in Figure 1.21.

Due to the extremely harsh radiation environment near the beam line, HF is constructed of a 1.2-m thick, 1.7-m long ring of steel absorber with radiation hard quartz fibers distributed within the steel and running parallel to the beam line. Hadronic showers develop in the steel and are sampled in the quartz fibers when charged shower particles hit the the fibers and emit Čerenkov light. The light is transmitted by total internal reflection down the fibers to a photomultiplier tube (PMT), where the signals from all fibers in an HF tower are merged into one. Since only relativistic particles emit Čerenkov light in these fibers, it is mostly the EM component of the hadronic shower, consisting of neutral pions decaying to photons that interact electromagnetically with the absorber, that is sampled [?]. The charged hadrons produced in hadronic showers are typically too slow to generate Čerenkov light. Figure 1.22

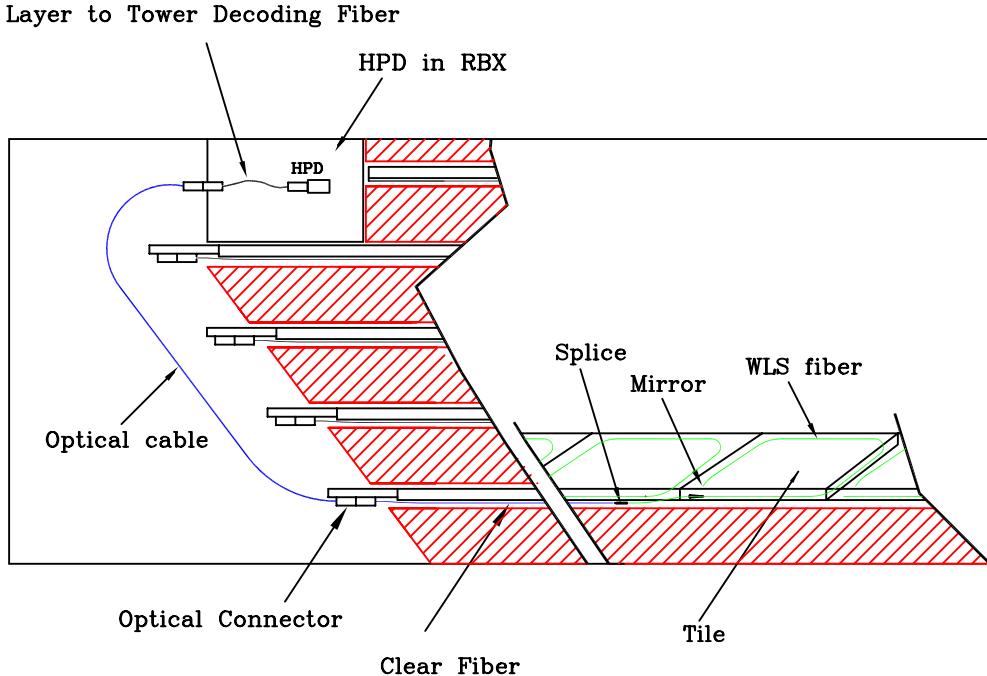


Figure 1.21: Diagram of the optical readout of HB. Reprinted from Fig. 5.7 of ref. [?].

shows a cross-sectional view of one side of HF.

Electrical signals from either HPDs (HB/HE/HO) or PMTs (HF) are digitized on the front ends by means of a fast charge-integrating ADC. The digitized signals are sent off-detector to the HCAL Trigger/Read-Out (HTR) boards, where they await a trigger decision. If the trigger is accepted, the signals are sent on to the HCAL data concentrator cards (DCCs), which interface to the global DAQ system. HCAL trigger primitives, consisting of transverse energy sums over an entire tower, are calculated in the HTR boards and sent to the global trigger system.

Selected HCAL performance results can be seen in Figure 1.23.

#### 1.1.4 Muon System

Beginning at a radius of  $\sim 10$  interaction lengths from the beam line, where all particles except muons should have been stopped by the HCAL, are the muon chambers, interspersed with the iron return yoke of the CMS magnetic field. Three technologies

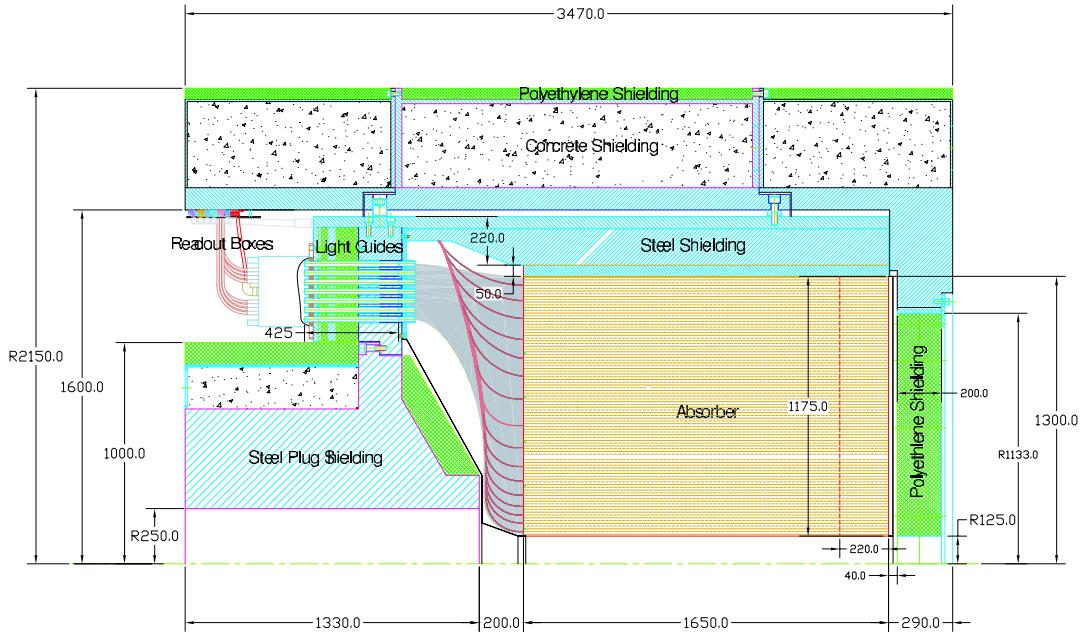
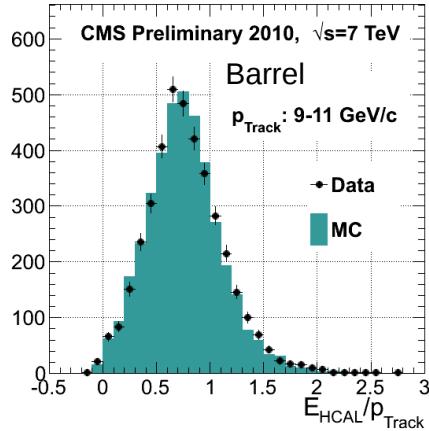


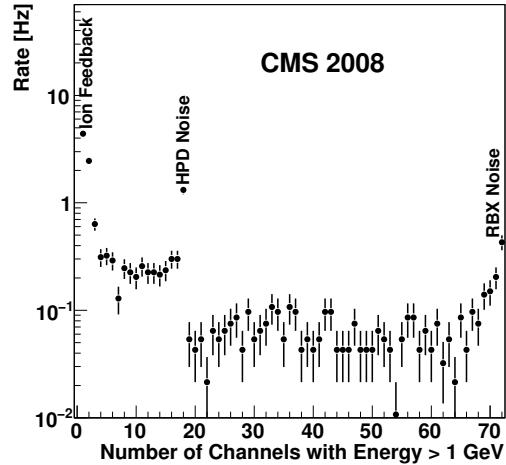
Figure 1.22: Cross-sectional view of one side of HF. The  $z$ -axis is horizontal. Reprinted from Fig. 5.28 of ref. [?].

are employed: drift tubes in the barrel section (MB), cathode strip chambers (CSCs) in the endcap section (ME), and resistive plate chambers (RPCs) in both sections to provide an independent trigger with superior time resolution. There are four barrel layers of stations extending out to  $|\eta| = 1.2$ . Each endcap consists of five disks of stations as shown in Figure 1.24b, covering  $1.4 < |\eta| < 2.4$ . RPCs populate the barrel and endcap muon systems alongside the DT chambers and CSCs. Since they have time resolution much better than a few ns, they are used to assign the bunch crossing of muon tracks and provide a  $p_T$  trigger with sharp turn-on.

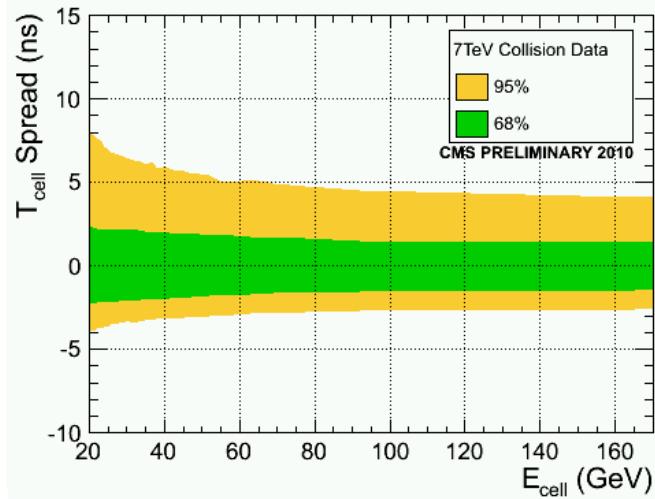
Each DT chamber consists of two  $r \cdot \phi$  superlayers (SLs) and optionally one  $z$  SL (in all chambers except those in the fourth layer). The SLs contain four rows of drift tubes, with the rows staggered such that there are no gaps in the coverage. The  $r \cdot \phi$  SLs have the tube axis parallel to the beam line, while the  $z$  SL is perpendicular to the beam line. The tubes are  $\sim 2.4$  m in length and  $13 \text{ mm} \times 42 \text{ mm}$  in cross section. Each chamber therefore records eight  $r \cdot \phi$  tracking points and optionally four



(a) Data/MC comparison of HB response to charged tracks of 9-11  $\text{GeV}/c$  momentum [?].

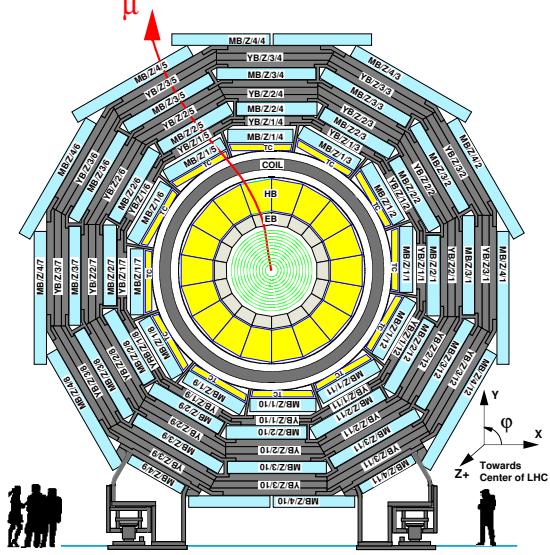


(b) Distribution of tower multiplicity, clearly showing three peaks in rate corresponding to noise sources (see Sec. ??) [?].

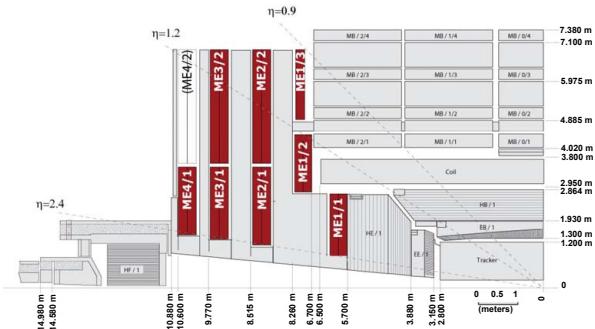


(c) Timing resolution vs. tower energy [?].

Figure 1.23: Selected HCAL performance results.



(a) One of the five wheels of MB, showing the four layers of muon stations. The five wheels are spaced symmetrically in  $z$  about  $z = 0$ . As a muon traverses the muon detectors, its curvature in the transverse plane changes direction and magnitude due to the magnetic field in the return yoke, which is of opposite sign and reduced strength relative to the field within the solenoid volume. Reprinted from Fig. 7.3 of ref. [?].



(b) Quarter longitudinal cross section of CMS highlighting the location of the ME disks. Reprinted from Fig. 7.47 of ref. [?].

Figure 1.24: View of the MB and ME layout in CMS.

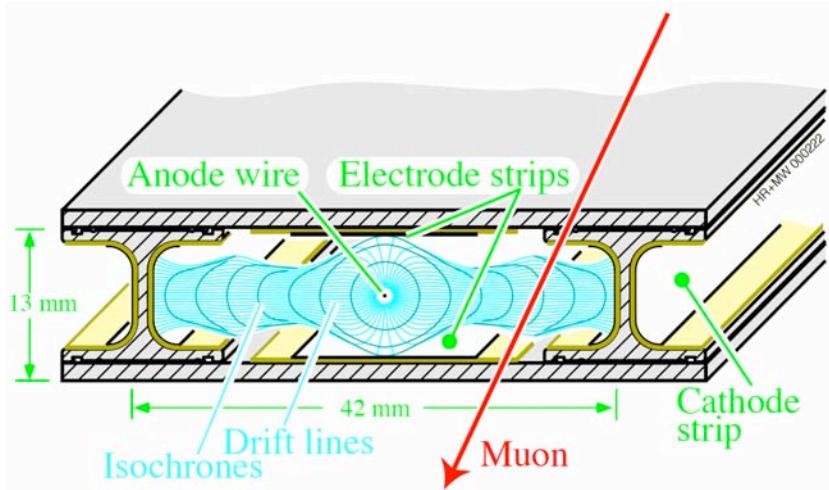


Figure 1.25: Electric field lines within a drift tube as well as the contours of equal drift time. Reprinted from Fig. 7.5 of ref. [?].

$z$  tracking points. The tubes are filled with an 85%Ar + 15% CO<sub>2</sub> gas mixture. An anode wire at 3600 V runs the length of the tube, while the walls are covered with electrodes held at 1800 V or -1200 V depending on wall. When a muon passes through the tube, it ionizes the gas atoms. The liberated electrons drift along the electric field lines created by the electrodes to the anode, which is read out. Figure 1.25 shows the electric field lines within a drift tube as well as the contours of equal drift time. The maximum drift time is 380 ns.

CSCs consist of alternating layers of cathode strips (four planes oriented along  $r$ ) and anode wires (three planes oriented along  $\phi$ ). A 40%Ar + 50%CO<sub>2</sub> + 10%CF<sub>4</sub> gas mixture fills the space between two successive planes, forming six gas gaps. When a muon ionizes the gas atoms, the positive ions drift toward the anode and are read out to provide a measurement of  $r$ , just as in the DTs. However, an image charge is induced on the cathode strips, which is also read out to provide a measurement of  $\phi$ . The wires are spaced 3.2 mm apart. The cathode strips have pitch varying from 8.4 mm at the end closest to the beam line to 16 mm at the other end, and are spaced 0.5 mm apart. A trapezoidal CSC is shown in Figure 1.26.

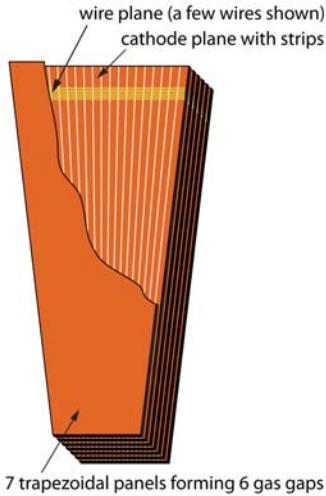


Figure 1.26: CSC wedge, showing the cathode and wire planes. Reprinted from Fig. 7.49 of ref. [?].

Track stubs from the muon system are combined with tracks from the tracking system to form more precise muon tracks than either system could form alone, as shown in Figure 1.27. This leads to extremely good di-muon invariant mass resolution (Figure 1.28) over a large  $p_T$  range.

## 1.2 Triggering, Data Acquisition, and Data Transfer

### 1.2.1 Level 1 and High Level Trigger Systems

The Level-1 (L1) trigger system, which encompasses dedicated hardware processors to construct trigger objects (typically high  $p_T$  jets, electrons, photons, taus, and muons) out of the calorimeter and muon hits, distributes a L1 accept or reject to all subdetectors at the LHC bunch crossing frequency of 40 MHz. Further data filtering is performed by the High Level Trigger (HLT) system, a farm of  $\sim 1000$  commercially available processors running a slimmed down version of the CMS event reconstruction

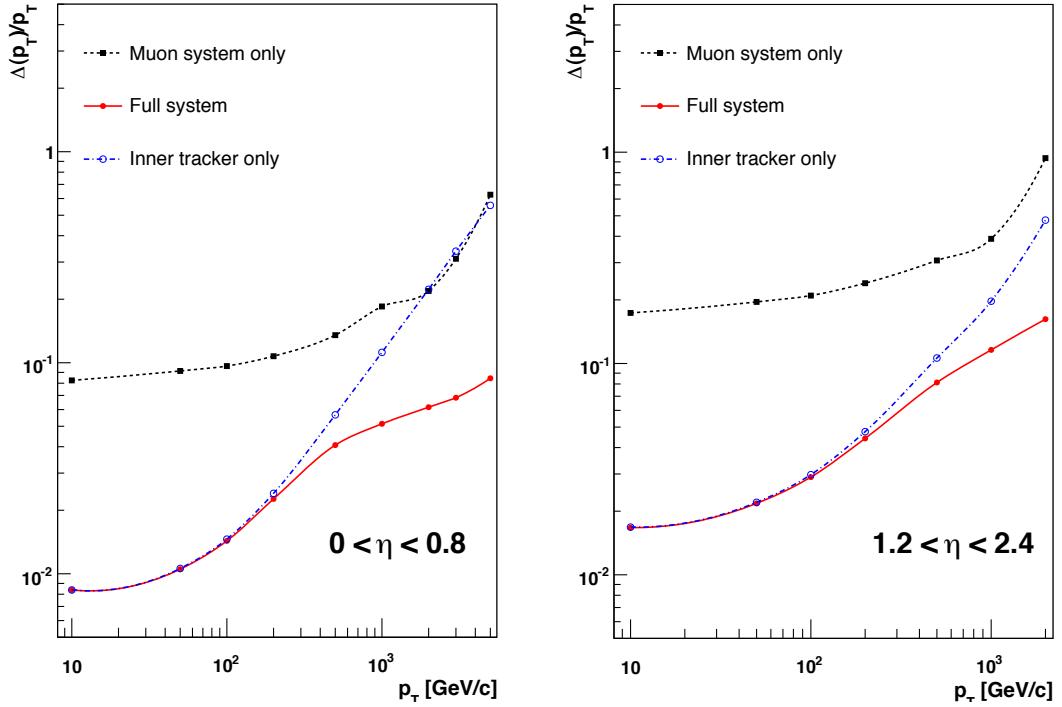


Figure 1.27: Muon  $p_T$  resolution as a function of muon  $p_T$  for tracker information only (blue), muon information only (black), and both tracker and muon information combined (red). Reprinted from Fig. 1.2 of ref. [?].

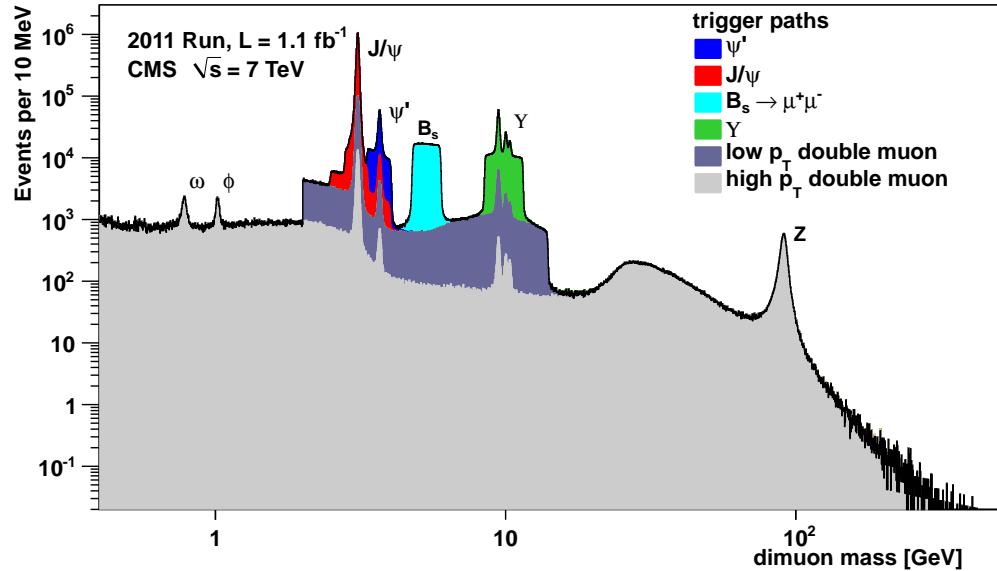


Figure 1.28: Di-muon invariant mass spectrum broken down by trigger path [?]. The light(dark) gray regions show the contribution from high- $p_T$ (low- $p_T$ ) di-muon triggers. Note that no  $B_s \rightarrow \mu^+\mu^-$  decays have been observed [?]; the light blue region just shows the amount of triggers dedicated to the  $B_s \rightarrow \mu^+\mu^-$  search.

software CMSSW. The data rate received by the HLT is  $\sim 100$  kHz; the output rate of events permanently written to disk is  $\sim 100$  Hz. An L1 trigger *latency* (time between the collision and the distribution of the L1 decision to the subdetectors) of  $3.2\ \mu\text{s}$  is achieved via the use of fast electronics and sufficiently deep buffers to pipeline trigger primitives waiting to be analyzed. This latency corresponds to the length of the LHC abort gap, so in principle CMS may be operated with zero *dead time* (during which LHC bunches are missed because the L1 system is blocked while processing other triggers).

At the bottom, the L1 trigger consists of trigger primitive generators (TPGs) in the calorimeter and muon systems that send  $E_T$  sums or muon track stubs to the regional calorimeter trigger (RCT) or muon track finders, respectively. The RCT, DT track finder (DTTF), and CSC track finder (CSCTF) sort and rank the regional trigger primitives based on  $p_T$  and quality. The ranked RCT candidates and muon track stubs are sent to the global calorimeter trigger (GMT) and global muon trigger (GMT), respectively, where high-level objects like isolated and non-isolated muons and EM candidates, jets, taus, and  $\cancel{E}_T$  are constructed from all the regional inputs and ranked. Calorimeter isolation sums for muons are also sent from the RCT to the GMT. The highest ranked global objects are sent to the global trigger (GT), which sits at the top of the L1 trigger. The GT issues the final L1 accept or reject to all subdetectors based on a comparison of the GMT and GCT candidates with the requirements of its programmed trigger menu. A block diagram of the L1 trigger is shown in Figure 1.29.

A region in the RCT consists of a matrix of  $4 \times 4$  trigger towers. A trigger tower in EB/HB is one HCAL tower + the  $5 \times 5$  matrix of ECAL crystals in front of it; in EE/HE the idea is similar but the counting of crystals and HE towers is slightly more complicated. An EM RCT candidate is built around a high  $E_T$  seed tower. The  $E_T$  of the candidate is the sum of the tower  $E_T$  and the  $E_T$  of its highest- $E_T$

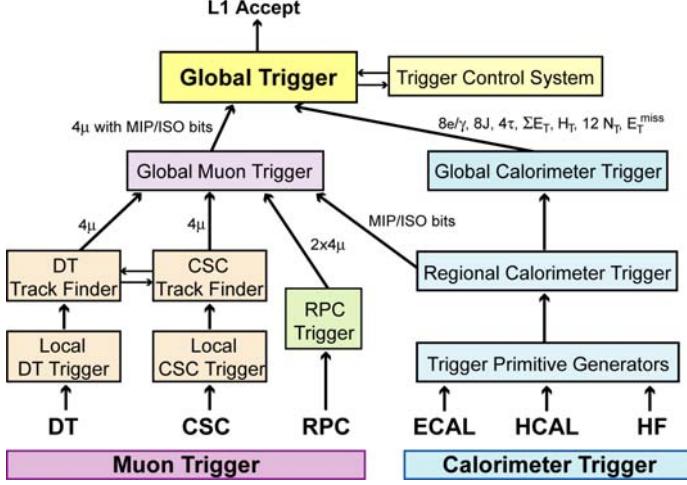


Figure 1.29: Block diagram of the L1 trigger. Reprinted from Fig. 8.1 of ref. [?].

broad side neighbor (see Figure 1.30 for a definition of the broad side neighbors). Two isolation criteria are defined based on (a) the ratio of the EM energy to the HCAL energy in the tower and (b) the shower shape. For a non-isolated EM candidate, the highest- $E_T$  broad side neighboring tower must pass these two isolation criteria; for an isolated EM candidate, all eight neighboring towers must the criteria, and there must also be at least one quiet corner with the  $E_T$  of all five towers in the corner below some threshold (see Fig. 1.30). The process is repeated until four isolated and four non-isolated EM candidates are found, starting with the highest- $E_T$  tower in the region and moving down in tower  $E_T$ . An RCT region is flagged as consistent with tau decay only if the pattern of tower transverse energy sums defines at most a  $2 \times 2$  matrix of energetic towers within the  $4 \times 4$  RCT region.

From the tower transverse energy sums, eight EM candidates, and tau flag received from each RCT, the GCT computes the total  $E_T$  in the calorimeter (and the total  $E_T$  above some programmable threshold, called  $H_T$ ), and the  $\cancel{E}_T$ . It also classifies the towers into jets and determines the globally highest ranked isolated and non-isolated EM candidates. The jet finding uses a clustering algorithm based on the energy of a sub-cluster with respect to its neighbors [?]. Jets are classified as tau decays if all

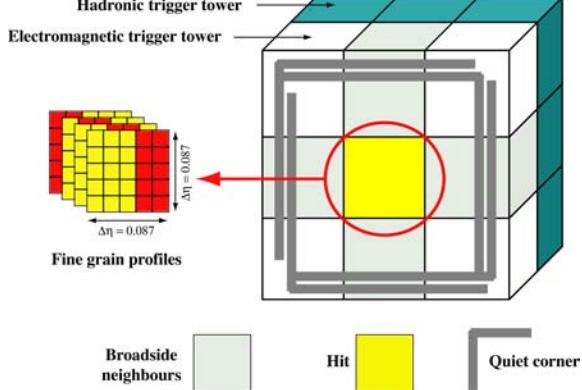


Figure 1.30: Geometry of an EM RCT candidate. Reprinted from Fig. 8.2 of ref. [?].

of the RCT regions participating in the jet clustering had energy patterns consistent with tau decay. Counts of jets above 12 different programmable  $E_T$  thresholds are calculated. The jet counts, energy sums,  $\cancel{E}_T$ , and highest ranked EM candidates are sent to the GT, where the final L1 decision is taken and transmitted to the subdetectors. The GT can execute a maximum of 128 trigger algorithms in parallel. If any one of these algorithms yields an accept, the event is accepted, and all trigger information is sent on to the HLT for further filtering. The double-photon HLT paths used in this analysis (see Sec. ??) require isolated L1 seeds (i.e. EM candidates built by the RCT) with  $E_T > 12$  or 20 GeV, depending on path.

No muon triggers are used in the two-photon analysis. A description of the muon trigger system can be found in ref.[?].

### 1.2.2 Data Acquisition System

The CMS data acquisition (DAQ) system takes event fragments (calorimeter hits, track hits, etc.) from each of the 626 subdetector front end drivers (FEDs), assembles them into a data structure representing the full event, and sends the event on to the HLT for further filtering. The DAQ must operate at an input rate of  $\sim 100$  GB/s, corresponding to an input rate from the L1 trigger of  $\sim 100$  kHz. To facilitate

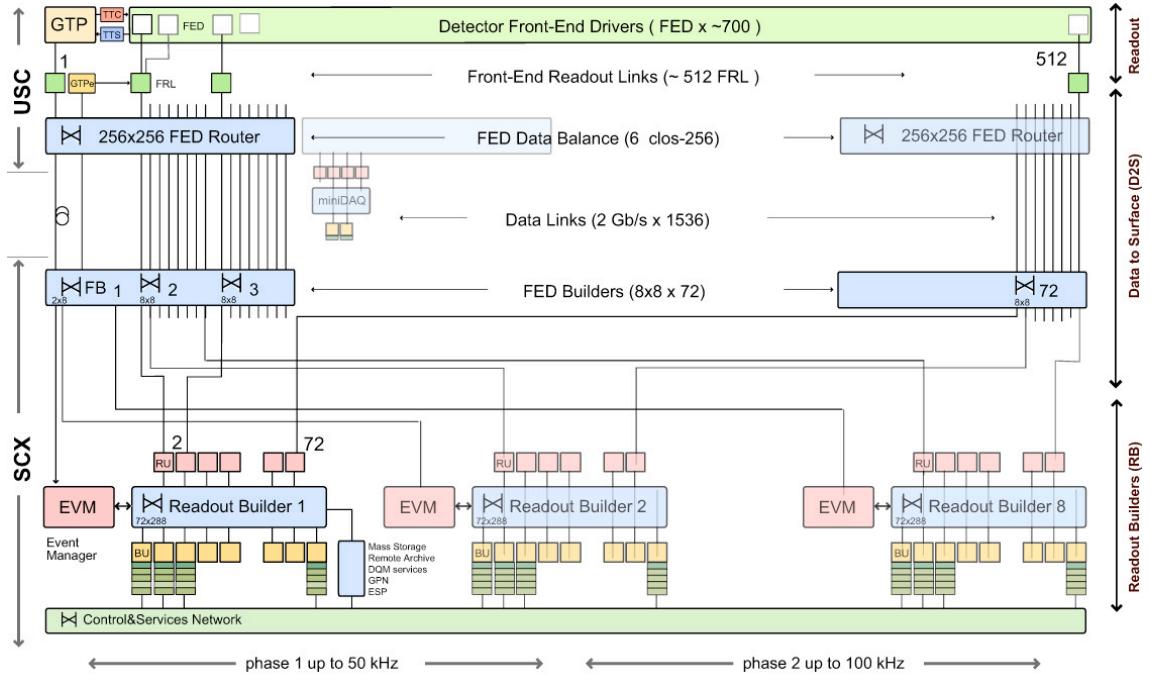


Figure 1.31: Diagram of the DAQ system. The identical event builder systems, shown as inputs and outputs to the boxes labeled “Readout Builder 1”, “Readout Builder 2”, etc., represent the eight slices. Within one slice, data can flow from the detector front ends to the readout systems to the builder network (which assembles the event fragments) to the filter systems (HLT) independently of the other slices. Reprinted from Fig. 9.8 of ref. [?].

expansion of the system as the need arises, the DAQ is composed of eight nearly independent slices. Each slice functions as a smaller version of the whole DAQ that can handle an input event rate up to  $\sim 12.5$  kHz. A diagram of the DAQ system, showing schematically the eight slices, is given in Figure 1.31.

Data from the front ends is collected by the FEDs and pushed to the front end readout links (FRLs), which may take inputs from up to two FEDs simultaneously. The FRLs check for transmission errors, generate event fragments with size  $\sim 2$  kB, buffer the fragments in 64 kB memories, and finally send them to the FED builders. The FEDs, FRLs, and FED builders are located in the underground control room. The 72 FED builders each construct one  $\sim 16$  kB *super-fragment* from the input event fragments, then send the super-fragment on to a readout unit (RU) located in

the surface control room  $\sim 80$  m away. Super-fragments belonging to the same event are sent to RUs in the same DAQ slice. There are 72 RUs per readout builder, one for each super-fragment of an event, with each DAQ slice built around one readout builder (see Fig. 1.31). Each readout builder hosts a number of builder units (BUs) that perform the final integration of super-fragments into complete events.

Resource brokers (RBs) in the HLT filter farm request complete events from the BUs and distribute those events to the filter units (FUs) for HLT selection. If an event passes any one of the HLT paths in the predefined menu, it is sent back to the RB for transfer to the storage manager (SM). The SM nodes transfer accepted events to the CERN Tier-0 prompt reconstruction facility for unpacking of the raw data into ROOT [?] files that can be accessed by physicists wishing analyze the data. The lag time between recording of an event in the DAQ and availability of the fully reconstructed event for analysis is typically 48 hours.

If the buffers of the upstream DAQ elements (the filter farm, readout builders, FED builders, or FRLs) are full, those elements will not request new events from downstream. This can lead to a buildup of events in the downstream element buffers, *back-pressuring* all the way down to the FEDs themselves. The CMS trigger throttling system (TTS) consists of dedicated lines between the FEDs and the GT for the purpose of sending predefined signals to the GT about the state of the FED buffers. If the buffer of a particular FED is getting full, it can alert the GT to reduce the trigger rate so as to prevent FED buffer overflows and loss of time synchronization between event fragments. The TTS latency is  $\sim 1\mu\text{s}$ . Causes of back-pressure (hence dead time) include: problems with the FED electronics (in this case, the upstream elements request events but the FEDs have trouble sending them), increases in the L1 accept rate (perhaps due to a noisy detector channel) beyond what the upstream DAQ elements can handle, increases in the event size due to high pileup or a poor quality beam that scrapes against the beam pipe, failures of the DAQ transmission

lines or DAQ hardware such that events are not requested from the FEDs fast enough, or bottlenecks at the SM nodes or filter farm due to hardware failures or large event sizes.

All components of the DAQ, from the FEDs up to the SMs, are controlled by cross-platform DAQ (XDAQ) [?] processes, or *executives*. The Simple Object Access Protocol (SOAP) [?] protocol is used to transmit control and monitoring data between XDAQ-enabled devices and to the end user, who can view the running of a XDAQ executive via a Web interface called HyperDAQ [?]. The Run Control and Monitoring System (RCMS) handles the configuration and control of all XDAQ executives via a hierarchical structure. At the top of the hierarchy is the Level-0 *function manager* (FM), controlling the Level-1 sub-detector FMs, which in turn control their Level-2 system-specific XDAQ executives. The central DAQ and L1 trigger each have their own Level-1 FM. A unit of data acquisition, called a *run*, may be configured, started, and stopped by an end user interacting with the RCMS Web interface.

### 1.2.3 Data Processing and Transfer to Computing Centers

Data leaving the filter farm are grouped into datasets based on HLT path, i.e. there are different datasets for events passing diphoton triggers, jet triggers, muon + electron triggers, etc. At the Tier-0 facility, the datasets are go through three levels of processing to create three *data tiers*. The first layer produces RAW data by unpacking the detector byte streams sent from the DAQ and L1 trigger into data structures holding the ADC counts recorded for each channel of the detector, digitized trigger primitives, and the L1 decision. A single event has  $\sim 1.5$  MB of RAW data. The next layer of processing is the reconstruction, which forms channel energies in GeV, applies calibrations, and creates high-level objects like photons, electrons, muons, taus, jets,  $E_T$ , and charged tracks. The RECO data tier occupies  $\sim 0.5$  MB per event. Finally, analysis object data (AOD) is a subset of the RECO data, comprising the

high-level objects but usually excluding the individual channel hit information if it is not associated to a physics object. This tier occupies  $\sim 0.1$  MB per event. One copy of the RAW data is stored permanently at CERN and another copy is distributed amongst the Tier-1 facilities (see below) for permanent storage. Changes in the reconstruction algorithms periodically require reprocessing of the RAW data to form a new RECO tier. In general, only the AOD tier is available to physicists wishing to perform analyses due to the smaller size and faster replication and transfer time of AOD with respect to RAW or RECO.

There are three tiers of computing and data storage sites within the Worldwide LHC Computing Grid (WLCG) [?]. The tier closest to CMS is Tier-0, which is located at CERN and performs archiving of the RAW data, prompt reconstruction of the data within  $\sim 48$  hours of its being collected, and transferral of copies of the RECO datasets to Tier-1 facilities. There are a few Tier-1 centers worldwide, hosted by national computing facilities and laboratories. They store parts of the RAW dataset and copies of the RECO datasets, participate in subsequent reconstruction passes after the prompt reconstruction at Tier-0, and ship AOD datasets upon request to the Tier-2 centers. Analysts interact primarily with the Tier-2 centers, which store AOD datasets and run batch processing queues for running analysis jobs over the datasets. Different layers of WLCG software control data transfer between sites, data storage, and batch processing. A diagram of the WLCG tier system is given in Figure 1.32.

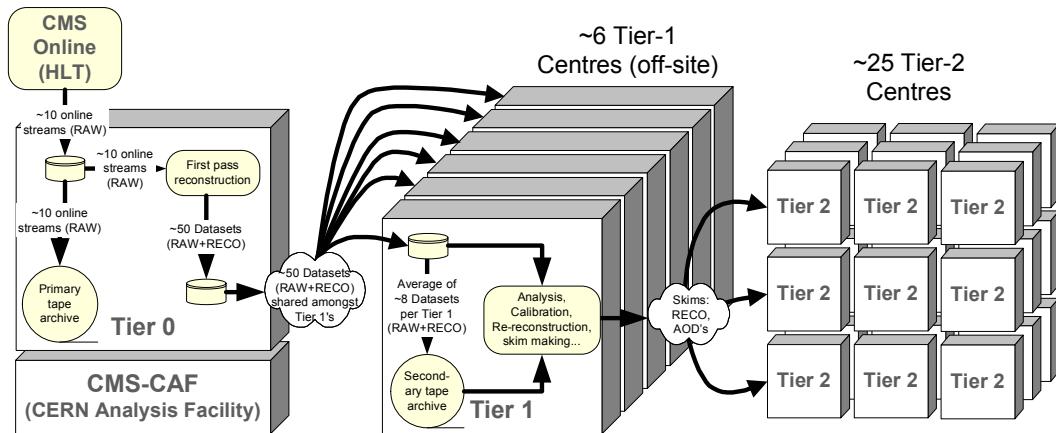


Figure 1.32: Diagram of the WLCG tier system showing data archival and reconstruction at each tier along with data transfer between tiers. Reprinted from Fig. 11.2 of ref. [?].