3.2.1.2 The strip tracker

Figure 3.5: A 3D view of the barrel and forward pixel detector.

The outer part of the tracker consists of 15 148 strip modules, which are distributed among multiple barrel layers and endcap disks and make up a total active area of 198 m². The inner part is composed of 4 Tracker Inner Barrel (TIB) layers with 3 Tracker Inner Disks (TID) on each side. Surrounding these are 6 Tracker Outer Barrel (TOB) layers and the 2 Tracker EndCaps (TEC), which are composed of 9 disks. This geometric arrangement is shown in Figure 3.4, with double lines to indicate back-to-back modules. These so-called double-sided modules are mounted with a stereo-angle of 100 mrad to improve the 3D point resolution by providing a measurement of the z and r co-ordinate in the barrel and disks, respectively.

In the TOB and the 3 outermost rings of the TEC two silicon sensors are daisy chained, while single sensors are used in the inner part. This is done to limit the number of read-out channels, since the area that had to be instrumented is larger in the outer region. However, the noise of the sensors also increases with strip length, so thicker silicon sensors, 500 μm compared to 320μm in the inner part, are used in order to collect more signal per traversing particle.

The strip sensors are single sided p-on-n type silicon. The signals from the sensors are amplified, shaped, and stored by 4 or 6 custom APV25 chips per module. When the trigger has made a positive decision, the analog signals from two APV25 chips are multiplexed and sent to the FED boards in the service cavern via optical fibers, where they are converted to digital signals. The FEDs then perform pedestal and common mode subtraction as well as cluster finding. Additionally, the data is sparsified in these off-detector electronics, before being sent to the CMS central data acquisition (DAQ). Due to charge sharing, this analog read-out scheme also results in an improved spatial resolution of 15 to 40 μm, depending on the position of the modules and the strip pitch.

3.2.1.3 Tracking

The tracks of charged particles going through the CMS tracker are reconstructed with an iterative tracking approach. This is used to cope with the high occupancy and consequently high combinatorics. First, the first iterations search for tracks with less possible combinations, such as tracks with many pixel hits or a high momentum. After every iteration, the hits associated with the found track are removed to reduce the combinatorics. Each iteration consists of four steps:

1. **Seed generation.** In this first step hits are combined into seeds for the subsequent track finding. In the initial iterations pixel triplets are used, then pixel pairs, in order to take gaps or non-working modules into account. Next, mixed pixel/strip triplets are taken, and finally strip-only seeds are used. These additional iterations improve the acceptance in p_T and in displacement with respect to the primary vertex.

2. **Track finding.** The seeds are used as starting point for a Kalman filter algorithm. This method

extrapolates the seed trajectory outward to the next layer, taking into account potential energy loss and multiple scattering. If compatible hits are found in the next layer, the parameters of the trajectory are updated. This process continues until the outermost layer of the tracking system. A Using this method, a given seed can generate multiple tracks, or different tracks can share hits. A trajectory cleaner therefore determines the fraction of hits the tracks have in common and discards the track with the lowest number of hits when there are too many shared hits. If both tracks have the same number of hits, the track with the largest χ^2 value is removed.

3. **Track fitting.** The track parameters are then refitted using a Kalman filter and smoother, taking all hits determined in the track finding step into account.

4. **Track selection.** Finally, the tracks are selected based on quality requirements, such as the number of layers that have hits, the χ^2/dof, and the distance to a primary vertex. This greatly reduces the fraction of reconstructed tracks that are fake.

The performance of the track reconstruction is excellent, and a high track-finding efficiency is obtained [1] while keeping the rate of fake tracks negligible. The highest tracking efficiency is obtained for muons, which traverse the full detector volume and have an improved momentum resolution due to tracking information from the muon detectors giving a long lever arm. For isolated muons with p_T between 1 and 100 GeV the tracking efficiency is higher than 99% for the entire η coverage of the tracker, as can be seen from the left plot in Figure 3.6. The p_T resolution is about 2-3% for a muon with $p_T = 100$ GeV up to $|\eta| < 1.6$, but worsens for higher pseudorapidities. Different types of particles interact differently with the detector material. Charged hadrons, for example, are also subject to elastic and inelastic nuclear interactions and have a tracking efficiency of 80-95% depending on pseudorapidity and transverse momentum, as shown in the right plot of Figure 3.6.
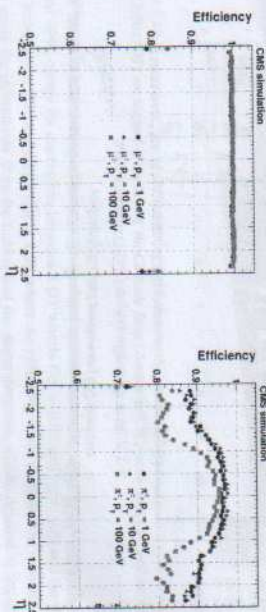

Figure 3.6: The muon efficiency (left) and pion efficiency (right) as a function of pseudorapidity, for multiple transverse momenta. [1]

Finally, the primary vertex is reconstructed from the tracks. Since the collisions happen between bunches of protons, multiple protons will be colliding at the same time. The extra collisions, next to the potentially interesting collision, are referred to as pile-up interactions. The particles generated in these collisions are all detected simultaneously and form a challenge to disentangle them from the particles coming from the to be studied interaction.

The reconstruction is done in 2 steps: first the tracks that appear to originate from the same interaction vertex are clustered, then a fitting procedure computes the vertex parameters and assigns a weight to each associated track, reflecting the probability that it corresponds to the considered vertex. Figure 3.7 shows the reconstruction efficiency and the resolution of the primary vertex. The more tracks, the better the vertex is constrained and thus the better the resolution.
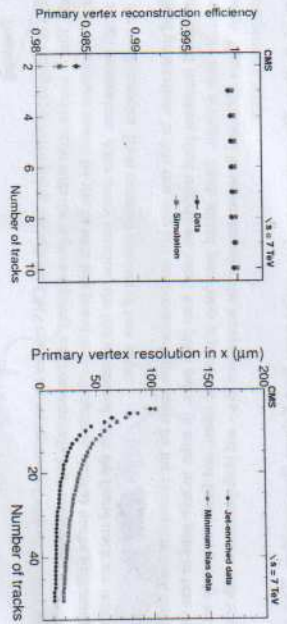
Figure 3.7: The primary vertex reconstruction efficiency (left) and resolution (right) as a function of the number of tracks associated to it. [11]

3.2.2   The electromagnetic calorimeter

Surrounding the tracker, the CMS electromagnetic calorimeter (ECAL) is designed to measure the energy of photons and electrons. It is composed of 75 848 lead tungstate (PbWO$_4$) crystals arranged in a cylindrical barrel and 2 endcaps. The barrel crystals measure 22 × 22 mm² at the front face of crystal, and 26 × 26 mm² at the rear face, which corresponds to approximately 0.0174 × 0.0174 in η-φ. The length of the crystal is 230 mm, corresponding to 25.8 radiation lengths. In the endcaps, the crystals have a rear face cross section of 30 × 30 mm², front face cross section of 28.62 × 28.62 mm², and a length of 220 mm, corresponding to 24.7 radiation lengths.

The high density material was chosen due to its short radiation length and small Molière radius, resulting in a small spread of the electromagnetic shower generated by an incoming photon or electron. This allows for a fine granularity, a better shower separation, and a compact calorimeter. Additionally, this scintillating material has a fast response, as about 80% of the light is emitted during the first 25 ns. The scintillation light is collected by photodetectors, digitized, and read out.

The layout of the ECAL is shown in Figure 3.8, with the barrel (EB) extending up to |η| < 1.470 and the endcaps (EE) on each side covering the range 1.479 < |η| < 3.0. A preshower detector (ES) is positioned in front of the endcap crystals, covering the pseudorapidity range between |η| = 1.653 and |η| = 2.6. This detector consists of a layer of lead which initiates an electromagnetic shower from incoming photons or electrons, and a layer of silicon sensors which measures the deposited energy. The main goal of this 20 cm thick detector is to discriminate between photons and neutral pions.

The energy resolution of calorimeters can be parametrized by the following stochastic (S), noise (N), and constant (C) terms:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \qquad (3.5)$$

The stochastic term represents contributions from the shower containment, the number of photoelectrons and the fluctuations in the gain process. The noise term takes into account all noise components, such as electronics and digitization noise. Finally, the constant term characterizes among others energy leakage from the back of the calorimeter crystals and non-uniformities of the longitudinal light collection. The latter term dominates the energy resolution for high-energy electron and photon showers. Figure 3.9 shows the energy dependence of this resolution for incident electrons as measured in a beam test, as well as the determined stochastic, noise, and constant terms obtained by fitting equation 3.5 to the data.

A more recent measurement of the energy resolution was performed using electrons from Z boson decays in collision data. In the central region, up to |η| < 0.8, it was measured to be better than 2%. Outside of this region, in the more forward direction, the energy resolution is 2-5% [2]. The reconstruction of the electrons and photons will be discussed in Section 4.3.1.
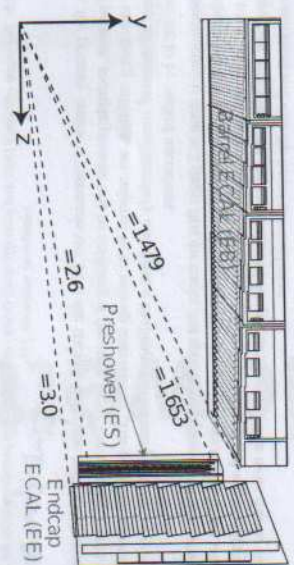
Figure 3.8: A transverse view parallel to the beamline showing one quarter of the ECAL, with its barrel (EB), endcap (EE), and preshower (ES) detectors.
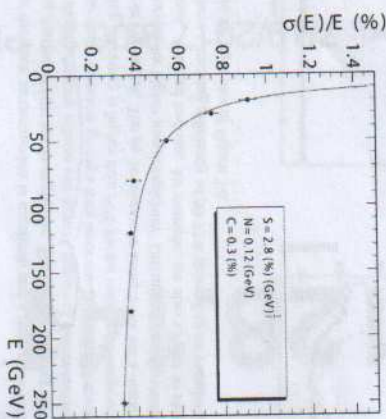
Figure 3.9: The ECAL energy resolution as a function of the electron energy, measured from a beam test. The stochastic (S), noise (N), and constant (C) are given as well.

3.2.3   The hadronic calorimeter

The hadronic calorimeter (HCAL) surrounds the ECAL with the aim to measure the energy of charged and neutral hadrons. The missing transverse energy can then be inferred from this measurement together with the measured energy in the ECAL, in order to identify neutrinos or exotic particles. The HCAL consists of brass absorber plates interleaved with plastic scintillator tiles.

Figure 3.10 shows a longitudinal quarter view of the different HCAL components. A cylindrical barrel (HB) covers the region up to |η| < 1.4 and is complemented by endcaps (HE) on each side, extending the pseudorapidity range to |η| < 3.0. In the central region, the stopping power of the ECAL and HCAL barrel is not sufficient to contain the entire hadron showers. The HCAL was therefore extended outside the solenoid with an outer calorimeter (HO), which uses the the magnet coil as absorber and consists of

scintillators. Two layers are positioned at $\eta = 0$, where the absorber depth is minimal, and only 1 layer is used for the 2 rings on each side of the central ring. Finally, a forward calorimeter (HF) is positioned at 11.2 m from the IP covering $3.0 < |\eta| < 5.2$. Unlike the other HCAL components, this detector consists of iron and quartz fibers. Cherenkov-based radiation-hard technology, since it is exposed to very large particle fluxes.
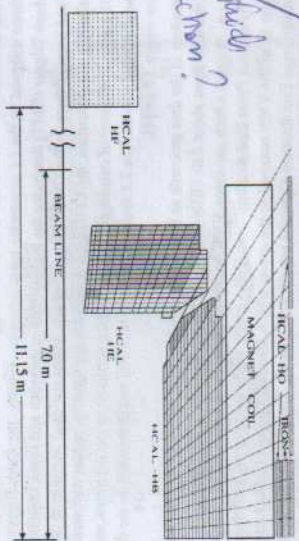

Figure 3.10: A quarter view of the hadronic calorimeter (HCAL), parallel to the beamline. The barrel (HB), endcap (HE), outer (HO), and forward (HF) detectors are indicated.

The optical signals from the scintillators in the HB and HE are converted to electrical signals by multichannel hybrid photodiodes, while silicon photomultipliers (SiPMs) are used in the HO. In the HF, the Cherenkov light emitted in the quartz fibers is detected by standard photomultiplier tubes (PMTs), since the magnetic field is much smaller in this region.

The expected transverse energy resolution for jets is shown in Figure 3.11 for various pseudorapidity regions: barrel jets ($|\eta| < 1.4$), endcap jets ($1.4 < |\eta| < 3.0$), and very forward jets ($3.0 < |\eta| < 5.0$). Details about the reconstruction of jets from calorimeter and tracking information will be given in Section 4.3.3.
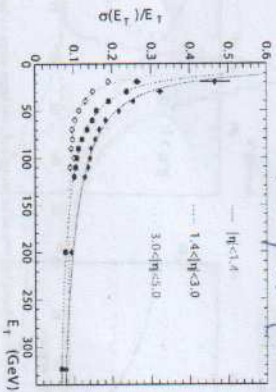

Figure 3.11: The jet transverse energy resolution as a function of the jet transverse energy, for barrel jets ($|\eta| < 1.4$), endcap jets ($1.4 < |\eta| < 3.0$), and very forward jets ($3.0 < |\eta| < 5.0$).

3.2.4 The muon system

The outermost detector, located entirely on the outside of the solenoid, is a dedicated muon detection system. The purpose of this subsystem is muon identification, momentum measurement, and triggering.

As illustrated in Figure 3.12, the layers of muon chambers are embedded in the iron yoke constraining the magnetic field lines. The strong magnetic field completely saturates the return yoke with a field of about 2 T, in opposite direction with respect to the field inside the magnet.

Three different types of gaseous detectors are used. In the barrel, 4 layers of Drift Tubes (DT) are installed, covering the pseudorapidity range up to $|\eta| < 1.2$. Due the higher flux and the larger and non-uniform magnetic field at larger pseudorapidities, Cathode Strip Chambers (CSC) are used in the endcap region ($0.9 < |\eta| < 2.4$). The DTs are designed for the low muon rates that are expected in the barrel and thus have a slower response time than the CSCs. Resistive Plate Chambers (RPCs) complement the DT and CSC systems in the pseudorapidity region up to $|\eta| < 1.8$. They provide a fast response, with a good time resolution but a worse spatial resolution than the DTs or CSCs. The RPCs are therefore very well suited to trigger on muons.
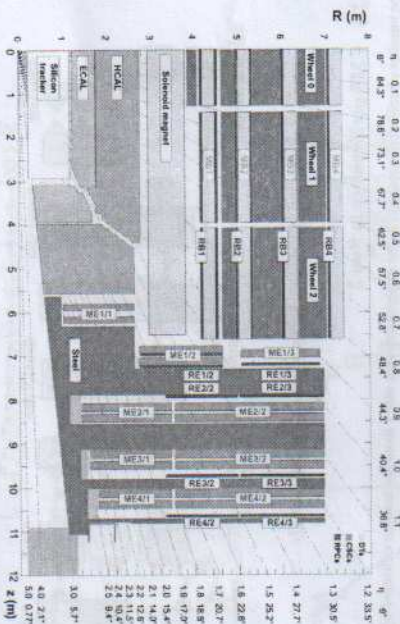

Figure 3.12: A transverse view of one quarter of CMS showing the position of the 3 types of muon detectors. The Drift Tubes (DT) are located in the barrel, the Cathode Strip Chambers (CSC) in the endcaps, and the Resistive Plate Chambers (RPC) in both regions up to $|\eta| < 1.8$.

The offline reconstruction efficiency of simulated events containing one muon is typically between 95% and 99%, except for the regions between 2 DT wheels ($|\eta| = 0.25$ and $|\eta| = 0.8$) and the transition region between the DTs and CSCs ($|\eta| = 1.2$), where the efficiency drops to 92%. The reconstruction of muons using the information from the tracker and the muon detectors will be detailed in Section 4.3.2. For low pseudorapidities and small momenta, the offline momentum resolution of the standalone muon system is about 9%. At momenta around 1 TeV, the resolution varies from 15% to 40%, depending on the pseudorapidity. As demonstrated in Figure 3.13, performing a global momentum fit using the tracker as well improves the resolution by an order of magnitude at low muon momenta. At high momenta the resolution of the full system is about 5%.

3.2.5 Trigger and data acquisition

Collisions are provided by the LHC at high interaction rates, with an interval of 25 ns between bunch crossings. This corresponds to a frequency of 40 MHz. Additionally, multiple collisions occur at the same time, depending on the luminosity. Since it is impossible to store and process the large amount of data produced at this high rate, a severe rate reduction is needed. This rate reduction is performed by the trigger system, which decides whether to store or reject an event. Since this decision process is constrained in time, the computing time is optimized by rejecting uninteresting events
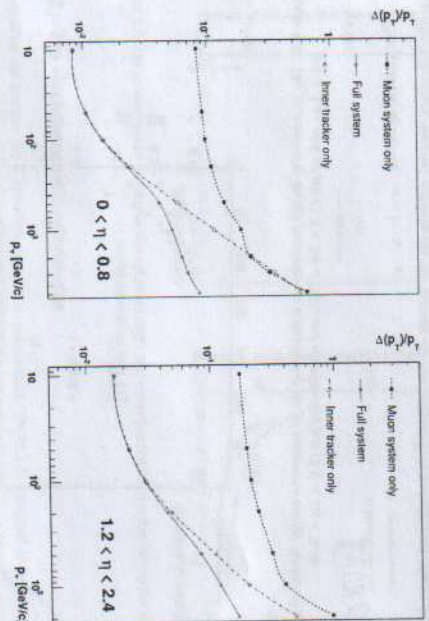
Figure 3.13: *The muon transverse momentum resolution as a function of transverse momentum for low (left) and (high) pseudorapidities. The resolution is shown for the muon system and the tracker separately, and for the full system.* [3]

1 as quickly as possible. The rate is reduced to 1 kHz in two steps by the Level-1 (L1) Trigger and the
2 High-Level Trigger (HLT).

3 The L1 Trigger decision is based on information from the calorimeters and muon systems, following
4 the structure illustrated in Figure 3.14. At the lowest level, the Local Triggers are based on energy deposits
5 in calorimeter towers and track segments or hit patterns in the muon system. Regional triggers, indicated
6 as Calo Trigger Layer 1 and Muon Track-Finder Layer in the figure, then combine this information
7 and use pattern logic to determine trigger objects such as jet or muon candidates in separated spatial
8 regions. The candidates are ranked based on their energy or momentum and quality, reflecting the level
9 of confidence assigned to the L1 parameter measurements. Finally, the Calo Trigger Layer 2 and the
10 Global Muon Trigger (GMT) determine the highest-rank calorimeter and muon objects across the whole
11 detector and transfer them to the Global Trigger, which makes the final decision to accept or reject an
12 event. Following this procedure, the L1 Trigger thresholds are tuned to reduce the event rate to 100 kHz.
13 The L1 Trigger is composed of custom electronics located partially on the detectors, and partially in the
14 underground service cavern. The L1 decision needs to be made and distributed to the detector front-end
15 electronics within 3.8 μs [4].

16 The readout of the data proceeds as illustrated in Figure 3.15. When an event is accepted by the L1
17 Trigger, the data from about 740 FEDs is read out by the Readout Units (RUs). For so-called *legacy*
18 systems, i.e. systems which are using VME-based hardware from the initial installation, the FEDs are
19 read out by custom Front-End-Readout-Link (FRL) cards, while for systems that changed their readout
20 architecture from the VME standard to the newer μTCA standard during or after LS1 they are read out
21 via the newer Front-End-Readout-Optical-Link (FEROL) cards. The event fragments are then sent over
22 the event-builder switch to the Builder Units (BUs), which assemble the events. Next, the events are
23 distributed to the Filter Units (FUs) by a large switch network.

24 The HLT software system is implemented in this filter farm, which uses more than 15000 CPU cores
25 for the final event selection. In this second step, the HLT reduces the event rate further to 1 kHz. The
26 complete read-out data, including information from the pixel and strip tracker, are available for this step.
27 New objects can therefore be reconstructed such as e.g. tau leptons and b-jets, as is done in the offline
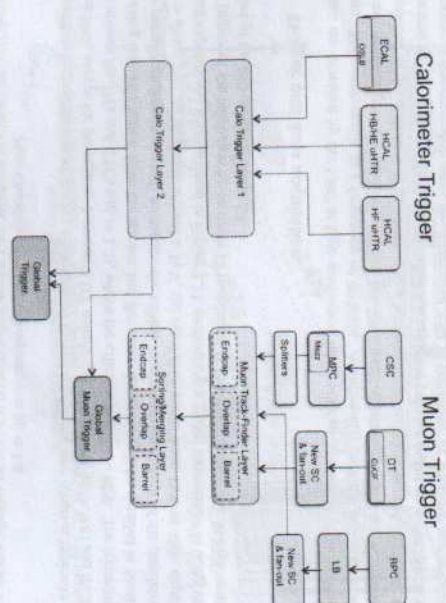28 software, but speed-optimized.
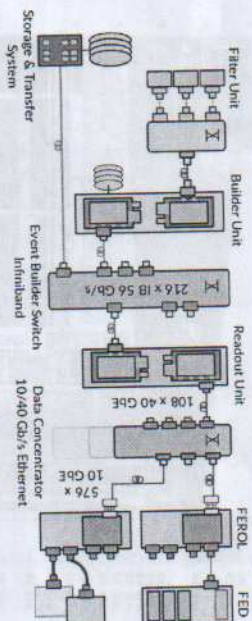
Figure 3.14: *Schematic overview of the L1 Trigger* [4]

Figure 3.15: *Schematic of the data acquisition (DAQ) system.* [5]

3.2.6 CMS performance in Run 2

1 The number of collisions recorded at the experiments will differ from the amount delivered by the LHC.
2 Data loss can be caused by e.g. problems with a particular subdetector, the trigger rate, the data acqui-
3 sition, or the infrastructure. During Run 2, CMS achieved a data taking efficiency of 89% and 92% in
4 2015 and 2016, respectively. The comparison between the delivered and recorded cumulative integrated
5 luminosity in 2016 is shown in Figure 3.16. Subsequently, the recorded data is certified by the offline
6 Data Quality Monitoring (DQM), to ensure that the data are suited for physics analysis.

3.2.6.1 Pre-amplifier saturation in the APV25 chip

8 During Run 2, the instantaneous luminosity delivered by the LHC increased continuously, and even ex-
9 ceeded the design luminosity of 10^{34} cm^{-2}s^{-1} in 2016. As the luminosity increased, a dynamic in-
10 efficiency appeared in the strip tracker, which was most noticeable in the first layer of the TOB. The
11 symptoms were a change in the signal-to-noise ratio and loss of hits. As can be seen from Figure 3.17,
12 the most probable value (MPV) of the signal-to-noise ratio is shifted towards lower values and the low

CMS Integrated Luminosity, pp, 2016, $\sqrt{s}$ = 13 TeV

Data included from 2016-04-22 22:48 to 2016-10-27 14:12 UTC

LHC Delivered: 40.82 fb⁻¹
CMS Recorded: 37.76 fb⁻¹
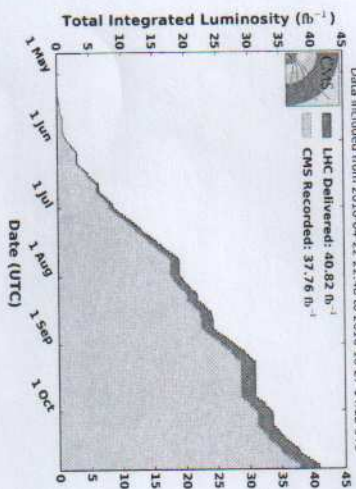
Total Integrated Luminosity (fb⁻¹)

Date (UTC)

Figure 3.16: The cumulative distribution of the instantaneous luminosity delivered by the LHC (blue) and recorded by CMS (yellow) in 2016.

1　tail increased as well. The loss of hits is clearly visible in Figure 3.18, showing the change in number
2　of hits per track for increasing instantaneous luminosities. The run periods indicated in the plot refer
3　to a subset of the data taken over the course of the year. Run period boundaries are typically defined
4　by changes in the LHC running conditions, changes to the detector configuration or calibration, or other
5　parameters. The number of hits decreases for later run periods such as D and F, as the instantaneous
6　luminosity increases. This loss of hits results in less and shorter tracks.

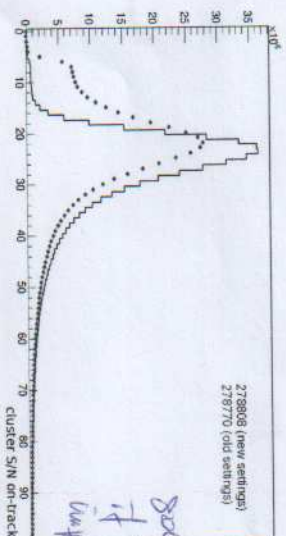273808 (new settings)
278770 (old settings)

cluster S/N on-track

Figure 3.17: The signal-to-noise ratio for clusters on reconstructed tracks in the first layer of the TOB for a run before (blue) and after (black) the change of pre-amplifier drain speed.

source? self-modi: if improve x-axis title and legend

7　　The origin of this inefficiency was eventually tracked down to saturation effects in the pre-amplifier of
8　the APV25 chip. The pre-amplifier decay time changes significantly with temperature. As the operating
9　temperature of the strip tracker was lowered from +4°C to −15°C coolant temperature during LS1, the
10　decay time was no longer sufficient to cope with the high luminosities. The dynamic inefficiency was
11　cured in August 2016 by changing the pre-amplifier drain speed. This led among others to the recovery
12　of the muon efficiency, which showed a large drop for the highest luminosities before the change and an
13　essentially flat behavior afterwards, as demonstrated in Figure 3.19.

Number of tracks

Number of hits

— Run2016B
— Run2016D
— Run2016F

Figure 3.18: The number of hits per track for run periods B, D, and F, showing the effect of the increasing instantaneous luminosity.

Efficiency

Old data
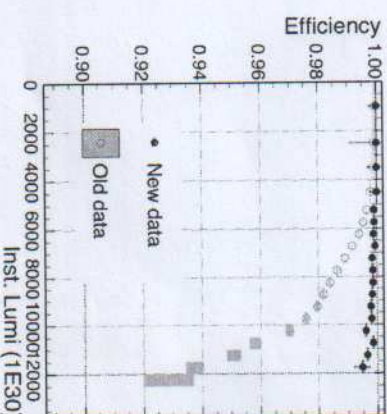New data

Inst. Lumi (1E30)

Figure 3.19: The muon efficiency as a function of the instantaneous luminosity for before (orange) and after (black) the change of pre-amplifier drain speed which cured the dynamic inefficiency.
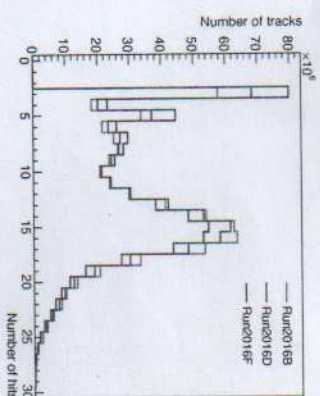
I would add a sentence on how this is going to be a crucial effect in the analysis described in chapter ... such that your readers keep it in mind.