The CMS tracker calibration workflow: experience with cosmic ray data.

Simone Frosali, on behalf of CMS Silicon Strip Tracker
Università degli Studi di Firenze and Istituto Nazionale di Fisica Nucleare (INFN) sez.
Firenze, 50019, Via Sansone 1, Sesto Fiorentino, Firenze, Italy.
E-mail: sfrosali@cern.ch

Abstract. During the second part of 2008 a CMS commissioning was performed with the acquisition of cosmic events in global runs. Cosmic rays detected in the muon chambers were used to trigger the readout of all CMS subdetectors in the general data acquisition system. A total of about 300M of tracks were collected by the CMS Muon Chambers with a 3.8T magnetic field produced by the CMS superconducting solenoid, 6M of which pointing to the tracker region and reconstructed by the Si-Strip Tracker (SST) detectors. Other 1M of cosmic tracks were collected with the magnetic field off. Using the cosmic data available it was possible to validate the performances of the CMS tracker calibration workflows. In this paper the adopted calibration workflow is described. In particular, the three main calibration workflows requested for the low level reconstruction of the SST, i.e. gain calibration, Lorentz angle calibration and bad components identification, are described. The results obtained using cosmic tracks for these three calibration workflows are also presented.

1. Introduction
Compact Muon Solenoid (CMS) is one of the four main detectors placed at the Large Hadron Collider (LHC) based at CERN, Switzerland. It is a general purpose collider detector optimized to discover the Higgs Boson and signatures of physics beyond the Standard Model (SM). The heart of CMS consists in a 13 m long, 5.9 m inner diameter superconducting solenoid, which can produce 4 T magnetic field. Within the solenoid are placed the LHC beam pipe, the tracking detector and the calorimeter. CMS is provided with a system of highly granular silicon microstrip detectors and three layers of silicon pixel detectors are also placed close to the interaction region for the precise spatial measurement of particle vertices. The calorimeter consists of both electromagnetic (Ecal) and hadronic (Hcal) components. Beyond the magnet an iron return yoke and a muon tracking system are placed.

1.1. CMS Microstrip Tracker
The CMS Si-Strip Tracker (SST) [1] covers the pseudorapidity region |η| < 2.5. It is 5.6 m long with a 2.4 m diameter, and consists of 1548 microstrip detectors, named “modules”, for a total of 198 m² of Silicon active area. The tracker is organized in four subdetectors, two barrels and two endcaps, with different characteristics:

- Tracker Inner Barrel (TIB), the innermost cylindrical part, coaxial with the beam axis;
- Tracker Inner Discs (TID), the end-cap of TIB;
- **Tracker Outer Barrel** (TOB), the outermost cylindrical structure, coaxial with the beam;
- **Tracker EndCaps** (TEC), the end-cap of the whole SST.

The barrel subdetectors are divided in 10 layers with different radii (4 layers for TIB, and 6 for TOB) while the endcaps are divided in discs (or "wheels") orthogonal to beam axis. Detectors placed on the endcap discs are organized in *rings* with different radii. Some barrel layers and some rings of the end-cap discs lodge single-sided modules named "mono" (or "rφ") , while the other layers and rings lodge double-sided modules. The latter consist in mono modules glued back-to-back with an other module, named "stereo", tilted with respect to the mono by an angle of 100 mrad. In this way the double-sided detectors can provide precisely measured coordinates of the particle impact point in the sensor plane. The microstrip pitch varies from a minimum of 80 µm, to a maximum of 183 µm. Detector thickness is 320 µm for the modules in the inner part of the tracker (TIB, TID and first four rings of TEC), while the modules in tracker outer part have a thickness of 500 µm (TOB and outermost rings of TEC).

The readout of each detector is performed by the chip APV-25 [2]. The chip takes in input 128 analog channels, each one connected to a strip, and sends the sampled signals, stored temporarily in a pipeline, to the Analog Opto-Hybrid device (AOH) [3]. Each AOH converts the multiplexed electronic signals coming from two APV-25 in optic signals, and transmits them via optical fibres to the front-end readout system (FED) [4] where digitization occurs. On each SST detector 4 or 6 APV-25 are mounted, depending on the number of detector strips, and connected to 2 or 3 AOH respectively.

### 1.2. Low level reconstruction

The CMS offline reconstruction framework, named CMSSW (*Compact Muon Solenoid SoftWare*) is structured in a modular architecture and designed around the Event Data Module [5]. All the user-defined types are contained within a single object, called the "event". All event products (e.g. digitized raw data, clusters, reconstructed tracks etc.) relate to a single triggered readout of the CMS detectors. The access to auxiliary conditions data (e.g. informations about gain and Lorentz angle calibration, noise, bad components etc.) is given via an independent object

![Tracker reconstruction chain](image)

**Figure 1.** Block diagram of the low level reconstruction in the SST. Hit reconstruction requires several calibration tasks that need to be performed using data. In particular the three main calibration workflows requested for the low level reconstruction of the SST are the gain and Lorentz angle calibration, and the bad component identification.
called the “event setup”. The reconstruction chain processes the event products by passing them through a series of user-defined modules, with a schedule defined within a CMSSW path. The modules interact directly with the event and event setup, and no interaction between software modules is allowed.

Low level reconstruction chain [6] is schematically represented in figure 1. In the first step it digitizes the raw data, after cabling calibration applied. Then it proceeds with the clusterization [6], in which the clusters are reconstructed looking for those strips where the collected charge, opportune calibrated by using the gain information, is higher than some definite thresholds of the signal-to-noise ratio. The clusters related to modules, fibres or APVs identified as bad (i.e. noisy or disconnected from readout) are removed from the reconstruction chain. Finally the hits are produced calculating the centroid of the clusters in terms of the local module coordinates. In this last step of the low level reconstruction the Lorentz angle information are needed to correct opportunely the hit position. The hit coordinates can then be converted in CMS global coordinates and used for the subsequent tracking procedure.

In the next paragraphs the algorithms used to evaluate the gain and Lorentz angle calibration, and to identify the bad components are described. In particular the results obtained validating the calibration workflow by using the collected cosmic tracks are shown.

2. Calibration workflow
Raw data are transferred to the first computing center (T0) to be reconstructed. To have a fast calibration feedback, needed by low level reconstruction and alignment, a dedicated Calibration Stream (CS) is used, in addition to the physics events stream. The CS data are reconstructed as soon as they arrive at T0, together with the Express Stream (ES) which contains physics events of particular importance. The output of this prompt reconstruction is kept on a local pool of disks to be available for further processing in the CERN Analysis Facility (CAF). Therefore it is possible in this way to provide the up-to-date calibration constants for the reconstruction of the full statistics. In figure 2 a block diagram of the CMS calibration workflow is shown.

The processing of CS is organized in AlCaReco Producer, i.e. CMSSW jobs which select events according to the High Level Trigger (HLT) bits and skim the data to save disk space in the CAF disk pool. After the first data reprocessing, the calibration constants can be updated

![Figure 2. Block diagram of the CMS calibration workflow. The events of the CS are prompt reconstructed at T0 and are available at CAF for further processing. The calibration constants stored at CAF require transfer via Offline Reconstruction Conditions database ONline/Offline subset (ORCON/ORCOF) for the reprocessing. For the commissioning with cosmic data a simplified version was used, due to the low event rate.](image-url)
in an iterative procedure, using now the full statistics available. The CS contains up to 20% of the total amount of events in case of p-p collisions. On the contrary, in the case of cosmic tracks discussed in this paper, all the available statistics was used as CS from the beginning of the calibration workflow, due to the low event rate.

3. Gain calibration

The charge released in a silicon sensor by a charged particle is digitized into ADC counts assigned to a set of channels making up a cluster. Non-uniformities in the charge collection and in the readout chain can affect the correct amplification and linearity of the response to the released charge. The main components involved in this chain are the silicon wafers, the strips, the APV and AOH chips, the optical fibres and the FED. A significant contribution to possible gain non-uniformities is expected to come from the Linear Laser Driver (LLD) [7] on the AOH. The LLD has been designed with four gain settings, allowing a certain amount of gain equalization. The residual non-uniformity after optimal settings are applied is still expected to be at the level of 15%. Therefore the ultimate equalization of the charge response can only be obtained by looking at signals produced by particles.

The gain calibration using cosmic data was performed at module level, instead of the APV level requested for the LHC scenario. This choice was due to the low statistics available. The procedure that has been developed consists of the following steps:

- Take the charge of the clusters associated to any reconstructed track.
- Normalize the charge to the distance travelled by the particle through the active material of the detector.
- Produce a separate normalized cluster charge distribution for each SST module by excluding clusters spanning across the border between two APV chips or involving the first or last strip of the detector.
- Fit the obtained distributions to a Landau curve. At least 50 hits are required in a module to perform the fit.

![Figure 3. MPV (ADC/mm) distribution for the two thickness-type detectors of the tracker, before gain calibration. The number of entries is related to the APVs, to which is assigned the MPV value estimated for their own module.](image1)

![Figure 4. MPV (ADC/mm) distribution for reprocessed data after gain calibration (larger data sample used). The calibration computed from the reprocessed data are well centered on the chosen reference value (300 ADC/mm) and the width of the distribution is compatible with the error on the Landau fit.](image2)
• For each module, take the ratio of the Most Probable Value (MPV) of the fitted curve to 300 ADC/mm (taken as reference equalization point) and use this ratio as the inter-calibration constant to be applied in the software to correct the charge read on all the strips of the given module. If a module has not enough hits, or if the fit fails (i.e. it provides as MPV a negative number or a value at the border of the fit interval), its calibration constant is set to 1.

The evaluated calibration constants are then stored in a database and can be passed to clusterization module via event setup.

The cosmic tracks used for the gain calibration were reconstructed by the Combinatorial Track Finder algorithm (CTF) \[8\]. The CTF is a Kalman Filter based track finder, which uses the Kalman Filter both for the trajectory building and the estimation of the track parameters. Only events with a track with \(p > 1\) GeV/c were skimmed. It was found that to extract the gain calibration constants from 90% of the SST modules, about 1M of cosmic tracks were necessary. The distribution of the MPV values obtained before the gain calibration is shown in figure 3, separately for 320 \(\mu\)m and 500 \(\mu\)m thickness detectors. The same distribution is shown in figure 4, for data reprocessed by using the evaluated calibration constants.

4. Lorentz angle

Due to the presence of the \(B=3.8\) T magnetic field in the tracker region, the drift direction of the holes collected on the strips of the SST modules is deflected by an angle \(\theta_L\), named “Lorentz angle”. This deflection produces a shift of the cluster position of about \(15 - 25\) \(\mu\)m, depending on the detector thickness and orientation. A correction on the hit position is therefore necessary to evaluate the real traversing point of the particle. The Lorentz angle is related to the Hall mobility \(\mu_H\) of the holes by the equation \(\tan \theta_L = \mu_H B\), and depends (via \(\mu_H\)) on the detector temperature, electric field inside the sensor and absorbed dose.

The Lorentz deflection of the charge carrier drift direction makes the minimum of the cluster width to be achieved no longer for tracks orthogonal to the detector, as for \(B=0\) T, but for tracks incident on the sensor with an angle, in the plane orthogonal to the strips, equal to \(\theta_L\) (see figure 5). The track incident angle for which minimum cluster width is achieved provides therefore a direct measurement of the Lorentz angle. The measurement is performed only for modules of the tracker barrel, since in the endcap modules electric and magnetic fields are parallel and there is no Lorentz drift.

The Lorentz angle calibration provides a database where the measured values of the Lorentz angle are stored module by module. This database is then passed to the hit reconstruction module via event setup to correct the effect of the Lorentz drift on the hit position. The

![Figure 5. Cluster formation in presence of a magnetic field. Tracks incident with a generic angle (a) and with an angle equal to the Lorentz angle (b) are considered. The cluster is represented by the rectangle.](image)
algorithm used to evaluate the Lorentz angle follows this scheme:

- Refit the tracks and evaluate for each associated hit the tangent of track incidence angle (tan θ) in the plane perpendicular to the strips.
- Multiply tan θ by the orientation factor 1/cos α, where α is the angle between the strips and the magnetic field direction. This correction is needed in order to take into account the different orientation of the modules with respect to the magnetic field, which also generates two different signs of θL. Moreover, since only the component of the cluster centroid displacement orthogonal to the module strips is measurable by the detector, the measured shift of the cluster centroid in the stereo detectors is less than the one observed in the mono detectors by the same angle of incidence, because of their 100 mrad inclination with respect to the mono detectors. The factor 1/cos α corrects this effect and makes the sign of θL uniform.
- For each barrel module make a profile with the cluster width vs the tangent of the incidence angle (corrected by the orientation factor). An example of these profiles is shown in figure 6.
- Fit each profile with the function (thickness/pitch) · p1 · |tan θ − p0| + p2, where p0, p1 and p2 are fit parameters, and extract the value of p0 = tan θL. At least 1000 entries are requested to fit the profile. Other quality cuts (tunable by the user) are applied on the fit results.
- If the fit fails or it is rejected by the quality cuts, another fit iteration is performed with a narrower range.
- Extract the value of μH = tan θL/B for each layer and update the database. For the profiles with less than 1000 entries or with fit results rejected by the quality cuts, the μH values already stored in the database are kept.

4.1. Lorentz angle measurement with cosmic tracks at 3.8T

The runs selected for the measurement were taken with a 3.8T magnetic field in the solenoid, for a total of about 5M cosmic events skimmed, each one containing a CTF track with at least 7 hits in the tracker and with pT > 5GeV/c.

For our measurement with cosmic tracks, the produced Lorentz angle database contains the mean value of μH evaluated for each layer by a gaussian fit on the distribution of the μH values...
measured for the modules belonging to that layer (in figure 7 the example of TOB layer 6 is shown). This choice is due basically to the not uniform illumination of the tracker by the cosmic rays. Indeed the use of cosmic tracks makes that only in the top and bottom regions of the tracker the statistics available for each module was sufficient to perform the measurement. On the contrary the modules in the lateral region are not enough illuminated, or have a statistics not uniformly distributed in the profiles of cluster width vs $\tan \theta$. In the former case the fit is not performed, in the latter the fit fails. For the same reason we limited our analysis in the phi region $1 < |\phi| < 2$. Where phi is the azimuthal angle, to avoid outliers (see figure 8).

With these constraints it was possible to perform the Lorentz angle measurement on the 66% of the SST modules. The results obtained by gaussian fits on the layer distributions of $\mu_H$ values are shown in figure 9. The difference bewteen TIB and TOB values is due to the different detector thickness, which causes different values of electric field inside the sensors on equal bias voltage applied. Variations among the layers of the same subdetector instead can be related to differences in temperature and depletion voltage of the modules lodged on each layer.

The Lorentz angle measurement was confirmed by the alignment group. Indeed a set of alignment constants obtained without magnetic field was used to reconstruct cosmic tracks taken at 3.8T. A double peak structure was found in the hit residual plot, due to a difference between an ideal Lorentz angle used in the reconstruction and the real value. Double peak disappears using the measured Lorentz angle, as shown in figure 10.

5. Bad component identification
Some tracker components (e.g. modules, strips, APVs etc.) may be found noisy or not read-out during some runs. It is important to spot these bad components in the reconstruction, since the track reconstruction is dropped if, for two consecutive layers, no hit is found where expected. A procedure has been set-up to transfer the information from commissioning and from power supply system to an offline condition database which is accessed during the reconstruction via event setup. Whenever a module is masked in the database no hit is expected during the reconstruction and therefore the module inefficiency is made available to the tracking. Even few noisy strips could be dangerous for the reconstruction because they may result in a large number of track seeds and therefore engender to a huge increase of the track reconstruction combinatorial. This is especially true for cosmic muons where no beam constraint can be applied to the seeding.
After the first prompt reconstruction of the events, the bad component identification is performed as part of the CMS calibration workflow. The procedure is organized in two steps:

(i) Identification of APVs having an anomalous high occupancy compared to the mean occupancy evaluated for all APVs of the same layer (High relative occupancy algorithm).

(ii) Identification of single, or limited groups of hot strips (Hot strips identification).

5.1. High relative occupancy algorithm

The goal of the high occupancy algorithm is to spot bad components using the cluster occupancy in random trigger events. It uses the assumption that the occupancy is uniform in a given layer/disc, since in random trigger only noise and no physics signals should be present. The procedure to identify bad components is the following:

- Calculate the median cluster occupancy of each single APV. Instead of the mean the median is used to avoid a bias in the APV occupancy if only a few strips have a very high occupancy, which can increase the mean value significantly but does not reflect the overall behaviour of this APV.

- The mean and RMS of all APV median values are computed for each layer/disc separately (e.g. TIB layer 2, TEC disc 6 ...). This is done in an iterative procedure to remove outliers from the calculation.

- By a comparison with the last updated mean and RMS of the APV median values belonging to the layer/disc under investigation, all corresponding APVs are checked for low/high occupancy.

- The resulting bad APVs, fibres and modules are stored in a database.

The applied cuts on the outliers as well as the number of iterations of the procedure are all configurable by the user.
5.2. Hot strips identification
Assuming a constant rate of incident particles and a specific hardware configuration, the distribution of cluster multiplicity can be used to determine the number of noisy or dead channels. The algorithm to identify a “hot” channel in the detector, which may give origin to fake tracks, is thus based on the distribution of the cluster occupancy for each APV. It performs the following steps:

- Calculate the mean and sigma of the number of clusters.
- For each strip plot the number of times it contributes to a cluster. If its contribution exceeds the average number of clusters by more than five times the RMS of the number of clusters, the strip is flagged as “hot”.
- Remove those strips from the calculation.
- Iterate the whole procedure.

In this way, a list of hot strips can be produced and stored in the database. In order to be statistically meaningful, the procedure requires at least 10 signals for a strip in a run and at least 100 signals for a module. An APV can be flagged as bad if a specific fraction (called granularity threshold) of strips are found to be hot. A module is flagged as bad if all APVs are bad. If an APV is flagged as bad, it is taken out of cluster, seeding and track reconstruction algorithms.

In figure 11 a summary of all the modules which were found to have problems in at least one run during the commissioning with cosmics is shown.

5.3. Hit efficiency
To calculate the hit efficiency with cosmic data, track reconstruction is performed using the standard tracking algorithms for cosmic events and excluding the hits of the layer under study, to avoid a bias. The efficiency for modules in this layer is then calculated by finding track trajectories that pass through a given module and looking for the presence of a hit on that module. The procedure to perform the hit efficiency calculation is then:

![Identified SST bad components during commissioning](image-url)
Figure 12. Hit efficiency in each layer/disc including (black) or excluding (red) silicon detectors disconnected or masked.

- Select high quality tracks by applying the following cuts: single track event; tracks with at least 8 hits; maximum of 4 lost hits.
- Require track trajectory passes through center of a module (> 5σ from the sensor edge).
- Check for a track hit on the module under study.

In figure 12 the hit efficiency calculated by using the commissioning cosmic runs is shown. As pointed out in the plot, removing from the hit efficiency calculation the problematic modules found by using the bad component identification algorithms described above, most layers have an efficiency higher than 99%. On the contrary, including all modules the study of hit efficiency may be another useful method to spot the bad components.

6. Conclusions
Commissioning with cosmic runs taken at 0T and 3.8T in the second part of 2008 allowed to validate the calibration workflow of the CMS Si-Strip Tracker, performed with a procedure that should be used for the CMS startup scenario. In particular the algorithms for gain calibration, Lorentz angle calibration and bad component identification were tested, each one providing a database used for the low level reconstruction. The use of this calibration information for data reprocessing showed a sensitive quality improvement, e.g. concerning the alignment and hit efficiency.

References
[1] CMS Collaboration 1998 The Tracker Project Technical Design Report CERN/LHCC 98-6
[2] Lawrence J 2001 APV25-S1 User Guide Version 2.2
[3] Troska J et al. 2003 IEEE Trans. Nucl. Sci. 50 1067
[4] Bell K W et al. 2002 User Requirements Document for the Final FED of the CMS Silicon Strip Tracker
[5] Innocente V et al. 2004 Composite Framework for CMS Applications, Proc. Computing in High Energy and Nuclear Physics, Interlaken (Switzerland)
[6] Bainbridge R J, Cole J, Fulcher J et al. 2008 J. Phys.: Conf. Series 119 022009
[7] Cervelli G et al. 2001 A radiation tolerant laser driver array for optical transmission in the lhc experiments 7th Workshop on Electronics for LHC Experiments, Stockholm (Sweden)
[8] Lenzi P, Genta C and Mangano B 2008 J. Phys.: Conf. Series 119 032030