Available on CMS information server

CMS CR -2009/105

The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

13 May 2009 (v3, 15 May 2009)

Commissioning the CMS Alignment and Calibration Framework

David Futyan

Abstract

The CMS experiment has developed a powerful framework to ensure the precise and prompt alignment and calibration of its components, which is a major prerequisite to achieve the optimal performance for physics analysis. The prompt alignment and calibration strategy harnesses computing resources both at the Tier-0 site and the CERN Analysis Facility (CAF) to ensure fast turnaround for updating the corresponding database payloads. An essential element is the creation of dedicated data streams concentrating the specific event information required by the various alignment and calibration workflows. The resulting low latency is required for feeding the resulting constants into the prompt reconstruction process, which is essential for achieving swift physics analysis of the LHC data. This report discusses the implementation and the computational aspects of the alignment and calibration framework. Recent commissioning campaigns with cosmic muons, beam halo and simulated data have been used to gain detailed experience with this framework, and results of this validation are reported.

Presented at *17th International Conference on Computing in High Energy and Nuclear Physics,21 - 27 March 2009,Prague,Czech Republic,15/05/2009*

1 Overview of calibration and alignment strategy at CMS

The accurate and timely determination of alignment and calibration constants is essential in order to achieve the optimal performance of the CMS detector for physics analysis, and will play a central role in the commissioning of the detector during early LHC data taking.

Calibration and alignment workflows are categorized by their required latency. There are three main categories:

1. Detector-near calibrations: Fast changing quantities such as electronics pedestals and gains, which can vary from one run to the next, are in general determined online, run by run, at the site of the experiment, either during collision data taking, or in dedicated pedestal or pulser runs.
2. Prompt calibration and alignment: Constants which can be determined offline within 24 hours of data taking and are required to be updated regularly are fed back into the conditions database to be used in first full reconstruction ("prompt reconstruction") of the RAW data arriving from the detector. This is referred to as the "prompt calibration loop", and is described in Section 5.
3. Long-range calibration and alignment: Constants determined using a data sample accumulated over a longer period of time, ranging from a few days to a few months, are typically applied to the data in re-reconstruction campaigns.

With the exception of detector-near calibrations which are carried out online, alignment and calibration algorithms are run at the CERN Analysis Facility (CAF) at CERN. An overview of the workflow for prompt calibration is illustrated in Figure 1. The workflow for long range calibration is illustrated in Figure 4. The main components, discussed in the sections which follow, consist of dedicated data streams from LHC Point 5 to the Tier-0, the transfer of dedicated skims of the reconstructed data from the Tier-0 to the CAF, the operation of the calibration and alignment algorithms at the CAF, and the uploading of the constants produced by these algorithms to the conditions database, ready for use in subsequent (re-)reconstruction.

A description of the CMS detector, and of the various alignment and calibration procedures for each sub-detector can be found in [1].

2 Transfer of data streams for calibration and alignment from CMS to the Tier-0

Data is transferred from LHC Point 5 to the Tier-0 in three principal streams:

1. Physics stream: The main data stream intended for physics analysis
2. Express stream: A specially selected subset of the physics stream (approximately 10%, initially more, of the total data bandwidth), chosen according to High Level Trigger (HLT) bits, intended for prompt physics analysis, calibration and alignment, detector and trigger commissioning, and fast physics validation.
3. Calibration Streams: Dedicated streams are set up for calibration and alignment (approximately 10% of the total data bandwidth). These include hardware calibration streams, which transfer events acquired using special triggers during the LHC orbit gap. Orbit gap data is important for pedestal and noise calibration, which needs to be carried out when no signal is present. Orbit gap triggers are also used to transfer laser data used for tracker alignment and for monitoring of electromagnetic calorimeter crystal transparency. In addition to hardware calibration streams, there are dedicated collision event data steams for calibration with special reduced event content. These are referred to as "AlCaRaw" streams.

2.1 AlCaRaw streams

Three calibration procedures require a very high event rate, of order 1kHz each, in order to perform the calibration within a reasonable time, which would saturate the affordable data bandwidth (approximately 300MB/s) if the full event content were transferred. The procedures are:

- calibration of the hadron calorimeter using ϕ-symmetry
- calibration of the electromagnetic calorimeter using ϕ-symmetry

- calibration of the electromagnetic calorimeter using neutral pions and eta mesons

Instead of transferring the full RAW event content in these cases, only the minimal information needed for the corresponding calibration algorithms is transferred. Special high rate triggers are processed on the High Level Trigger farm at Point 5, producing dedicated output streams with event content reduced by factor of around 100 with respect to the full RAW event.

3 Offline processing at the Tier-0: Production of AlCaReco skims

Calibration and alignment algorithms need to run over very large datasets, in some cases with many iterations, within a short period of time. AlCaReco datasets are skims of reconstructed data (both event selection and event content selection), containing the minimal information required as input to a given calibration or alignment algorithm. All AlCaReco skims perform an initial filtering of events based on a set of HLT bits. In some cases, additional processing is performed to produce specialized data collections needed as input to the algorithm. In general AlCaReco skims take reconstructed datasets as input. Some also require RAW information and these AlCaReco skims must be run in parallel with the reconstruction. A subset of AlCaReco skims is run on the express stream as needed for the prompt calibration loop, as described in section 5.

4 The conditions database

Conditions data for CMS are handled by three principal databases, all of which are based on oracle technology. The database used for online data acquisition is called the Online Master Data Storage (OMDS). The conditions subset needed for the HLT and for offline processing is cached at the experiment site in a database referred to as ORCON. The conditions database for use outside the experiment site is referred to as ORCOF. In order to ensure synchronization between ORCON and ORCOF, conditions updates can only be written to ORCON, with ORCOF being exclusively populated through automated streaming from ORCON. Following validation, new constants obtained from calibration and alignment algorithms on the CAF are therefore copied in sqlite form to Point 5, where they are uploaded via a drop-box to ORCON and streamed to ORCOF.

5 The prompt calibration loop

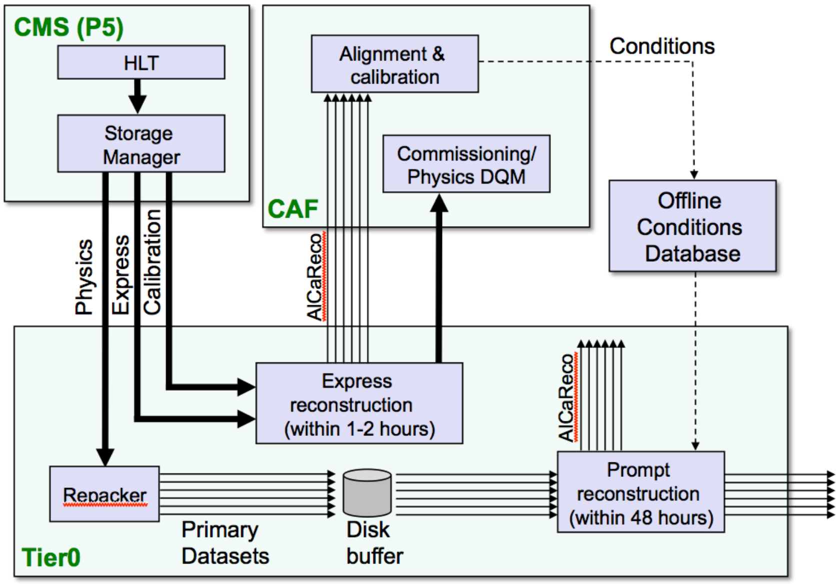

Figure 1: The prompt calibration loop: The main physics dataset is stored on a disk buffer at the Tier-0 for up to 48 hours, while a subset of the physics data set (express stream), together with dedicated calibration streams undergo a prioritized express reconstruction. AlCaReco skims of the express reconstructed data are used as input to alignment and calibration algorithms on the CAF. The resulting new constants are uploaded to the conditions database, ready to be used in the reconstruction of the full physics dataset.

The goal of the prompt calibration loop, illustrated in Figure 1, is to provide good alignment and calibration constants in time for the first (prompt) reconstruction of the RAW data arriving from Point 5. While events in the main physics data stream are buffered at the Tier-0, express and calibration streams are passed through an express reconstruction within one to two hours. Reconstructed data and AlCaReco skims are transferred to the CERN Analysis Facility (CAF), where they are used as input to prompt alignment and calibration workflows to produce new constants. These constants are validated and uploaded to the conditions database. When prompt alignment and calibration are completed, with a target latency of 24 hours, the full data sample is passed through prompt reconstruction using the updated conditions, and subsequently made available for analysis.

6 Commissioning with Monte Carlo

A full scale test of the offline part of calibration and alignment framework formed a major component of the 2008 Computing, Software and Analysis Challenge (CSA08) [2], which took place during May 2008. The overall goal of the challenge was to test the full scope of offline data handling and analysis activities needed for early LHC data-taking operations. Initial mis-alignments and mis-calibrations as expected before collisions were applied to the simulated data, making this the first full-scale challenge with large statistics under conditions similar to LHC startup. The challenge included the operation of the full end-to-end offline workflows for alignment and calibration, using both the Tier-0 and CAF facilities, together with the conditions database, as will be used for real data. The exercise was operated in “real-time”, with the schedule matched to the pace of the reconstruction. The resulting constants were fed into the conditions database and used to re-reconstruct the simulated data for physics analysis, which in turn was carried out as part of the challenge. The full complexity of around 20 concurrent alignment and calibration workflows were operated in parallel, with interdependencies taken into account. An example result from CSA08 is shown in Figure 2, which shows the reconstructed mass peak for Z bosons decaying to two electrons, before, and after calibration of the electromagnetic calorimeter using the ϕ-symmetry technique.

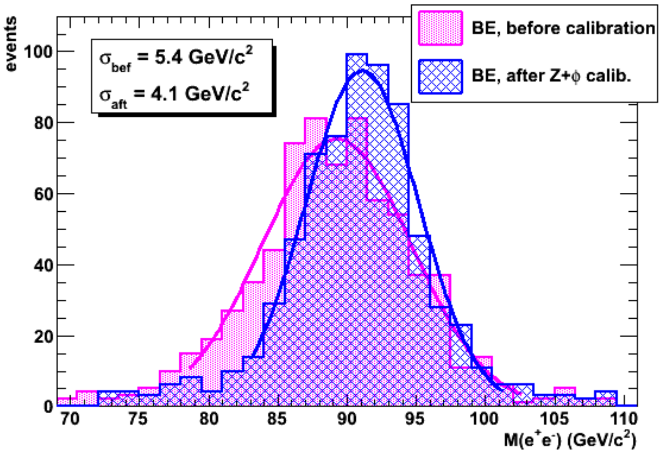

Figure 2: Reconstructed mass peak for Z bosons decaying to two electrons, before, and after calibration of the electromagnetic calorimeter using the ϕ-symmetry technique. Events used for the plot shown were required to have one electron in the barrel and one electron in the endcap. The Z mass resolution improves from 5.3 GeV to 4.1 GeV, validating the improvement in the intercalibration precision of the ECAL endcaps achieved with the CSA08 Monte Carlo sample.

7 Commissioning with Cosmic Data

Full alignment and calibration workflows have operated extensively during cosmic data taking periods in 2008. This included cosmic data taken with the CMS magnetic field off, between July and September 2008, referred to as the Cosmic Run at Zero Tesla (CRUZET), followed by a major cosmic data taking campaign during October and November 2008, the Cosmic Run at Four Tesla (CRAFT). The latter involved a continuous data taking period of four weeks, with the complete detector operating at full magnetic field (3.8 Tesla). A total of 375 million events were collected during this period. An example of a cosmic muon passing through the inner tracker volume and recorded during CRAFT is shown in Figure 3. In addition to cosmic data, data collected from commissioning of the LHC beam on 10th September 2008 was also important for calibration and alignment studies.

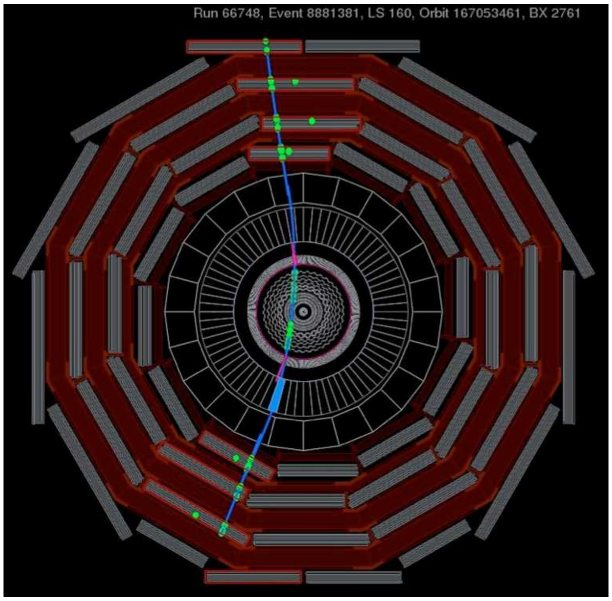

Figure 3: Cosmic ray muon passing through the inner tracker volume of the CMS detector.

A steady ramping up of alignment and calibration workflows was performed during CRUZET, as sub-detectors were increasingly included into global runs, and new workflows were commissioned. During CRAFT, alignment and calibration procedures operated in parallel for all sub-detectors, testing the complete workflow from collection at Point 5 to re-processing of reconstructed data at Tier-1s. The workflow operated during CRAFT is illustrated in Figure 4, and includes AlCaRaw streaming from Point 5 to Tier-0, centralized production of AlCaReco datasets at the Tier-0, and the transfer of AlCaReco datasets to the CAF for the determination of new alignment and calibration constants.

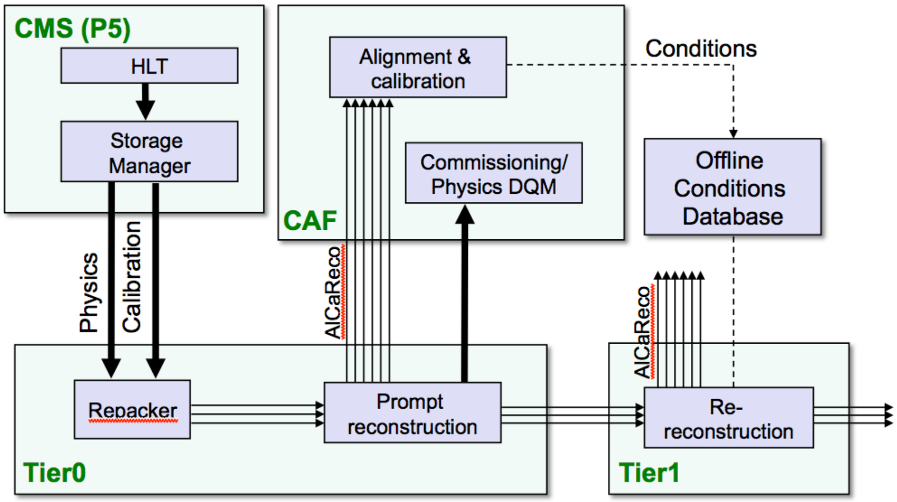

Figure 4: Overview of the workflow for calibration and alignment operated during CRAFT. The workflow includes AlCaRaw streaming from Point 5 to Tier-0, centralized production of AlCaReco datasets at the Tier-0, transfer of AlCaReco datasets to the CAF for the determination of new alignment and calibration constants, which are then used for re-processing of the full dataset at Tier-1 sites. The same workflow is intended for collision data taking for calibration and alignment procedures whose time-scale is too long to be applied in the prompt calibration loop.

Newly determined constants were used for two subsequent re-processings of the full dataset, following formal validation and sign-off procedures. A third full reprocessing is planned using the software version intended for 2009 collision data taking. Significant improvements have been obtained, for example for tracker alignment with respect to survey measurements, as illustrated in Figure 5. Calibration and alignment results from CRAFT have been used to define realistic mis-alignment and mis-calibration scenarios to be used for Monte Carlo simulations intended for comparison with first LHC data.

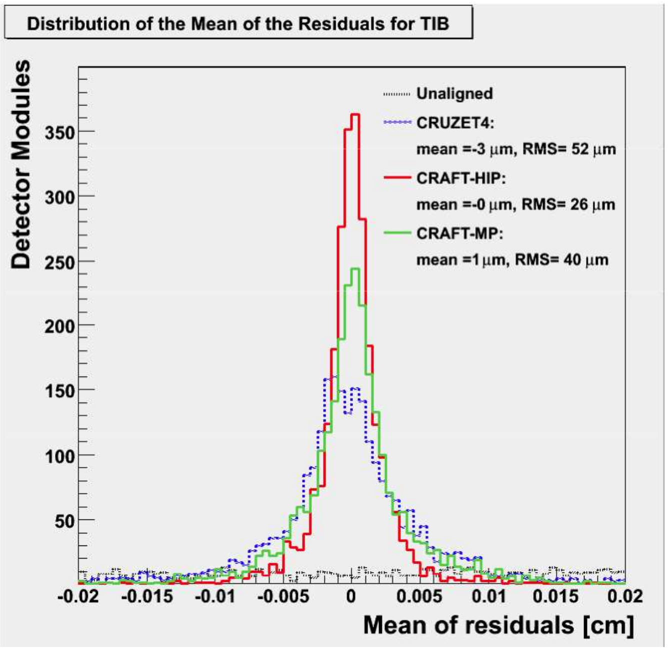

Figure 5: Mean of position residuals for the tracker inner barrel, following alignment using tracks from cosmics muons during CRAFT 2008. The black dashed line at the bottom of the plot shows the residuals obtained using the nominal detector geometry (un-aligned). The dashed blue line shows the result obtained after alignment using cosmics collected during CRUZET. The two solid lines compare the results from CRAFT obtained using two different alignment algorithms. The RMS of the residual distribution obtained using the Hit and Impact Point (HIP) algorithm is 26μm.

8 Summary

A powerful framework has been set up to ensure the precise and prompt alignment and calibration of the various components of the CMS detector. In order to ensure a fast turnaround for updating calibration and alignment constants, the strategy includes the transfer from CMS to the Tier-0 of dedicated high rate data streams for calibration, and the production at the Tier-0 of dedicated skims of reconstructed data, containing only the information needed as as input to calibration and alignment algorithms, which are run at the CERN Analysis Facility. The resulting low latency is required for feeding the updated constants into the prompt reconstruction process, in order to swiftly provide high quality data for physics analysis. The commissioning of this framework has been an important component of recent Monte Carlo and cosmic data taking campaigns. The full offline workflow was successfully demonstrated using simulated data as part of the CSA08 challenge. The complete chain was operated during the CRAFT cosmic data taking campaign, and significant improvements in the alignment and calibration of the detector were obtained. The updated constants have subsequently been applied in two full re-repocessings of the data.

References

[1] CMS collaboration (G.L. Bayatian et al.)., CMS technical design report, volume II: Physics performance, *J.Phys.*G34:995-1579,2007

[2] D. Futyan, R. Mankel and C. Paus, The CMS Computing, Software and Analysis Challenge, Proceedings of the 17th International Conference on Computing in High Energy and Nuclear Physics (CHEP'09), 21 - 27 March 2009 Prague, Czech Republic.