10 July 2003

Design and development of vertex reconstruction algorithm for CMS analysis

Studies of microstrip gas chambers (MSGC) and microstrip silicium detectors.

Stéphanie Moreau
(http://smoreau.home.cern.ch/smoreau/)

Abstract

The work presented in my thesis contributes to the **development of the tracking system** of the **Compact Muon Solenoid** detector (CMS).
My thesis is divided into three sections :

- **Study of microstrip gas chambers (MSGC) in a high intensity beam** : this study focussed on the final irradiation test before their substitution in the CMS tracker.
- The second part describes a **study of silicon microstrip detectors and its electronics exposed to a beam with a 40 MHz sampling rate**. The performance of the data acquisition and control system, as well as signal-to-noise ratio, bunch crossing identification and cluster finding efficiency were analysed.
- Finally, I was concerned with the **design and the development of an ORCA algorithm dedicated to secondary vertex reconstruction**, ORCA is the software framework for the reconstruction and analysis of CMS data. An "Elastic Arms" algorithm has been developped and evaluated in terms of resolution, efficiency, fake rate, purity and CPU time. It is an iterative algorithm usable for b tagging. I also started an analysis of primary vertex reconstruction in events with and without pile up.

1 Introduction

The main goal of particles physics is the study of both the elementary components of matter and of interactions between them, all this is currently described by the ”Standard Model”.

The standard model has been precisely checked by all LEP (Large Electron Prositon collider) experiments at CERN and at the Tevatron. However, several fundamental questions, such as the origin of particle masses, remain open.

My thesis contributes to optimisation of tracking detectors and to the analysis of their data in the framework of CMS (Compact Solenoid Muon), one of the four experiments at the proton collider LHC (Large Hadron Collider). The LHC will start operating in summer 2007. Initially, my thesis focussed on MSGC detectors (microstrip gas chambers) and their electronics. Following the substitution of MSGCs by microstrip silicon detectors (in December 1999), the thesis expanded in the following way :

- Study of microstrip gas chambers (MSGC) using a high intensity beam; the MSGC detectors were initially part of the CMS tracker (chapter 3 of my thesis);
- Tests of silicon microstrip detectors and their electronics with a pion/muon beam with a 40 MHz sampling rate (chapter 4 of my thesis);
- Design and the development of an ORCA (Object-oriented Reconstruction for CMS Analysis) algorithm dedicated to a very efficient secondary vertex reconstruction, in particular for tagging of b-jets (chapter 5 of my thesis) to improve on the quality of measurements in b-physics.

My thesis contains therefore various contributions to the CMS experiment : from a study of technical performances of detector modules to the reconstruction of data taken.

2 Gaseous detectors of the CMS tracker : MSGC+GEM

The initial version of the CMS tracker (before 2000) relied on MSGCs (Micro-Strips Gas Chambers) detectors with a total surface of about 120 m^2 that is to say more than one third of the surface of the CMS tracker.

To reduce the number of potentially damaged strips while keeping the same performance, MSGCs of the forward tracker were equipped by GEMs (Gas Electron Multiplier). The GEMs consist of a thin metal foil, with a high density of holes in the gas volume of the MSGCs permitting to transfer and multiply primary electrons and therefore to reduce the electric field near the strips.

To insure a sufficient performance of this type of detectors, the behaviour of eighteen MSGC+GEM modules was studied between the 19th of october and the 23th of november 1999 in a high intensity pion beam at 350 MeV/c from the PSI (Paul Scherrer Institut) in Villigen (Switzerland); its flux of $\sim$ 4 kHz/mm^2 correspond to the maximum flux estimated for MSGC+GEM modules at the LHC. These eighteen modules represent a sensitive area of 0.72 m^2 (that is to say 1 % of the total area foreseen).

I contributed to the data acquisition of this test ('MF2 test') and evaluated the number of damaged strips. The CMS collaboration set an upper limit of a 10 % loss of all strips for a period of time equivalent to 10 LHC years. In the analysis, 24 lost strips were found, corresponding to about 5.5 % of lost strips per 10 LHC years. This test lead therefore to the conclusion that the operational stability of MSGC+GEM detectors is good enough to include them into the CMS set-up.

After this test and the decision by the CMS collaboration to replace the MSGC+GEM detectors, I completed these studies by analysing :

- the temporal stability during 20 days of the signal-to-noise ratio (at high and at low beam intensity) of all MSGC+GEM detectors;
- the influence of the atmospheric pressure on the signal-to-noise ratio;
- the temporal variations before and after 12 hours of high beam intensity of the signal-to-noise ratio at low intensity which caracterise the polarization of the substrate for MSGC+GEM detectors, that is to say the apparition of charges in the substrate.

The analysis of the PSI-MF2 test data demonstrated stable operation of MSGC+GEM detectors under LHC conditions. The signal-to-noise ratio remains stable and no effect of polarization (charges on the substrates) have been observed over 20 days. Thus, MSGC+GEM detectors meet all specifications of the CMS tracker in terms of operation, construction and performance.

In spite of this success of the MF2 test, the CMS collaboration decided to unify the tracker technology by choosing a tracker entirely equipped with detectors based on silicon semiconductors. This decision is justified by the progress realized in the field of silicon technologies : manufacturing of big size modules, availability of large quantities from industry and consequently a non negligible reduction of costs.

3 Silicon microstrip detectors of the CMS tracker

As a consequence of the change of detector technology, I choose to also study silicon microstrip detectors in a beam test. Six silicon microstrip modules corresponding to a total of 3072 strips were exposed to a test beam (M200 test) at CERN. As this CERN beam was sampling with a rate of 40 MHz (that is to say one interaction per 25 ns) for the first time as at the future LHC, the associated detector electronics deserved special attention.

I contributed to the data acquisition of 6 silicon detectors during 10 days (between October 17 and November 3, 2001). In addition, I wrote part of the C++ analysis code in order to extract raw data (Zebra files) to be analysed in the framework of ORCA (Object-oriented Reconstruction for CMS Analysis).

Using this code, I investigated the signal-to-noise ratio, the synchronisation time (delay) and the efficiency of the detectors to reconstruct hits and tracks. These tests indicated a correct operation of the electronics and of the data acquisition system.

Having shown which of the silicon microstrip detectors resist strong radiation in terms of operation and quality of results, the subsequent phase consists of producing and testing 15 200 of these detectors modules.

4 Vertex reconstruction and "Elastic Arms" algorithm

Distinguishing jets containing b quarks is important for CP violation studies in $b\bar{b}$ events, for top quark measurements and for a discovery of the Higgs boson. The primary signature of b jets is a displaced (secondary) vertex relative to the primary vertex due to the relatively long lifetime (τ) of particles containing a b quark ($\sim 10^{-12}s$, corresponding to a distance of about $2 - 3mm$ at the LHC).

I developped an algorithm for secondary vertex reconstruction based on the 'Elastic Arms' method, especially to tag b jets. The primary vertex is generally reconstructed by a different algorithm; its spatial resolution is superior, as the number of tracks associated to it is rather large, and the reconstruction is done only on a plane transverse to the beam.

4.1 Primary vertex reconstruction

I adapted three different methods ('binning', 'clustering' and 'Gaussian' method) to optimize spatial resolution and efficiency of the primary vertex reconstruction. Resolutions (between 23 and 27 μm) and efficiencies (between 99 % and 97 %) are essentially the same for the three different methods in spite of theim different underlying logic.

The classical binning method is an histogramming technique where the z (coordinate along the beam) of the track at the impact point (point of closest approach to the beam axis) is considered. The clustering method makes use of the same z position and of its error in order to group the tracks along the z axis. Finally, we consider a method using the representation of the tracks by gaussian distributions.

In this framework, I also evaluated effects on the primary vertex reconstruction of pile up of 20-25 events occuring during the same beam crossing as the event of interest. This pile up turns out not to be of major concern for primary vertex reconstruction for the event which has been selected by a trigger. For this investigation it was necessary to take into account :

- the number of tracks associated to each of the 20-25 primary vertices;
- the sum of the transverse momenta of all tracks per primary vertex;
- the average transverse momenta of the tracks associated to each primary vertex.

The use of this last parameter permits to raise the efficiency of reconstruction to about 30 %.

4.2 Secondary vertex reconstruction by an "Elastic Arms" algorithm

I finally designed and developped an algorithm for reconstructing secondary vertices and integrated it into the official code of analysis and reconstruction for CMS data : ORCA. This algorithm is a data-processing program written in C++ based upon the "Elastic Arms" method. It is an iterative algorithm : Each vertex seed (a potentiel vertex defined as the intersection of two tracks) is attracted by all tracks; this attraction is parametrized by a potential of the analytic shape of a Boltzman distribution in χ^2; after typically 15 iterations the algorithm converges to the correctly reconstructed vertex. The closest vertex seeds can be merged such that the correct number of reconstructed vertices is obtained. After optimization of all parameters of this algorithm (by using Monte Carlo simulation), spatial resolution, reconstruction efficiency, fake rate and CPU time consuption were evaluated for simulated $b\ \bar{b}$ events, that is to say p-p collisions with a $b\bar{b}$ pair in the final state.

The spatial resolution obtained for secondary vertices (first results in ORCA) is about 100 μm for the transverse (to the beam) coordinate and $\sim$ 135 μm for the longitudinal coordinate. The secondary vertex efficiency of about 30 % with a purity of 60% is quite similar to that from other programs. The decisive property of this "Elastic Arms" algorithm is its very low fake rate of about few percent compared to about 50 % from competive codes. However, with a CPU time of 531 ms per event and per iteration, the "Elastic Arms" algorithm is twice slower for the time being than the other programs. An optimization is currently under way.

Application of this "Elastic Arms" algorithm to the analysis of $b\ \bar{b}$ events yields an estimation of the number of reconstructed secondary vertices per event. Based on this number, the relative number of jets containing b quarks which can be tagged only by this algorithm was evaluated : 50% of b jets can be identified so far against 3% of fake b jets.

Some more improvements should increase the efficiency and reduce the time CPU needed for vertex reconstruction; they are based upon :

- a quick elimination of a bad "vertex seed";
- optimization of the attractive potential to get a stable solution in a shorter time.

Last, not least, I have developped several C++ programs to visualize the position of primary and secondary vertices.

5 Conclusion

My thesis focussed first on the MF2 test of eighteen MSGC+GEM modules in a high intensity beam and yielded a small number of lost strips ($\sim$ 5 %) and a stable signal-to-noise ratio. But, in spite of the success of the MF2 test, the CMS collaboration has however prefer to unify technology and to choose an all-silicon tracker.

I pursued tests of CMS tracker modules and contributed to the test of six silicon microstrip detectors in a beam with a 40 MHz sampling rate. Inspection of the collected data showed reliable operations of detector electronics and of the data acquisition system.

Switching from detector studies to software developpment, I tested three different methods for primary vertex reconstruction. In spite of their different logic, primary vertex resolution and efficiency are comparable. Pile-up effects do not seem to be a major problem for primary vertex reconstruction. My ”Elastic Arms” algorithm yielded the first results for secondary vertex resolution, with an efficiency comparable to that from other programs, but with a very low fake rate of a few percent. This algorithm permits to identify about 50% of b jets with 3% of fake b jets.

The CMS Experiment is a remarkable adventure on the human, technological and physics level; I am really happy to have contributed my modest share and appreciate the help of my colleagues on hardware and software issues.