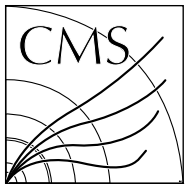
Available on CMS information server

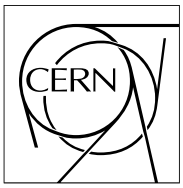
CMS NOTE-2008/005



The Compact Muon Solenoid Experiment

CMS Note

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



18 June 2007

Particle Identification with Energy Loss in the CMS Silicon Strip Tracker

A. Giammanco

Abstract

Some complementary methods are proposed for an identification of charged particles based on dE/dx in the silicon microstrip modules of the CMS Tracker. The performance of proton selection is discussed as benchmark, and the impact of the main systematic effects is estimated. Strategies are presented for the validation and calibration of the methods with several categories of data.

1 Introduction

The measurement of the energy lost per unit pathlength (dE/dx) by charged particles in a medium, in conjunction with the measurement of their momentum, is traditionally used in nuclear and particle physics as a powerful tool for particle identification.

In the CMS detector, although no specific sub-detector has been devised for mass-based particle identification, it is nevertheless possible to measure the energy deposited into the silicon layers of the Tracker. The signal height from a single microstrip or pixel is related to the number of electron-hole pairs created by the traversing particle in the bulk of the silicon sensor; an off-line analysis can make use of the corresponding number of ADC counts. The precision of the dE/dx determination in the CMS pixels is limited by the non-linearity of their response and by saturation effects, while for microstrips a good linearity is achieved.

The dE/dx of the track can provide additional information for the identification of electrons in jets, complementing the information from the Electromagnetic Calorimeter (ECAL), and most of all it is the only observable, among those accessible to CMS, able to discriminate between different charged hadron species.

Both the electron and the hadron identifications are of great importance in the development of a "Particle Flow Algorithm" (PFA), whose goal is to improve the jet identification and energy measurement by combining, instead of using solely the calorimetric tower information, the information from all CMS sub-detectors.

Ideally, in a PFA, one would like to identify and reconstruct each particle individually in an event, in view of getting the best possible determination of their origin, their four-momenta, and the various individual corrections and calibrations for each of these particles. In practice, the discrimination between different charged hadrons species is only possible at relatively low momenta, since it must be based on mass differences and thus is unfeasible for ultra-relativistic particles. Nevertheless, it is exactly under these conditions (i.e. when the mass of the particle is at least comparable to its momentum) that it is useful to identify the hadron in order to correct for its mass. This is particularly important for low-E_T jets, whose identification is crucial in analyses with "jet vetoes" and for which correcting the energy of a proton using its mass instead of assuming the pion mass can make a non-negligible difference of around 1 GeV for the individual jet constituents.

Counting the fraction of stable or long-lived hadron species in jets or in minimum-bias events, at least on a statistical basis, provides a test for fragmentation models. These are semi-empirical models tuned on measurements from other colliders, and reproduced in the Monte Carlos under the assumption of universality.

Although this paper deals with the identification of hadrons, an important motivation for the development of a tool for dE/dx measurement is that a large energy loss is one of the most characteristic signatures of long-lived massive charged particles as predicted in some "new physics" models [1].

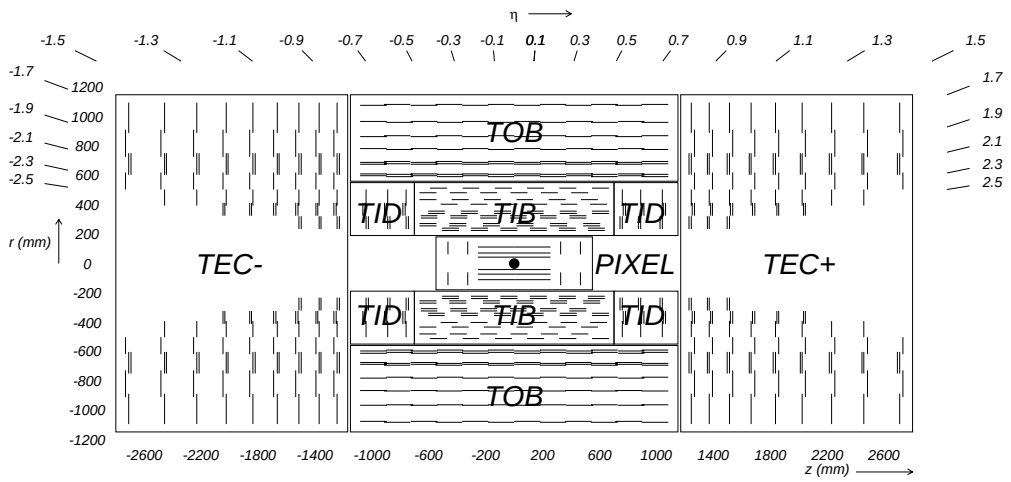


Figure 1: Layout of the Silicon Strip Tracker (r-z view) [3]. Double-sided modules are represented by double lines.

Section 2 describes how the dE/dx measurement is performed in the CMS Tracker, at the hit level. The validation of the simulation with cosmic data, in particular, is discussed in 2.2. The following two sections describe two complementary approaches to dE/dx-based particle identification at the track level: in Sec. 3 the construction of

simple most-probable-value estimators for the dE/dx to be associated with the track, compared on the basis of resolution and discriminating power at low momenta; in Sec. 4, instead, a method where the dE/dx measurement in any hit is used to estimate a compatibility (or incompatibility) with the given particle hypotheses, and all the hits in a track are combined to give a continuous discriminator. In Sec. 5 the two approaches are applied to proton identification in minimum bias and QCD events. Section 6 proposes some calibration strategies, valid both for absolute and relative calibration of the detector units, and based both on special runs and on *in situ* measurements. In Sec. 7 the main systematic uncertainties are discussed, and their impact on the performance of the presented methods is estimated. In Sec. 8 some conclusions are drawn, as well as some proposals for further applications of the methods.

2 Specific energy loss in the CMS Tracker

A study already exists, in CMS, for a dE/dx measurement with the Pixel Tracker [2], which is the innermost part of the CMS Tracker. This determination of dE/dx can provide a good discrimination between pions, kaons and protons in the regime of very low transverse momenta.

This paper, instead, focuses on the dE/dx measurement in the Silicon Strip Tracker (SST). The double-sided layers have pairs of modules glued back-to-back, so that each of them gives two independent dE/dx measurements. The overlaps between modules belonging to the same layer occasionally provide additional hits (the overlap area amounts to roughly 5%). The total number of hits per reconstructed track, without counting the pixels, is shown in Fig. 2.

All the modules of TIB and TID and the innermost rings of TEC are composed of silicon sensors with a thickness of 320 μm, while TOB and the outermost rings of TEC house sensors of 500 μm thickness. All the pixel sensors are 285 μm thick in the barrel and 275 μm in the endcaps.

The SST is composed, as seen in Fig. 1, by [3, 4]:

- the Tracker Inner Barrel (TIB), with four layers: two single-sided and two double-sided;
- the Tracker Outer Barrel (TOB), with six layers: four single-sided and two double-sided;
- the Tracker Inner Disks (TID), with three disks per side, each disk divided in three rings: one single-sided and two double-sided;
- the Tracker End-Caps (TEC), with nine wheels per side, each composed by a number of rings including single-sided and double-sided modules.

The Pixel Tracker is not used for dE/dx measurement in this study, but its clusters are used for the track reconstruction and momentum measurement; it consists of:

- the Pixel Barrel (PXB), composed of three layers;
- the Pixel Forward (PXF), with two disks per side.

2.1 dE/dx measurement for single hits

In this study the official CMS software (CMSSW, version 1.3.0) is used. The tracks are reconstructed with the Kalman Filter, with the standard CMS parameters, unless differently specified. Details can be found in Ref. [5]. The most relevant settings for what concerns this study are the lower p_T cut (0.9 GeV/c) at seeding level and the minimum number of hits per track (five, including hits in the Pixel Tracker). With this configuration, the fake rate is below 10% [5].

Each track crosses several pixel and microstrip modules, and each crossing gives an independent dE/dx measurement (Fig. 3). Here the derivative dE/dx is approximated by the quantity $\Delta E/(\Delta L \cdot \sec\theta)$, where ΔL is the effective thickness of the active volume of the silicon module at the particle entry point (taken as the nominal thickness of the module minus 30 μm of passivation layer[1)]), θ is the angle between the track direction and the

[1)] At the time of writing, this is not included in the simulation and all the nominal sensor volume is treated as active material. This is taken into account when comparing data and Monte Carlo in Sec. 2.2.

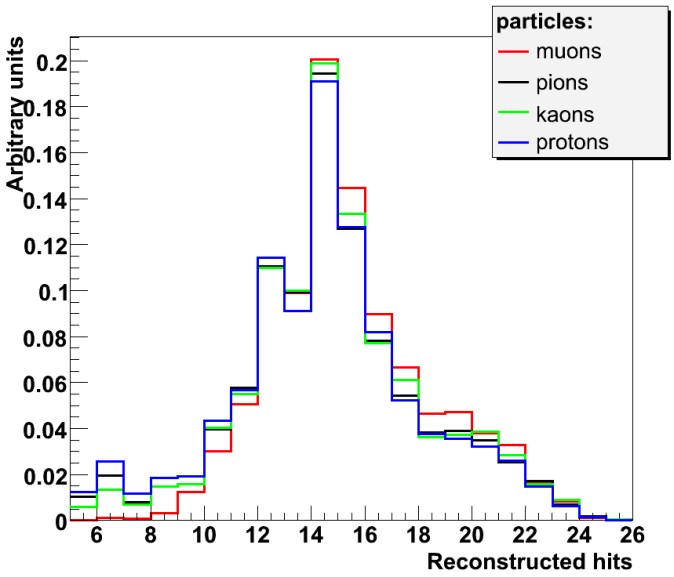


Figure 2: Number of reconstructed hits per track in the SST, for single particles with flat p_T between 3 and 5 GeV/c, reconstructed with the Kalman Filter. Double sided modules give two hits per crossing. The difference between the hadron and muon shapes is attributed to nuclear interactions.

axis normal to the module, and ΔE is the energy deposited by the charged particle in the module, measured by the integrated charge of the "cluster", i.e. the set of contiguous microstrips or pixels that measure the crossing point of the track in the module.

Since the signal is measured in ADC counts, a calibration factor has to be applied in order to have it in units of MeV. In this study the value of 250 electron-hole pairs per ADC count is used for the SST modules, unless differently stated. This is further multiplied by the 3.61×10^{-6} MeV energy needed for the creation of each electron-hole pair. This value holds for the operating temperature of the Tracker, i.e. $-10^{o}C$. At room temperature this value becomes 3.62×10^{-6} MeV.

The two modes of operation of the APV chip [6] are both simulated: "peak mode" and "deconvolution mode", described in Ref. [7]. The simulation of the detector response takes into account, among other things, the production of δ-rays (whose threshold is set to 120.4 keV, corresponding to a range of 0.1 mm in Silicon), the Lorentz angle, the noise (generated as the tail of a gaussian distribution, parameterized as a function of the strip length), a cross-talk of 12% between neighboring strips, and the CR-RC shaping curve [8], folded with the deconvolution algorithm, with time constant 12.06 ns [9]. In deconvolution mode the signal recorded for out of time particles is smaller, and the delay of a massive particle with respect to the time of flight of a relativistic particle is also taken into account in the simulation. This is anyway very small, e.g. ≈ 2.5 ns in the outermost Tracker layer for a particle with $\beta \approx 0.8$. The uncertainty due to of out-of-sync detector units is estimated in Sec. 7.2.

The hit-by-hit measurement of dE/dx in simulated single-particle events is shown in Fig. 3, separately for each sub-detector.

2.2 dE/dx of cosmic muons at MTCC

A first opportunity to test the outcome of the dE/dx extraction on real data came from the Magnet Test / Cosmic Challenge (MTCC), whose description can be found in Refs. [11, 12]. The validation and tuning of the CMS simulation of the silicon modules response with MTCC data, based in part on the energy loss measurement itself, is described in a dedicated section of Ref. [12]. Only the information most relevant for the present paper are repeated here.

The Tracker setup for the MTCC consisted of roughly 1% of the final number of channels. The instrumented parts were two layers of the TIB (with 15 double-sided 6-APV modules and 45 single-sided 4-APV modules respectively), two layers of the TOB (with a total of four "rods" inserted, two containing 6 single-sided 4-APV modules, the others 6 single-sided 6-APV modules), and one disk of TEC (with two "petals", each holding 17 silicon strip modules).

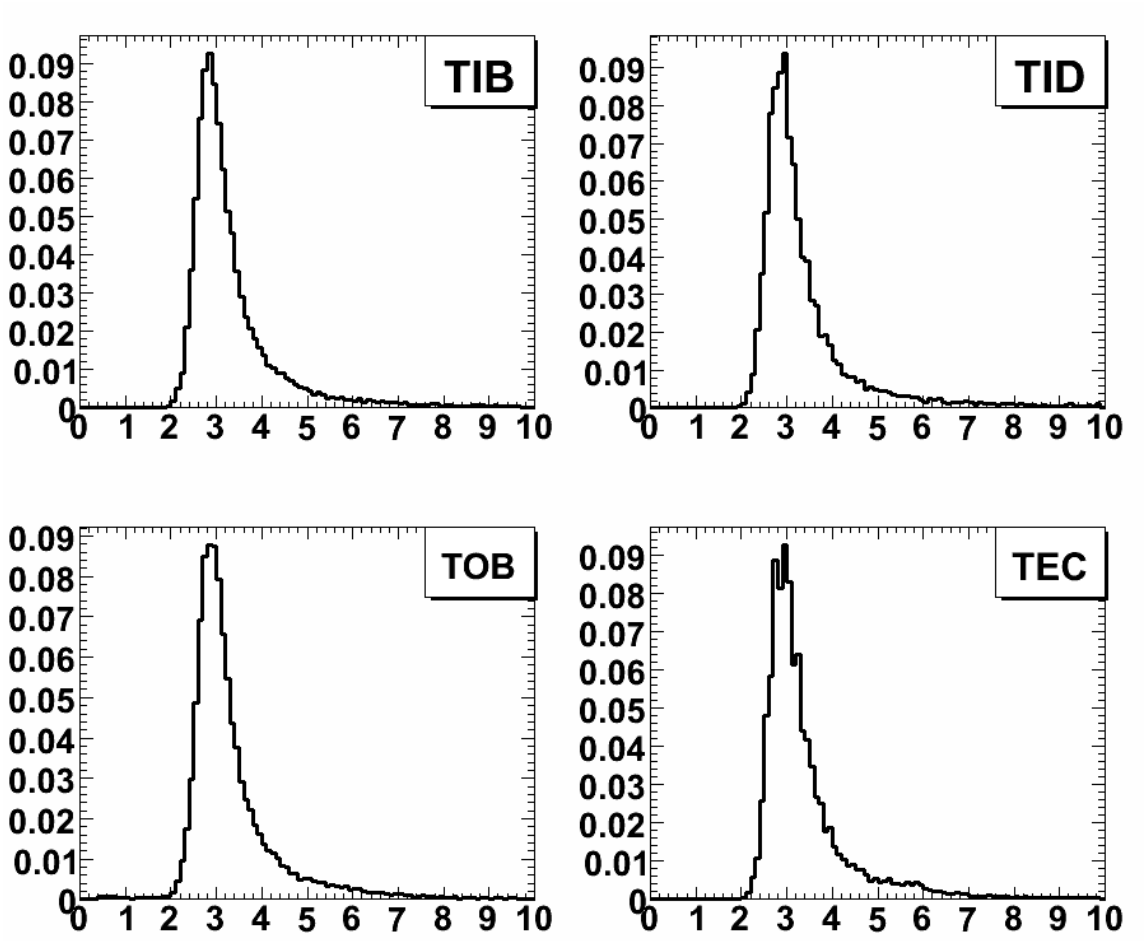


Figure 3: Hit-by-hit dE/dx in MeV/cm for simulated single muons with flat p_T between 3 and 5 GeV/c, separately for all the Silicon Strip Tracker sub-systems.

Only "grade B" modules were used in the Tracker setup for the MTCC. This means that the general performances of the system, in terms of noise and uniformity, are worse than the ones expected for the final setup.

Tracks are selected with the "cosmic track-finder", a specific algorithm optimized for muons not coming from the interaction point [12]. Runs with magnetic fields between 0 and 4 Tesla have been analyzed. A cosmic muon giving a trigger signal in the muon chambers, after crossing the Tracker modules, had to traverse the iron layers of the calorimeters and of the muon chambers themselves, resulting in an implicit cut on momentum that ensures that the particle can be treated as a MIP. Due to the very limited acceptance of this Tracker setup (at least three hits are required by the cosmic track-finder), only a small fraction of the triggered events contains a reconstructed track in the Tracker system (i.e., about 3500 tracks are reconstructed out of ≈ 15 million events triggered and registered during the runs with magnetic field ≥ 3.8 T).

The hit-by-hit measurement of dE/dx, for hits belonging to reconstructed tracks, is shown in Fig. 4 for TIB and TOB modules (TEC modules, due to their position, were hit by very few of the reconstructed tracks) in MTCC data and in a simulation based on GEANT 4.8.1 [13]. Data have been corrected via the 'tickmark calibration' procedure described in Sec. 6.1.

3 dE/dx most-probable-value estimators for tracks

This section describes several methods to attribute a dE/dx estimator to the track, by proper combinations of the dE/dx measurements in each of the Tracker modules traversed[2)]. The simple arithmetic mean is clearly not the

[2)] The loss of energy after each traversal is practically negligible with respect to the momenta considered, and this justifies the assumption of a unique dE/dx value along all the trajectory of the particle.

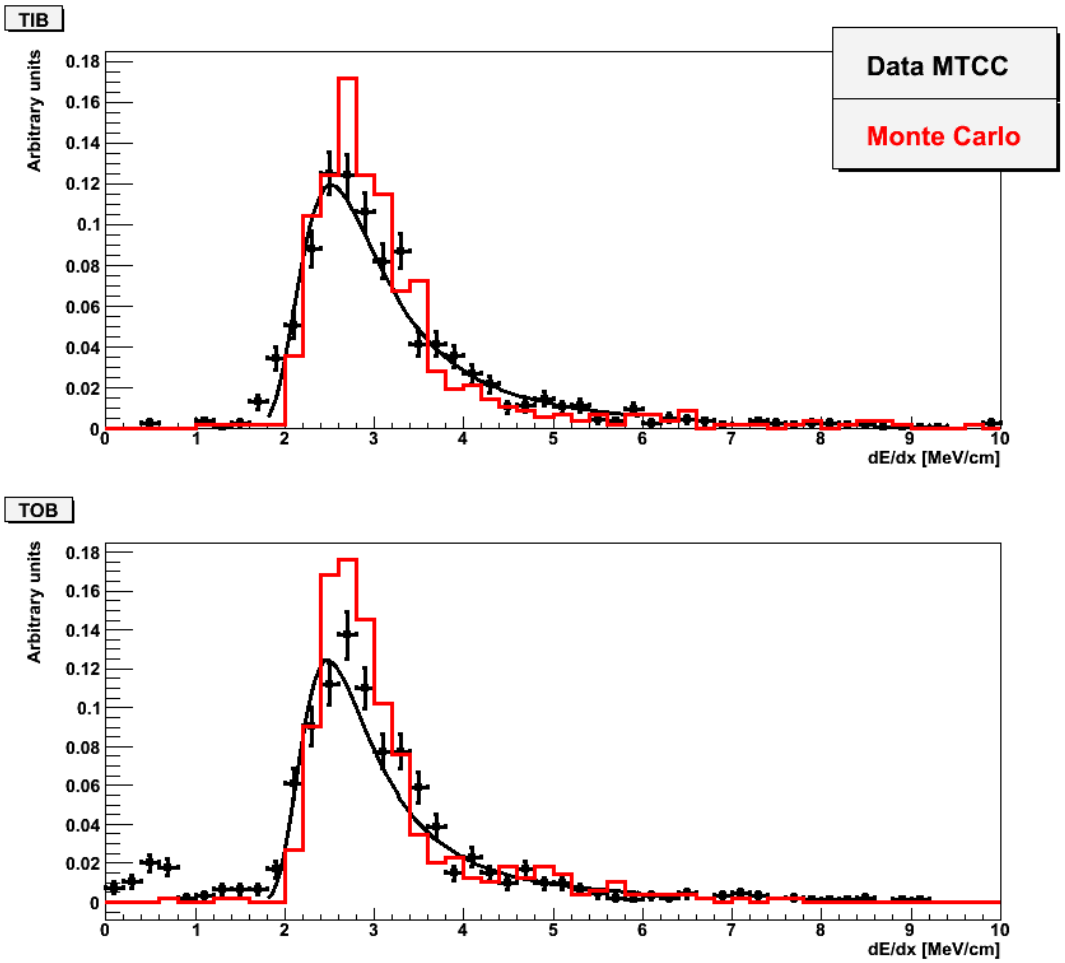


Figure 4: dE/dx measurements from single hits in TIB (top) and TOB (bottom) belonging to cosmic tracks, as recorded in the MTCC with $B = 3.8\ T$ and with the calibration described in the text (black dots with error bars, with a Landau distribution fitted superimposed) and in a simulation of the MTCC setup [12] (red histogram).

best estimator, due to the long tail of the energy loss distribution.

Several estimators are compared, applied to single-particle events originating from the LHC interaction region, with the goal of selecting a fair and robust estimation of the Most Probable Value (MPV) with such a small number of measured energy losses. With the only exception of the Landau fit, all of them are algorithmically very simple, so that they are suitable for being used in a new trigger path that could be envisaged specifically for the search of heavy stable charged particles (same selection as the non-isolated inclusive muon trigger, plus an anomalous energy loss).

For all these estimators at least three hits in the SST are required, otherwise the track is ignored.

In order to quantify the relative merits of the different estimators considered in this section, a gaussian fit is performed on the distribution of the logarithms (which tends to be fairly symmetric in the peak region) and the comparison is done in terms of standard deviation of this distribution, and of number of standard deviations of distance between different hadron species for momenta between 900 and 1100 MeV/c.

3.1 Median

The median of a distribution is taken by arranging all the observations from lowest value to highest value and picking the middle one. If the number of observations is even, the median is computed as the mean value of the two middle values.

It is often used as MPV estimator for skewed distributions (like the Landau distribution) and is robust against the presence of outliers. Nevertheless, when the statistics are very low it has a strong sensitivity to the tails of the

distribution.

The separation power for this estimator as a function of the reconstructed momentum is shown in Fig. 5(left). Fig. 5(right) shows the same estimator in a slice of momentum between 900 and 1100 MeV/c.

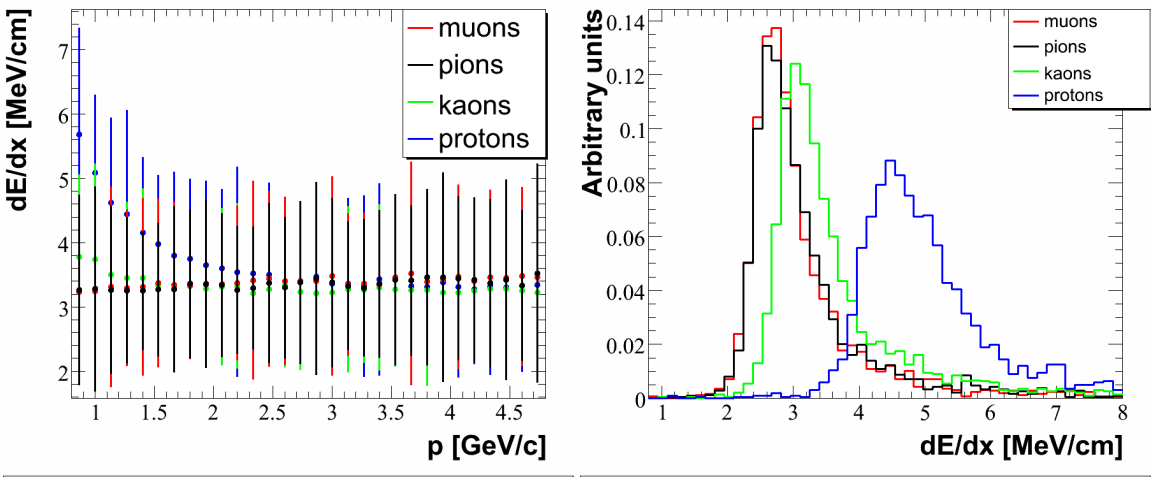


Figure 5: Median dE/dx as a function of the track momentum for different particle species (top left) and in the momentum range between 900 and 1100 MeV/c (right). Proton/pion separation as a function of momentum (bottom left). The error bars in top left are the RMS of the distribution.

In order to better quantify the separation power in an interesting kinematic regime, Fig. 6 shows the distribution of the natural logarithm of this estimator between 900 and 1100 MeV/c, with a gaussian fit superimposed.The resolution of this distribution (standard deviation divided by mean value) for muons is 14.5%.The distance between the kaon and pion distributions is 0.6 standard deviations, the proton/pion distance is 2.5 standard deviations.

3.2 Truncated mean

The truncated (or trimmed) mean is a crude but often effective way to estimate the MPV by cutting away part of the tail. The highest values (in a fixed fraction of the total number of observations) are discarded, and the arithmetic mean is computed for the remaining values.

It has little sensitivity to outliers, but it has some severe drawbacks: it introduces a bias which depends on an arbitrary parameter (the truncation fraction), and it wastes part of the already very limited statistics.

Here results are explicitly shown for two values of the truncation fraction: 20% (a value very often found in literature, see for instance Ref. [14]) and 40% (used as default in the old CMS software, where the truncated mean was the only estimator available for the track dE/dx).

The separation power for 40% truncation as a function of the reconstructed momentum is shown in Fig. 7(left). Fig. 7(right) shows the same estimator in a slice of momentum between 900 and 1100 MeV/c.

The resolution for muons is 6.8%(6.3%) for 20%(40%) truncation.The distance between the kaon and pion distributions is 1.3(1.5) standard deviations, the proton/pion distance is 4.8(5.3) standard deviations.

3.3 Generalized mean

The harmonic mean (or subcontrary mean) of a set of observations is defined as the number of terms divided by the sum of the terms' reciprocals.

Here the "generalized mean" of grade k of a variable x is defined as:

$$M_k(x_1, ..., x_n) = \left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^k \right)^{\frac{1}{k}}, \tag{1}$$

with the usual harmonic mean being the case $k = -1$. The case $k = -1/2$ was chosen by the H1 experiment at HERA [15]. This family of estimators has been considered in this study due to its good properties with respect

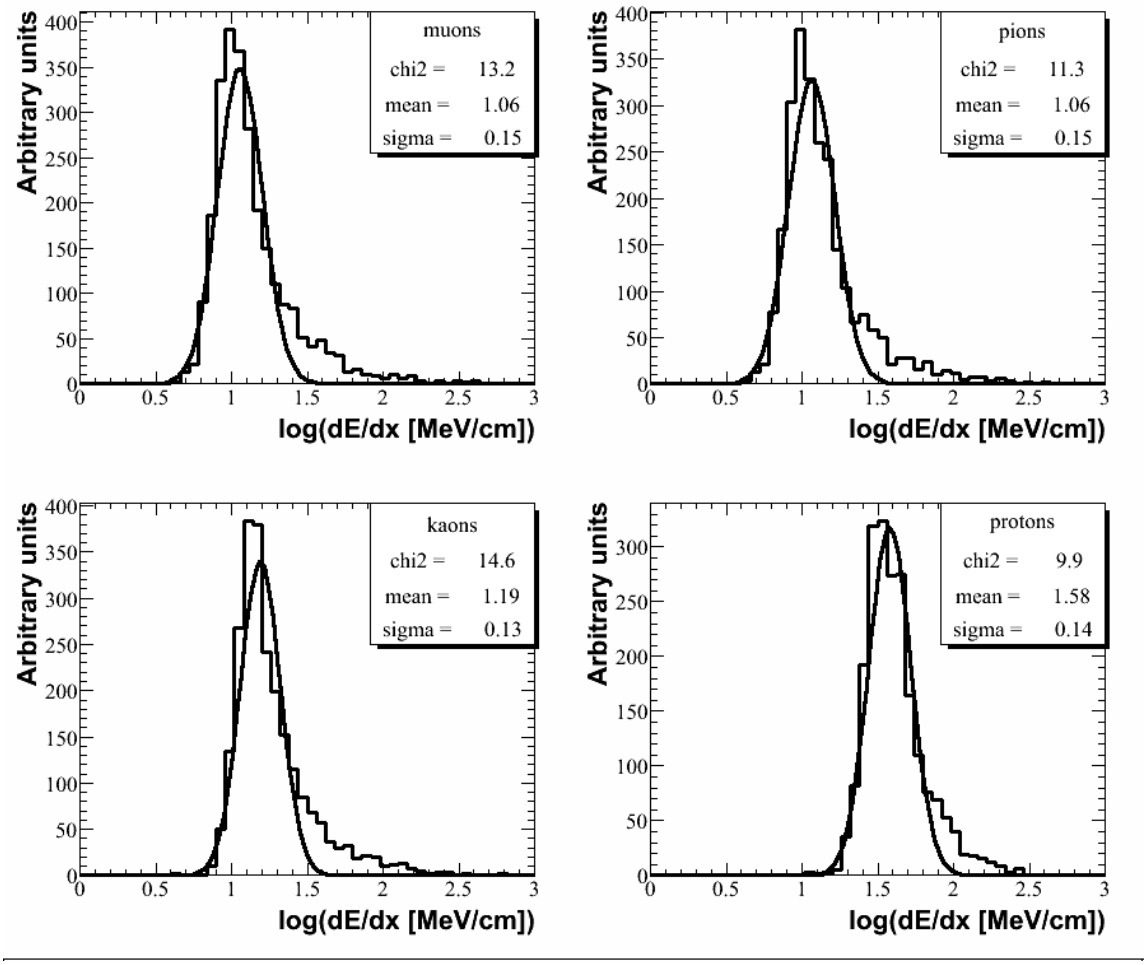


Figure 6: Logarithm of the median dE/dx for different particle species, for momentum between 900 and 1100 MeV/c.

to distributions with positive skewness. Its main drawback is the introduction of a bias depending on the arbitrary parameter k.

The separation power for the generalized mean of grade $k = -2$ as a function of the reconstructed momentum is shown in Fig. 8(left). Fig. 8(right) shows the same estimator in a slice of momentum between 900 and 1100 MeV/c.

The resolution for muons is 6.6% for the generalized mean of grade $k = -2$, while the distance between the kaon and pion distributions is 1.3 standard deviations, and the proton/pion distance is 4.8 standard deviations.

In Table 1, other generalized means of different grades are also compared.

3.4 Landau fit

Ideally, an unbinned fit to a Landau hypothesis would give the best estimation of the MPV of the distribution, although at the price of a much longer computation time (due to the iterative minimization of the errors). In practice, it has been found that this estimation doesn't perform significantly better than the best other discriminators under consideration, as shown in Table 1.

The unbinned fit is performed using ROOT [16], and the MPV and the width are extracted simultaneously. The separation power of the estimated MPV as a function of the reconstructed momentum is shown in Fig. 9(left). Fig. 9(right) shows the same estimator in a slice of momentum between 900 and 1100 MeV/c.

Fig. 10 shows the distribution of the natural logarithm for the estimated MPV between 900 and 1100 MeV/c. The resolution for muons is 6.3%, the distance between the kaon and pion distributions is 1.6 standard deviations, and the proton/pion distance is 5.4 standard deviations.

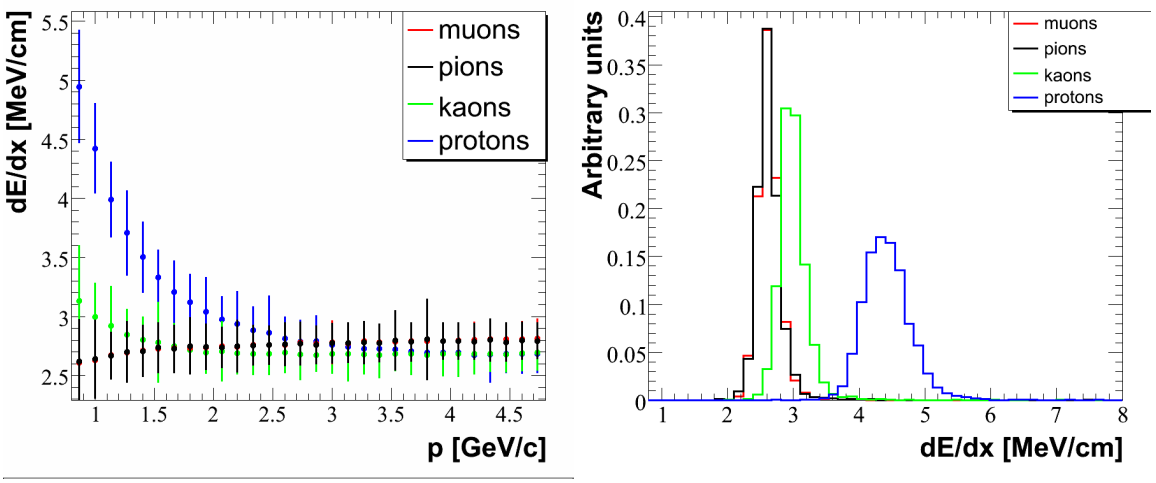


Figure 7: Truncated mean (40% truncation) of dE/dx as a function of the track momentum for different particle species (top left) and in the momentum range between 900 and 1100 MeV/c (right). Proton/pion separation as a function of momentum (bottom left). The error bars in top left are the RMS of the distribution.

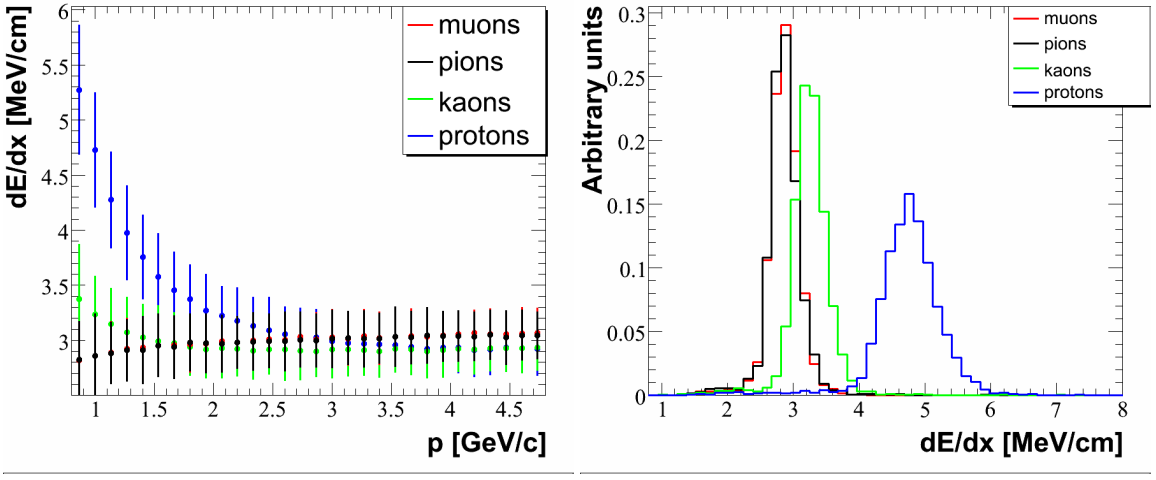


Figure 8: Generalized mean of grade $k = -2$ of dE/dx as a function of the track momentum for different particle species (top left) and in the momentum range between 900 and 1100 MeV/c (right). Proton/pion separation as a function of momentum (bottom left).

3.5 Comparison in single-particle events

These different estimators have been compared, with the goal of finding an "optimal" dE/dx for particle identification.

A benchmark is the resolution obtained for muons in the momentum range between 900 and 1100 MeV/c. The other two criteria considered in the table are the maximum separation between hadron species (K/π and p/π) in the log(dE/dx) variable (which presents the good property of approximate gaussianity in the surrounding of the peak of these estimators). No attention has been given to the bias of the estimator (which is a function of the truncation fraction for the truncated mean and of the grade for the generalized mean), i.e., to how much the MPV of the resulting distribution differs from the expectation for a given particle. This is not a concern since the correct peak position can be restored with a particle calibration, as devised in Sec. 6.3.

The results are shown in Table 1. Other criteria are described in Sec. 5 in the specific context of proton identification. In general, the "generalized means" seem to perform very well under the point of view of resolution and hadron separation. In the following, the generalized mean of grade $k = -2$ is taken as representative of the family.

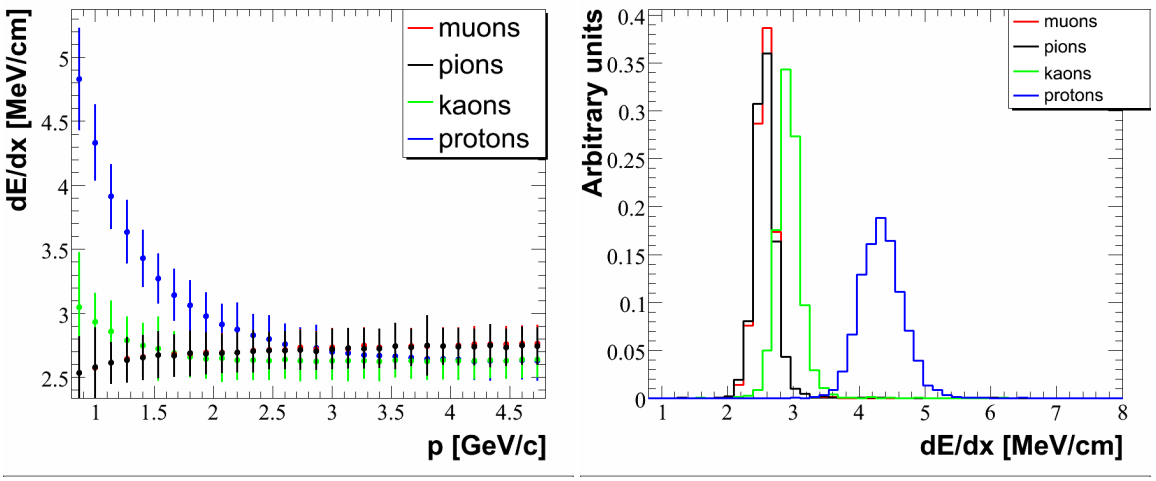


Figure 9: Most probable values for dE/dx, as extracted from a Landau fit, as a function of the track momentum for different particle species (top left) and in the momentum range between 900 and 1100 MeV/c (right). Proton/pion separation as a function of momentum (bottom left). The error bars in top left are the RMS of the distribution.

Table 1: Performances of some dE/dx estimators for particles of momentum between 900 and 1100 MeV/c.

Estimator	muon σ (%)	K/π separation	p/π separation
Median	14.5	0.61	2.45
Truncated mean (20 %)	6.8	1.28	4.76
Truncated mean (40 %)	6.3	1.50	5.29
Harmonic mean	7.4	1.11	4.26
Generalized mean ($k = -1/2$)	8.0	1.03	3.97
Generalized mean ($k = -1/3$)	8.1	0.98	3.97
Generalized mean ($k = -2$)	6.6	1.29	4.76
Generalized mean ($k = -4$)	6.0	1.50	5.38
Generalized mean ($k = -6$)	6.0	1.58	5.58
Landau fit	6.3	1.57	5.44

4 Combined compatibility method

This section describes a complementary strategy for particle identification, making the maximal use of the information content of the set of dE/dx samplings associated with a track.

For each hit, the compatibility with a given particle hypothesis at a given momentum is defined as the probability that dE/dx values equal or less than the one observed can be the outcome of the measurement for that particle at that momentum:

$$l^{hit}_{hypo}(p, \Delta x_{hit}) \equiv P\left(\frac{dE}{dx} \leq \frac{dE}{dx}^{hit} \mid hypothesis,\ p,\ \Delta x_{hit}\right), \tag{2}$$

where $\Delta x_{hit} = t_{sensor} \cdot \sec\theta$, since the pathlength affects the width of the distributions. The combined compatibility for a track is given by the geometric mean:

$$L_{hypo}(p) = \sqrt[N]{\prod_{hit=1}^{N} l^{hit}_{hypo}(p, \Delta x_{hit})} \tag{3}$$

In order to estimate $l^{hit}_{hypo}(p)$, the only assumption is that the dE/dx measurements are Landau-distributed. The particle-gun events are used to plot the $(dE/dx)^{hit}$ distributions in 16 bins of momentum between 900 and 2500 MeV and 4 bins of Δx_{hit}. The MPVs and widths extracted from the Landau fit to these distributions are plotted as a function of momentum (see Fig. 11 for the example of $510\mu m < \Delta x_{hit} < 560\mu m$) and fitted by the functional form $f(p) = q_0\,(1 - q_1 e^{-q_2\ p})$.

For each hit, $l^{hit}_{\pi}(p, \Delta x_{hit})$, $l^{hit}_{K}(p, \Delta x_{hit})$ and $l^{hit}_{p}(p, \Delta x_{hit})$ are computed by taking the expected MPVs and widths corresponding to the measured Δx_{hit} and p from the fits of Fig. 11, and integrating the normalized Landau

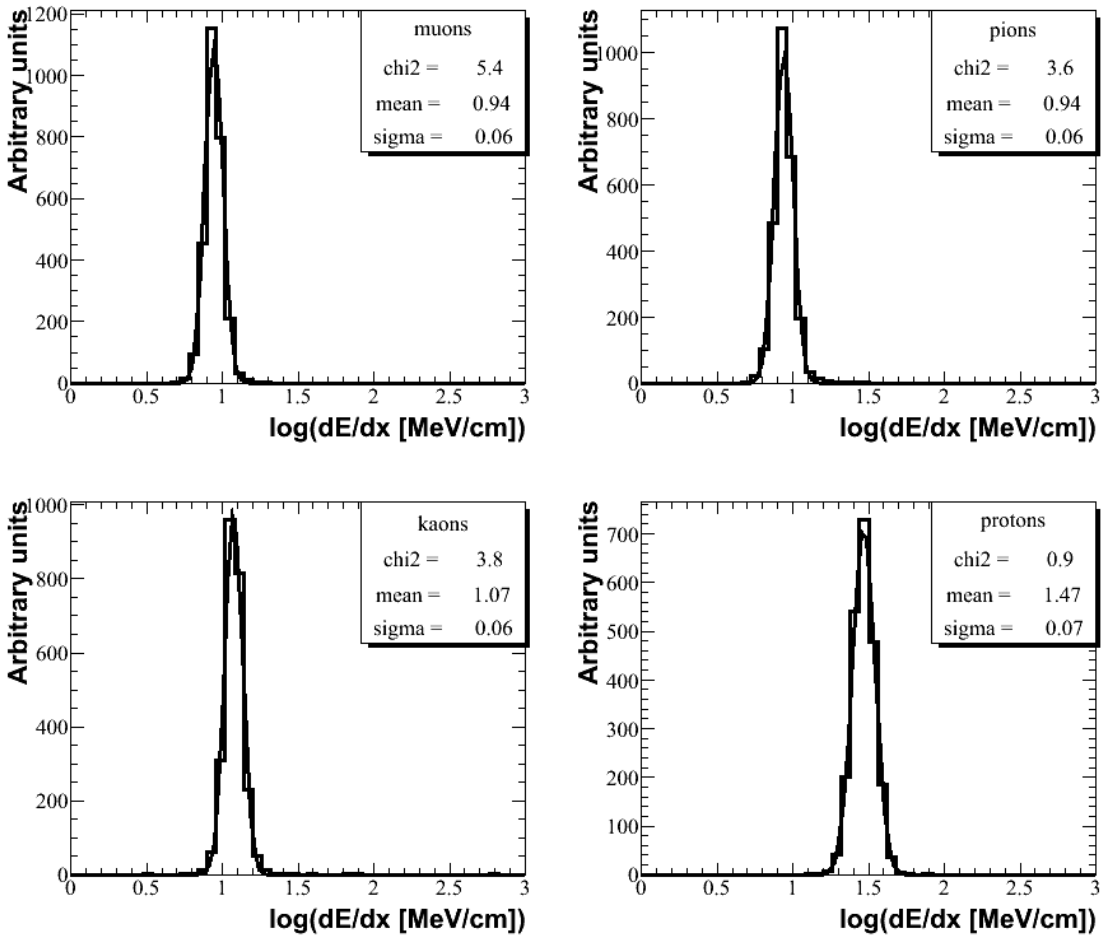


Figure 10: Logarithm of the most probable values (from a Landau fit) for dE/dx for different particle species, for momentum between 900 and 1100 MeV/c.

distribution from $-\infty$ to the measured value $(dE/dx)^{hit}$:

$$l^{hit}_{hypo}(p, \Delta x_{hit}) = \int_{-\infty}^{\frac{dE}{dx}^{hit}} Landau\left(y|MPV(p, \Delta x_{hit}), width(p, \Delta x_{hit})\right)\, dy\,. \tag{4}$$

For the purpose of particle identification, one can separately use the resulting $L_\pi(p)$, $L_K(p)$ and $L_p(p)$ distributions (or, equivalently, $1 - L_{hypo}$). The specific example of proton identification, by incompatibility with pion and kaon hypotheses, is illustrated in the following section.

5 Performance of proton selection

Two samples are used as benchmarks.

- A sample of minimum bias events, corresponding to a total cross section of 75 mb: hard scattering (50 mb), single diffractive (15 mb) and double diffractive (10 mb). The hard scattering part is the one contributing most tracks.
- A sample of QCD events with $80 < \hat{p}_T < 120$ GeV. Its estimated cross section is 3 μb.

A plot of dE/dx versus momentum is shown in Fig. 12 for minimum bias (left) and QCD (right) events, where the generalized mean of grade $k = -2$ is chosen as estimator of the most probable value. The incompatibilities with pion, kaon and proton hypotheses ($1 - L_\pi$, $1 - L_K$ and $1 - L_p$) are shown in Fig. 13 and Fig. 14, for minimum bias and QCD events respectively. The last variable ($1 - L_p$) has clearly no discriminating power when the aim

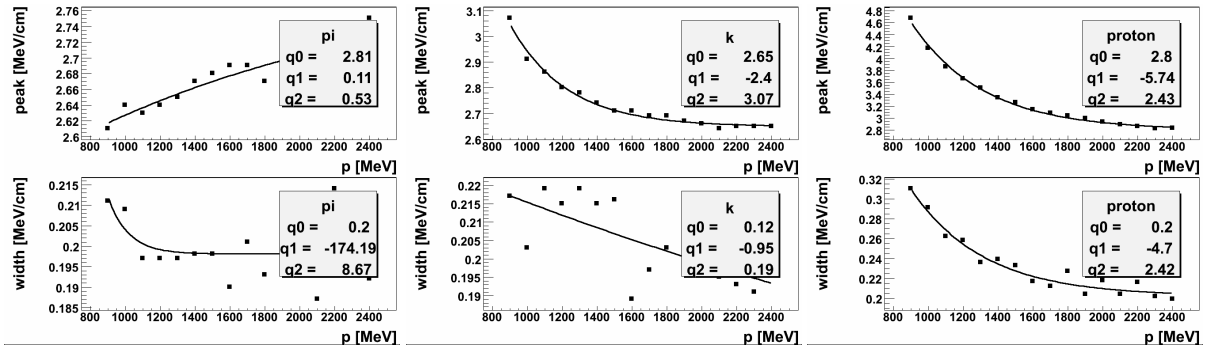


Figure 11: Most probable value (top row) and width over MPV (bottom row) from the Landau fits to $(dE/dx)^{hit}$ distributions, as a function of momentum, for pions (left), kaons (center) and protons (right), in the pathlength range $510\mu m < \Delta x_{hit} < 560\mu m$.

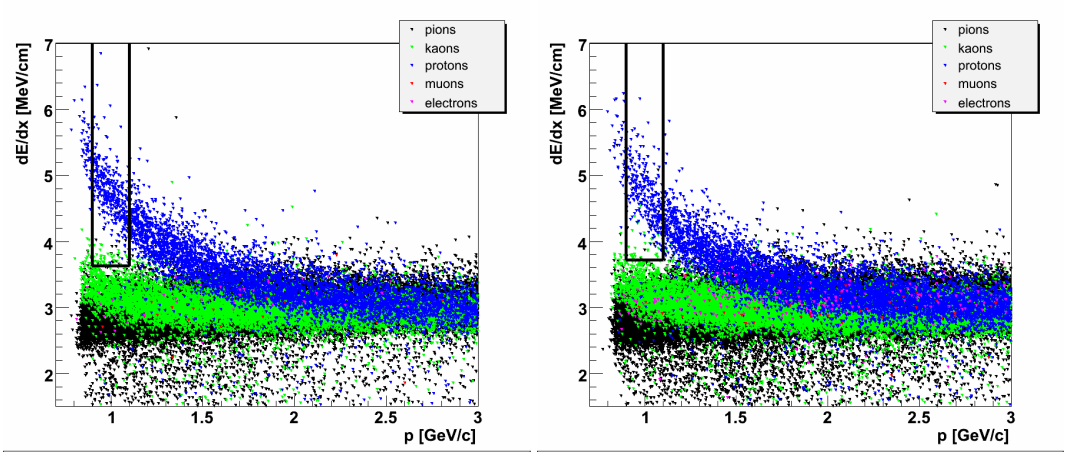


Figure 12: Plot of dE/dx versus momentum, where dE/dx is estimated with the generalized mean of grade $k = -2$, for minimum bias events (left) and QCD events with $80 < \hat{p}_T < 120$ GeV (right) at LHC. The rectangle indicates a selection region with $0.9 < p < 1.1$ GeV/c and proton purity higher than 80%.

is to select proton themselves, but it is shown here in order to provide an example of its behaviour for Standard Model particles, to be compared with an exotic signal in Fig. 21.

Table 2: Efficiencies of some proton discriminators for minimum bias (MB) and QCD events, in the momentum range between 900 and 1100 MeV/c, for a purity of 80%.

Discriminator	Efficiency for MB	Efficiency for QCD
dE/dx, generalized mean ($k = -2$)	97.2%	95.9%
dE/dx, generalized mean ($k = -4$)	95.8%	94.8%
dE/dx, generalized mean ($k = -6$)	94.6%	93.7%
Incompatibility with π	98.9%	98.2%
Incompatibility with K	99.7%	98.9%

As "goodness criterion" the following has been chosen: the efficiency of proton selection achieved, in the momentum range $0.9 < p < 1.1$ GeV/c, for a purity of 80%. The result is illustrated in Table 2, from which it is clear that the best discriminator for proton selection is the incompatibility with the kaon hypotesis.

6 Calibration strategy

When we talk about calibration, we are actually interested in some of the following:

- inter-calibration: all the detector sub-units are equalized to the same gain; otherwise, random deviations from one sub-unit to another result in a widening of the dE/dx distributions;

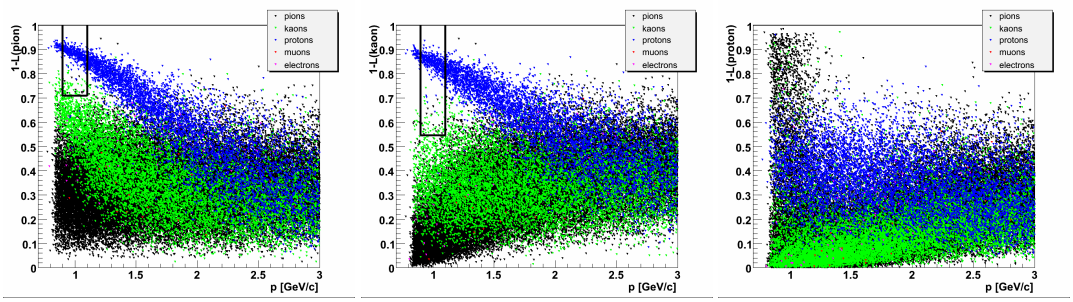


Figure 13: Plot of $1 - L_\pi$ (left), $1 - L_K$ (center) and $1 - L_p$ (right) versus momentum, for minimum bias events at LHC. The rectangles in the first two plots indicate a selection region with $0.9 < p < 1.1$ GeV/c and proton purity higher than 80%.

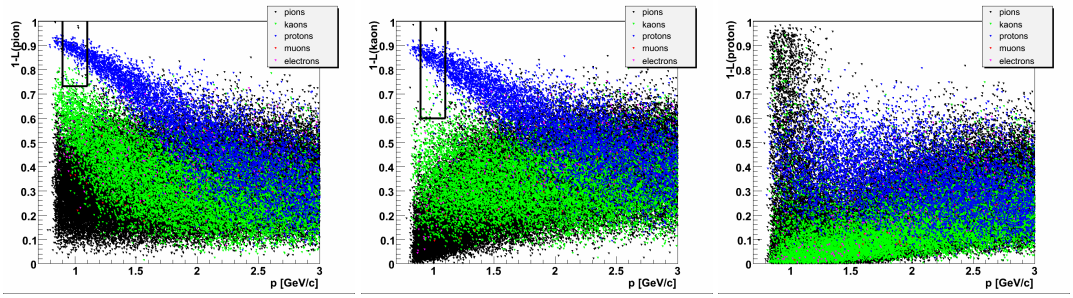


Figure 14: Plot of $1 - L_\pi$ (left), $1 - L_K$ (center) and $1 - L_p$ (right) versus momentum, for QCD events with $80 < \hat{p}_T < 120$ GeV at LHC. The rectangles in the first two plots indicate a selection region with $0.9 < p < 1.1$ GeV/c and proton purity higher than 80%.

- absolute calibration: this is done with pure samples of known particles (e.g., cosmic muons), whose specific energy loss as a function of momentum is known;
- N-points calibration curve: by selecting N samples, each with high purity for one specific particle, it is possible to validate directly the particle identification procedure with the first data from CMS.

It has to be noted that the absolute scale is not directly relevant for particle identification, where we deal only with differences: more important for the scope of this paper are the linearity of the response, the uniformity of the signal measurement across the sub-detectors, and the stability in time. Absolute calibration on the other hand is very important for keeping under control the performance of the system: regular surveys will permit to estimate and correct for any degradation of the gain which might occur with time.

6.1 Calibration with tickmarks

The principle of the method is described in Ref. [17]. During an optical link setup run (the 'gain scan'), the APV outputs synchronization pulses (or 'ticks'), identical in height to the digital header. What is relevant for the present study is the fact that the height of this digital header is fixed to a nominal value (800 mV between logic '0' and logic '1'). At the nominal APV gain, this is equivalent to an 8 MIP signal. This gain is assumed to be constant, but a 5% variation from chip to chip has been observed [18].

The 'tickmark calibration' permits the equalization of all the digital heights to a common value. In Fig. 15, where the method is compared with the one devised in the next section, values are rescaled according to the outcome of a dedicated calibration run at the MTCC [12].

A spread of 10% is obtained as standard deviation of a gaussian fit.

6.2 Calibration with cosmics

While the 'tickmark calibration' has the advantage of the simplicity, and of being available from the very first operation, it should be recalled that it doesn't take into account several factors affecting the overall amplification of the signals produced by particles traversing the detector. Non-uniformities in the charge collection, in the signal

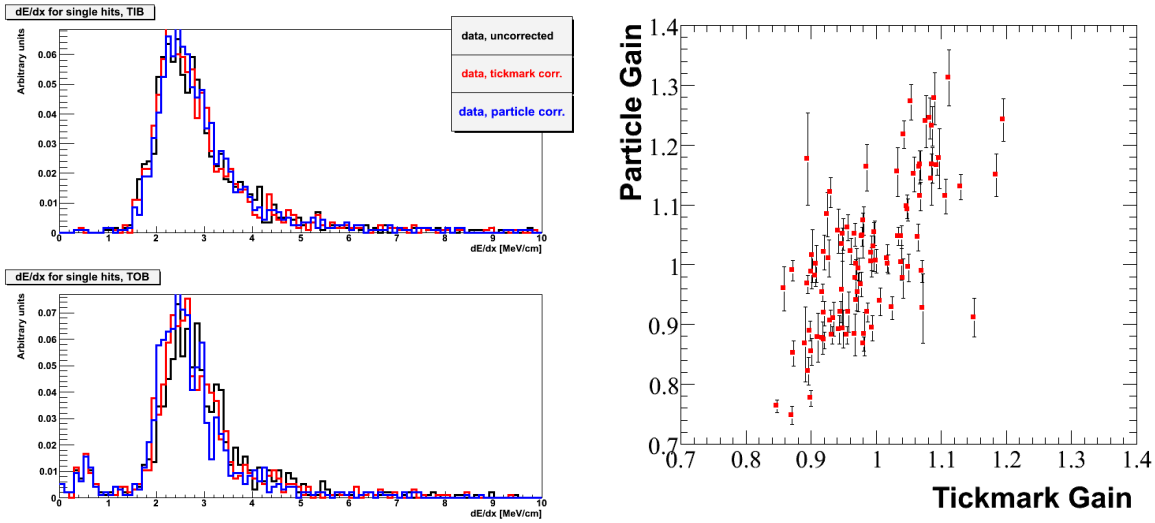


Figure 15: Left: Effect of tickmark and particle corrections on MTCC data. Right: Correlation between the tickmark and the particle calibrations, in MTCC data, from Ref. [12]. Only the APVs having at least 50 charge entries have been considered for this plot. The error bar corresponds to the statistical uncertainty on the peak position of the signal distribution in the particle case.

generation and in several steps of the readout chain affect the amplification and the linearity of the response to the primary charge. For this reason, when enough statistics are available it is convenient to use a 'particle calibration' from data.

Here, a particle calibration based on the dE/dx of the cosmic tracks from the MTCC is described and compared with the tickmark calibration from the same data (further details are presented in Ref. [12]).

For each APV, the charge of the clusters associated with tracks (in ADC counts) and the angle of incidence between these tracks and the surface of the module are used to plot a dE/dx distribution. For APVs crossed by at least 50 cosmic tracks during the MTCC data taking, the Most Probable Value (MPV) of the Landau curve fitted on the distribution is used for calibration. The correlation between the 'tickmark calibration' and this 'particle calibration' is apparent from Fig. 15(right).

The spread obtained with this method is $\approx 10\%$, compatible with what obtained with the tickmark method. Because the APVs, at the MTCC, were operated in peak mode, not much can be concluded on the non-uniformities introduced by bad synchronization when using deconvolution mode (as in the normal LHC data taking conditions), since the signal will be much shorter in time.

For what concerns the absolute calibration, the number of electrons corresponding to one ADC count has been extracted from the MTCC data, assuming 22.500 electron-hole pairs created in $300\ \mu m$ of silicon. The measured peak value of the signal (corrected to normal incidence on modules) is 90 ADC counts for cosmic tracks in TIB modules, corresponding to a gain factor e/ADC = 250 [12]. This value has been used for all the results in this paper.

The same calibration method can be used with the final CMS configuration in the underground cavern. A Monte Carlo estimation taking into account the angular distribution of cosmics and the rock shielding gives a rate of ≈ 10 Hz for muons of $E > 5\ \mathrm{GeV}/c$ in the cavern, across all the Tracker volume, to be compared with roughly two orders of magnitude more when the detector is at ground level, as in the MTCC [19]. On the other hand, the tracks will have much higher geometrical acceptance than in the MTCC setup, where only 1% of the final setup was instrumented.

A straightforward variant of this method makes use of the beam-halo muons[3)]. The measurements with cosmic and beam-halo muons will be complementary, since the former will traverse more often the Barrel modules while the latter will illuminate more the Endcaps.

[3)] This analysis will benefit from the acquisition of cosmic and beam-halo muons in specific runs motivated by the needs of the alignment procedures.

6.3 Calibration with V^0's

The ideal way to calibrate the dE/dx is "in situ", selecting subsamples of the data recorded in LHC collisions, in the same runs as the events that one is going to analyze. This way it is possible to disentangle time-dependent effects on the gains.

One possibility is to monitor pions, which are by far the most abundant particles produced in hadron collisions, at their ionization minimum (650 MeV/c), in order to set the absolute scale. This is below the standard seeding cut (see Sec. 2.1) but well within the range of specific tracking algorithms optimized for low-$p_T$particles [2], which can be applied for the purposes of calibration.

Another important issue is to check, and correct for, the linearity of the response with momentum, and to this end it is desirable to have a sample where it is possible to verify the dE/dx versus p curve in a broad momentum range. One possibility is to exploit the protons in the very abundant minimum bias and jet events; but, as shown in Fig. 12-14, this gives a very pure sample only for momenta below ≈ 1.5 GeV/c. Another possibility, discussed in this section, is offered by the exclusive decays of known resonances, where the final states have fixed composition.

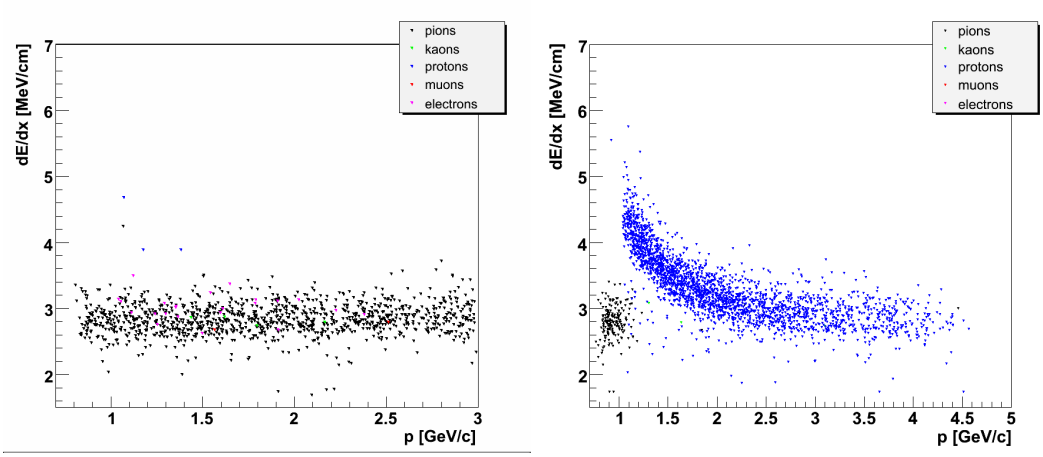


Figure 16: Plot of dE/dx versus momentum, where dE/dx is estimated with the generalized mean of grade $k = -2$, for tracks coming from K^0_S (left) and from Λ and $\bar{\Lambda}$ (right) decays at LHC.

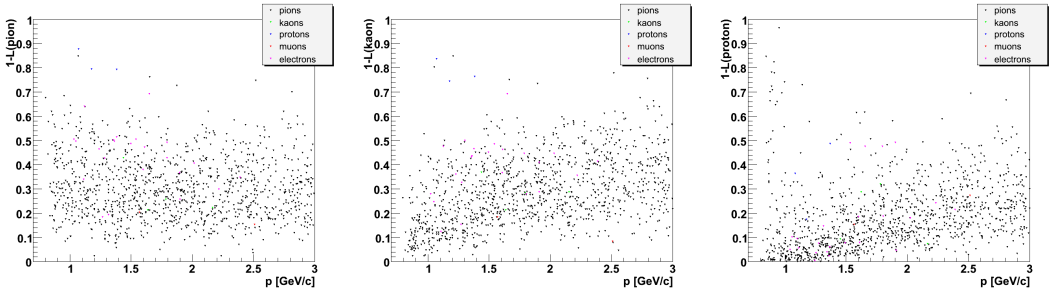


Figure 17: Plot of $1 - L_\pi$ (left), $1 - L_K$ (center) and $1 - L_p$ (right) versus momentum, for tracks coming from K^0_S decays at LHC.

A very pure sample of pion tracks can be obtained by selecting tracks compatible with the K^0 hypothesis. The $K^0_S \to \pi^+\pi^-$ has a branching ratio of 68% (the rest is mostly $\pi^0\pi^0$) and the lifetime of the K^0_S is short enough to always decay inside the Tracker volume. The result is shown in Fig. 16(left) and Fig. 17 for a sample of K^0_S events generated with a particle gun (p_T= 5 GeV/c, $\eta = 0$), where the tracks are reconstructed with a specific pixel-less seeding and with mass and secondary vertex constraints, tuned for K^0 reconstruction [20, 21]. The contamination from other hadrons is expected to be negligible.

While this "pion-enriched" sample would be sufficient to set the absolute calibration and equalize the APVs, it is also important to check the performance of the particle identification with samples containing other particles. A "proton-enriched" sample would permit to test the linearity of the response, by checking if their dE/dx versus p distribution reproduces the Bethe-Bloch expectation. Such an opportunity is offered by the decay $\Lambda \to p\pi^-$, $\bar{\Lambda} \to \bar{p}\pi^+$ (64% of the Λ decays, the rest being mostly into $n\pi^0$).

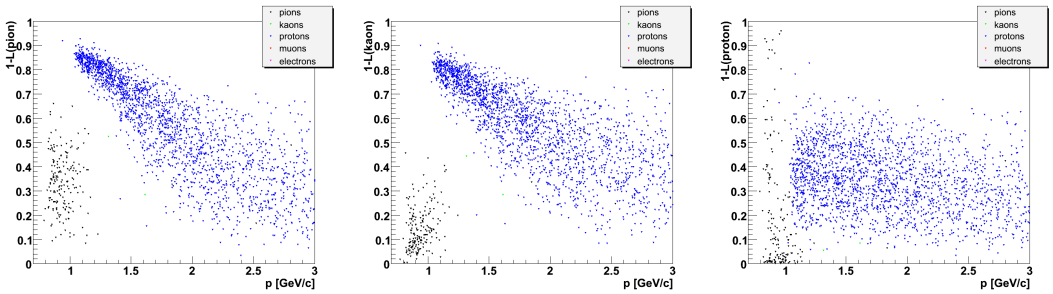


Figure 18: Plot of $1 - L_\pi$ (left), $1 - L_K$ (center) and $1 - L_p$ (right) versus momentum, for tracks coming from Λ and $\bar{\Lambda}$ decays at LHC.

Due to the kinematics of the decay, the proton is predominantly found going off in the direction of the spin of the Λ. A high-purity sample of protons can be selected by taking the highest-momentum track of the V^0 pair, after a selection similar to the K^0_S case or after a more sophisticated analysis based on the Podolanski-Armenteros variables [22, 23]. The dE/dx versus momentum is shown in Fig. 16(right) and the $1 - L_i$ variables in Fig. 18 for a sample of tracks from Λ and $\bar{\Lambda}$ decays generated with a particle gun (p_T flat between 1 and 5 GeV/c, $\eta = 0$), where the tracks are reconstructed with the same procedure used for K^0_S's.

7 Systematics

7.1 Gain linearity and stability

The linearity of the optical link gain, and its stability against several nuisance sources have been extensively studied in Ref. [17]. Here the results most relevant for the dE/dx measurement are summarized.

It is found that the maximum charge that can be transferred without any loss of information (i.e., that can be accommodated into 8 bits) is 80.000 electrons per strip, corresponding to 3.2 MIPs in the 320 μm modules and 2 MIPs in the 500 μm modules; above this limit, saturation occurs [24]. This effect is already included in the simulation, but has been found to have negligible effects on standard particles, protons included. As seen from Fig. 12, the signal is very rarely larger than roughly twice the muon, in the momentum range accessible to the standard tracking. Moreover, the charge of a cluster is typically shared between more than one single strip.From the point of view of the search for exotic particles, saturation is not a concern: at high momentum, a charged particle leaving a trail of saturated hits would be an unmistakable signal of non-standard physics. The only problem would be the impossibility to extract a mass from the Bethe-Bloch; in this case the mass measurement will have to rely completely on the time of flight method, exploiting the muon detectors.

For what concerns the stability, by far the main effect on the gains comes from temperature. It is estimated that the average gain increase with temperature is $-0.64\%\ ^oC^{-1}$ [17].

The cooling system of the Tracker provides great stability of the silicon temperature at the operating conditions around -10^oC [25]. Local temperatures may vary a lot with respect to this value but, apart from the first ≈ 5 minutes after being switched on, their values are expected to be linked to those of the cooling pipes and to be as stable as them.

Assuming, conservatively, a random temperature variation of $\pm 3^oC$ for each APV's optohybrid during data-taking, this leads to a $\pm 2\%$ uncertainty on the charge measured in each hit. The impact of this uncertainty has been estimated by smearing the dE/dx value for each hit of the track before estimating the most probable value.

The result of this smearing is negligible on the considered estimators, as can be seen in Fig. 19 and Table 1. A more pessimistic scenario of $\pm 6.4\%$ uncertainty, corresponding to $\pm 10^oC$ uncertainty in temperature, is also shown. Although appreciable, the differences are still small, with the generalized means of power larger than 2 showing much less stability than the powers up to 2.

7.2 Synchronization

The detectors units in the Tracker are synchronized to better than 1 ns [26]; such a delay corresponds to a loss of signal of 0.34% with respect to the peak of the pulse shape curve, approximating the pulse shape curve in deconvolution mode with a gaussian having 12.06 ns width [9], as in the current simulation. This has been approximated

Table 3: Performances of some dE/dx estimators for particles of momentum between 900 and 1100 MeV/c, in different smearing scenarios.

	Ideal detector	$\pm 2\%$ smearing	$\pm 6.4\%$ smearing
muon σ (%), Truncated mean (40 %)	6.27	6.35	6.77
muon σ (%), Generalized mean ($k = -2$)	6.58	6.68	6.95
muon σ (%), Generalized mean ($k = -4$)	6.03	6.04	6.45
muon σ (%), Generalized mean ($k = -6$)	5.98	6.06	6.68
muon σ (%), Landau fit	6.30	6.37	6.90
p/π separation, Truncated mean (40 %)	5.29	5.26	5.01
p/π separation, Generalized mean ($k = -2$)	4.76	4.74	4.54
p/π separation, Generalized mean ($k = -4$)	5.38	5.38	5.02
p/π separation, Generalized mean ($k = -6$)	5.58	5.48	5.09
p/π separation, Landau fit	5.44	5.42	5.14

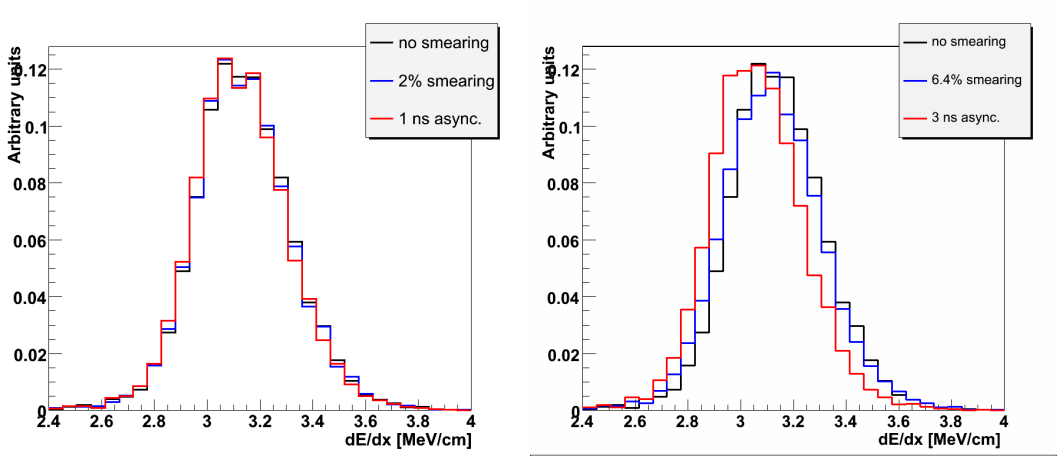


Figure 19: Generalized mean of grade $k = -2$ of dE/dx with different smearing scenarios, for simulated single muons with flat p_T between 3 and 5 GeV/c.

conservatively to $\approx 0.4\%$ in the following. To estimate the impact of this effect on dE/dx, a correction factor of $-|Gauss(0, 0.004)|$ has been randomly assigned to each hit measurement.

This results in a bias on the dE/dx determination, as shown in Fig. 19 for single muons with p_T between 3 and 5 GeV/c, which has to be corrected with the calibration procedures proposed in Sec. 6. This bias amounts to $\approx -0.3\%$ for all the MPV estimators considered in this note, independently of momentum.

The impact on the resolution and on the separation performances can be seen in Table 4. A more pessimistic scenario of 3% width of the gaussian smearing, corresponding to ± 3 ns uncertainty on synchronization, is also shown. The bias in this scenario ranges between -2.4% and -2.6% for all the considered estimators.

7.3 Capacitive coupling and cluster thresholds

The cluster-finding algorithm in the SST [10] takes three thresholds, expressed in units of noise: the intermediate threshold corresponds to the minimum S/N for a cluster seed, the lowest corresponds to the minimum S/N required for a strip adjacent to a candidate cluster in order to be merged to it, and the highest corresponds to the minimum S/N for accepting the cluster. Two threshold settings are being considered for CMS SST operation: "3,2,5" and "4,3,5", where the first digit refers to the seed threshold, the second to the merging threshold, and the third to the overall charge of the cluster.

The AC coupling of the strips readout influences the charge seen on neighboring strips adjacent to the hit one, and different cluster-finding thresholds are differently sensitive to its effect. The inter-strip capacitive coupling can be extracted from data; the MTCC data [11] favour a value of 6%, while all the results shown in this note assume a value of 12% in simulation.

The proton/pion separation power in the momentum range between 900 and 1100 MeV/c at $\eta = 0$, for the

Table 4: Performances of some dE/dx estimators for particles of momentum between 900 and 1100 MeV/c, in different synchronization scenarios.

	Ideal detector	± 1 ns synchrony	± 3 ns synchrony
muon σ (%), Truncated mean (40 %)	6.27	6.32	6.54
muon σ (%), Generalized mean ($k = -2$)	6.58	6.60	6.92
muon σ (%), Generalized mean ($k = -4$)	6.03	6.06	6.26
muon σ (%), Generalized mean ($k = -6$)	5.98	5.99	6.18
muon σ (%), Landau fit	6.30	6.30	6.52
p/π separation, Truncated mean (40 %)	5.29	5.29	4.86
p/π separation, Generalized mean ($k = -2$)	4.76	4.74	4.45
p/π separation, Generalized mean ($k = -4$)	5.38	5.38	5.08
p/π separation, Generalized mean ($k = -6$)	5.58	5.57	5.14
p/π separation, Landau fit	5.44	5.50	5.48

generalized mean of grade $k = -4$, is compared in Table 5 for these two different scenarios in the two cluster-finding settings. The result seems to be quite stable in general, and in particular with the "435" setting, as expected, there is a negligible dependence on the capacitive coupling.

Table 5: Separation power between protons and pions in the momentum range between 900 and 1100 MeV/c at $\eta = 0$, for the generalized mean of grade $k = -4$, in different capacitive coupling (c.c.) and cluster-finding threshold scenarios.

c.c.	"3,2,5"	"4,3,5"
12%	5.49	5.44
6%	5.43	5.42

8 Conclusions and Outlook

Some generic procedures for particle identification based on the specific energy loss in the SST have been discussed in this paper, divided in two classes:

- estimators of the most probable value for the dE/dx of the track, among which the most powerful are the generalized means, due to their property of suppressing the outliers;
- combined compatibilities of the hits of a track with specific mass hypotheses.

The effectiveness of both methods have been demonstrated for proton identification at low momentum and are expected to be exploited by the Particle Flow Algorithm, presently under development, as well as in the search of exotic charged particles using the CMS Detector.

Fig. 20 and Fig. 21 shows how R-hadron [1] events (simulated in GEANT with the extension described in Ref. [27]) would look like with the application of some of the methods developed in this note, and the requirement of at least 7 hits in the SST. In the bottom-left corner one finds the Standard Model background (mostly pions, kaons and protons) while a few outstanding events at high momentum and high energy deposition would be the signature for new physics.

A serious concern, when dealing with a dE/dx measurement, comes from the reliability of the signal height information, on which the measurement is based. It has been shown in this paper that, taking into account realistic values of the uncertainty on the gains and on the timing, the performance of the methods is only slightly affected.

A calibration strategy in three steps (tickmarks method, cosmic and/or beam-halo muons, V^0s in LHC collisions) will permit to exploit reliably the dE/dx information with the very first data.

9 Acknowledgements

Many thanks, for several useful discussions and advices, to Ferenc Siklér, Fabrizio Palla, Gigi Rolandi, Ian Tomalin, Giacomo Bruno, Geoff Hall and Andrea Rizzi. Andrea also produced the R-hadrons events that I used for

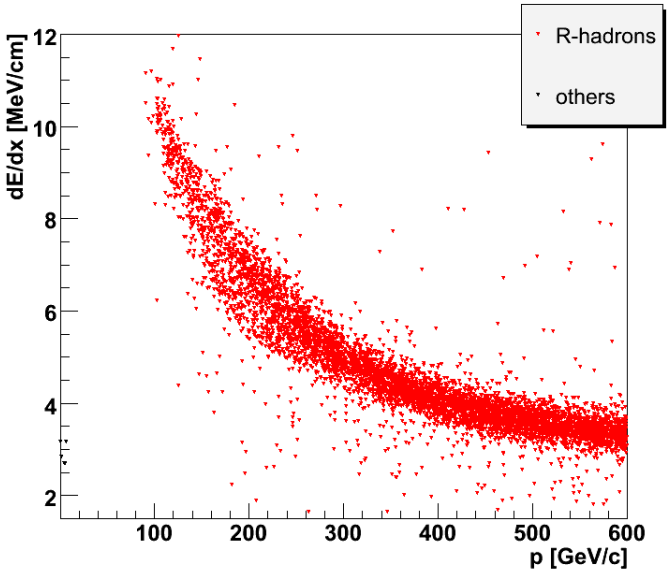


Figure 20: Plot of dE/dx versus momentum, where dE/dx is estimated with the generalized mean of grade $k = -2$, for events with R-hadrons of 300 GeV/c^2 mass, simulated in GEANT with the extension described in Ref. [27].

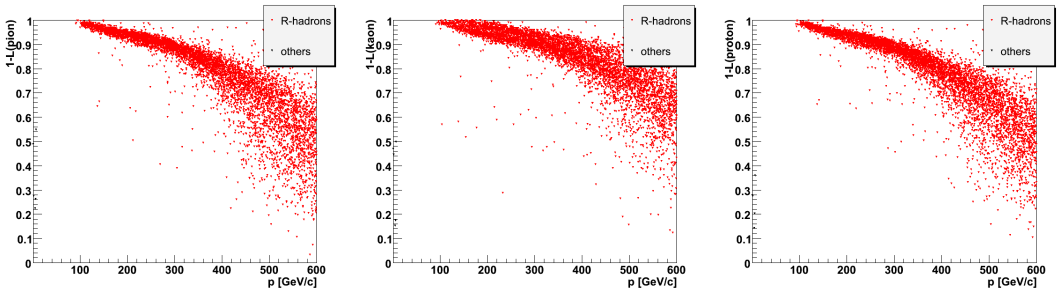


Figure 21: Plot of $1 - L_\pi$ (left), $1 - L_K$ (center) and $1 - L_p$ (right) versus momentum for events with R-hadrons of 300 GeV/c^2 mass, simulated in GEANT with the extension described in Ref. [27].

Fig. 20 and Fig. 21. Gero Flucke pointed me to the method used in the H1 experiment, which motivated my exploration of the generalized mean as estimator of dE/dx for tracks. Massimiliano Chiorboli and Livio Fanò are to be acknowledged for the simulation of cosmic muons with the MTCC setup, tuning the response of the Tracker and simulating the Muon System trigger. Massimiliano also produced the V^0 samples, including the reconstruction and selection of the V^0's themselves. Giacomo Bruno and Dorian Kcira are to be acknowledged for the application of the method of 'particle calibration' to the MTCC setup. Boris Mangano, Frédéric Ronga, Chiara Genta, Giuseppe Cerati were very helpful in several instances and in particular for the issue of the access to the trajectory information in CMSSW. Duccio Abbaneo suggested to check the influence of the cluster thresholds on the particle identification.

This work is partially supported by the Belgian Federal Science Policy (IAP 6/11).

References

[1] M. Fairbairn, A.C. Kraan, D.A. Milstead, T. Sjöstrand, P. Skands, T. Sloan, *Stable massive particles at Colliders*, hep-ph/0611040.

[2] F. Siklér, K. Krajczár *Measurement of charged hadron spectra in proton-proton collisions at* $\sqrt{s} = 14$ *TeV*, CMS AN-2007/021.

[3] CMS collaboration, *The CMS experiment at the CERN LHC*, paper in preparation.

[4] CMS collaboration, *Tracker Technical Design Report*, CERN/LHCC 98-6; and CMS collaboration, *Addendum to the CMS Tracker TDR*, CERN/LHCC 2000-016.

[5] CMS collaboration, *Physics Technical Design Report, Vol.I - Detector Performance and Software*, CERN/LHCC 2006-001, Sec.6.4.

W. Adam, B. Mangano, Th. Speer, T. Todorov, *Track Reconstruction in the CMS tracker*, CMS NOTE 2006/041.

[6] M.J. French et al., *Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker*, Nucl. Instrum. Meth., **A466**, (2001) 359.

[7] S. Gadomski et al., *The deconvolution method of fast pulse shaping at hadron colliders*, Nucl. Instrum. Meth., **A320**, (1992) 217.

[8] C. Delaere, L. Mirabito, *Timing of the CMS tracker. Study of module properties*, CMS NOTE 2007/027.

[9] M. Raymond et al., *The CMS Tracker APV25 0.25 μm CMOS Readout Chip*, Proceedings of the 6th Workshop on Electronics for LHC Experiments, CERN/LHCC 2000-041.

[10] CMS collaboration, *Physics Technical Design Report, Vol.I - Detector Performance and Software*, CERN/LHCC 2006-001, Sec.6.4.1.

[11] CMS collaboration, *The CMS Magnet Test and Cosmic Challenge (MTCC Phase I and II)*, CMS NOTE 2007/005.

[12] D. Abbaneo et al., *Tracker Operation and Performance at the Magnet Test and Cosmic Challenge*, CMS NOTE 2007/029.

[13] J. Allison et al., *Geant4 developments and applications*, IEEE Transactions on Nuclear Science 53 No. 1 (2006) 270-278.

[14] R.K. Bock and A. Vasilescu, *The Particle Detector BriefBook*, Springer (1998).

[15] H1 collaboration, *The tracking, calorimeter and muon detectors of the H1 experiment at HERA*, Nucl. Instrum. Meth., **A386**, (1997) 348.

[16] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Instrum. Meth., **A389**, (1997) 81-86. See also $http://root.cern.ch/$.

[17] S. Dris, K. Gill, J. Troska, F. Vasey, *Predicting the Gain Spread of the CMS Tracker Analog Readout Optical Links*, CMS NOTE 2006/145.

[18] M. Raymond, et al., *Final results from the APV25 production wafer testing*, Proceedings of the 11th Workshop on Electronics for LHC experiments, CERN/LHCC 2005-038, pp. 453-457, 2005.

[19] P. Biallass, private communication.

[20] A.-S. Giolo-Nicollerat, F.-P. Schilling, *Reconstruction of K^0_S and Λ^0 Mesons with the CMS Silicon Tracker*, CMS IN 2007/017.

[21] M. Chiorboli, CMS NOTE in preparation.

[22] J. Podolanski, R. Armenteros, *Analysis of V-events*, Phil. Mag. **45** (1954) 13.

[23] F. Siklér, *Reconstruction of V^0s and photon conversions with the pixel detector*, CMS AN-2006/101.

[24] S.A. Baird et al., *The Front-End Driver card for the CMS Silicon Strip Tracker Readout*, Proceedings of the 8th Workshop on Electronics for LHC experiments, CERN/LHCC 2002-034.

[25] L. Feld, W.D. Glessing, R. Hammarström, *Thermal Properties of the Silicon Microstrip Endcap Detector*, CMS NOTE 1998/018.

[26] C. Delaere, *Procedure for fine delay adjustement of the CMS tracker*, CMS NOTE in preparation

[27] R. Mackeprang and A. Rizzi, *Interactions of Coloured Heavy Stable Particles in Matter*, hep-ph/0612161.