Track Reconstruction with Maximum Likelihood:
an application to Silicon Microstrip detectors.

Gregorio Landi,

Depart. of Physics and Astronomy, University of Firenze; INFN Sezione di Firenze

1. G. Landi, Nucl. Instrum. Meth. A **485**, 698 (2002) .
2. G. Landi, Nucl. Instrum. Meth. A **497** 511 (2003).
3. G. Landi, Nucl. Instrum. Meth. A **554** 226 (2005).

The least squares method (K. Gauss 1809)

Is a linear method (or easily linearized) fast to implement for estimating a large number of unknown parameters.

Limitations

1. For generic identical probability distribution of errors, the least squares method is optimal in the class of linear estimators.

2. For the gaussian error distributions, the least squares method coincides with the maximum likelihood estimator and it is optimal in the class of all (linear and nonlinear) unbiased estimators.

What happen to the non-gaussian probability distributions?

Do we have an estimate of the error for the inappropriate use of the least squares method?

Let us compare the track reconstruction starting from different probability distributions.

Linear tracks, 5 points (3 degree of freedom), orthogonal incidence, impact points in zero, detector separation similar to the PAMELA tracker.

$$\text{Gaussian probability distribution } P_G(x) = \frac{\exp(-\frac{x^2}{2\sigma^2})}{\sqrt{2\pi}\sigma}$$
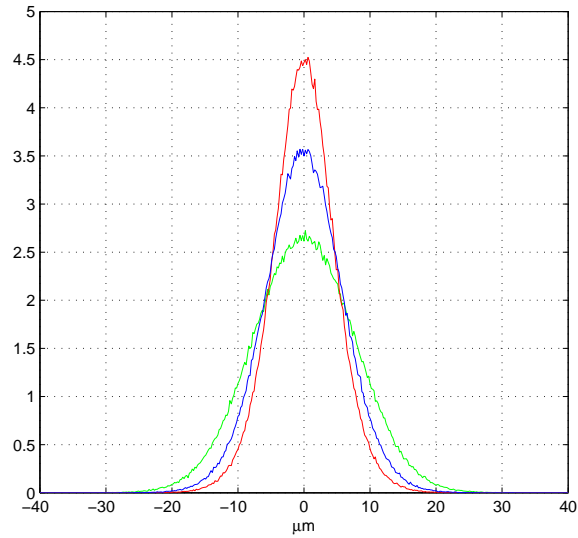

Figure 1: Green line: original data distribution, blue line: residuals, red line: differences of the reconstructed positions and the exact ones.

$$\text{Cauchy distribution (Breit-Wigner) } P_C(x) = \frac{1}{\pi}\frac{\sigma}{x^2+\sigma^2}$$
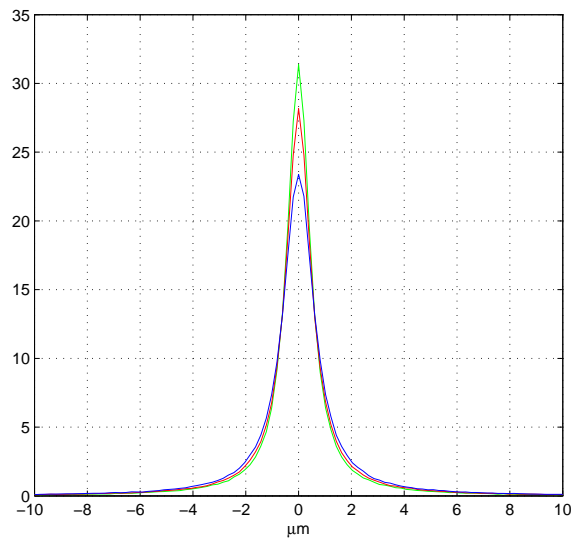

Figure 2: Cauchy error distribution, green line: original data distribution, blue line: residuals, red line: differences of the reconstructed positions and the exact ones.
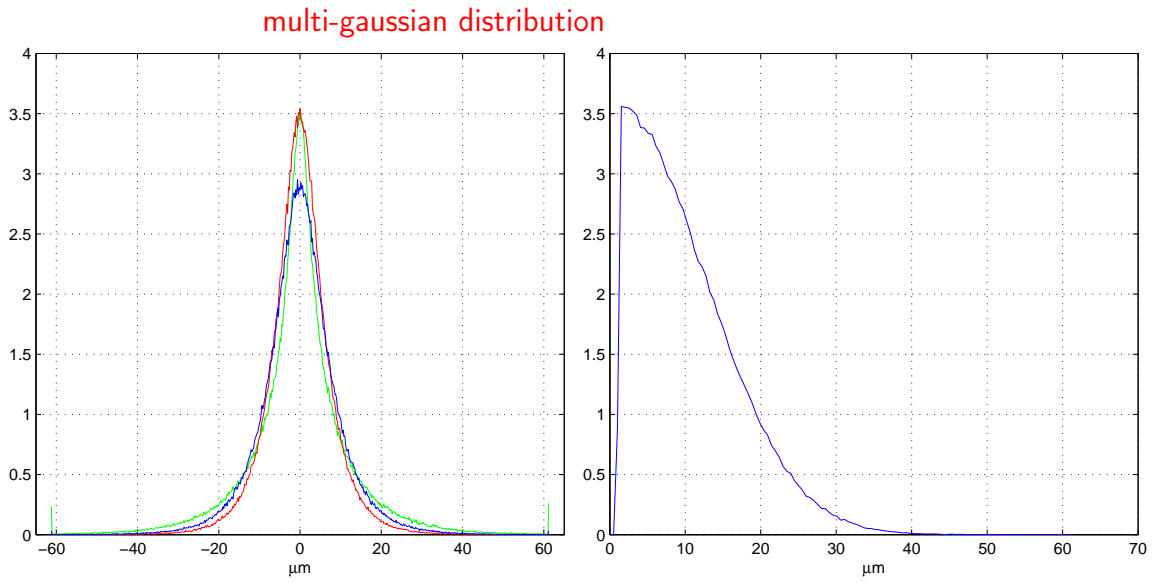
multi-gaussian distribution

Figure 3: The left plot is for a multi-gaussian error distribution, green line: original data distribution, blue line: residuals, red line: differences of the reconstructed positions and the exact ones. The right plot is the distribution of the σ used

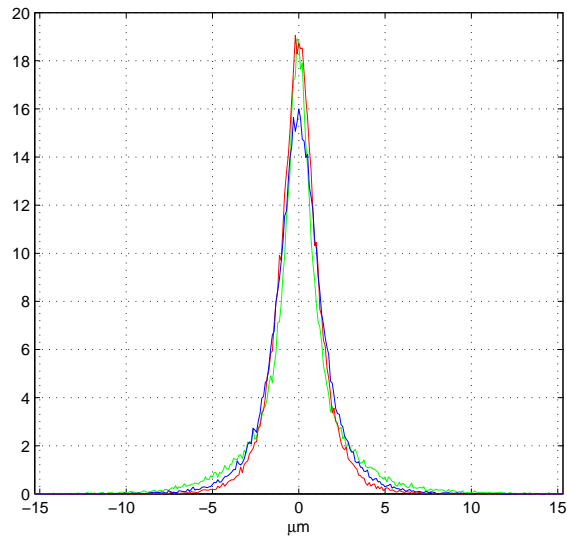Simulation Microstrip error distribution (PAMELA-like x-side, η_2 reconstruction)

Figure 4: Green line: original data distribution, blue line: residuals, red line: differences of the reconstructed positions and the exact ones.

4

The redundancy of 3 degree of freedom improves the gaussian case (around $\sqrt{3}$), it has no effect at all for the Cauchy distribution and a negligible effect for multi-gaussian simulation and the microstrip case

We have to remind that the least squares method for data extracted from a single gaussian distribution coincides with the maximum likelihood estimation

The maximum likelihood estimation

ML Cauchy distribution $F(\gamma, \beta) = \sum_i \ln[(x_i - \gamma Y_i - \beta)^2 + \sigma^2]$
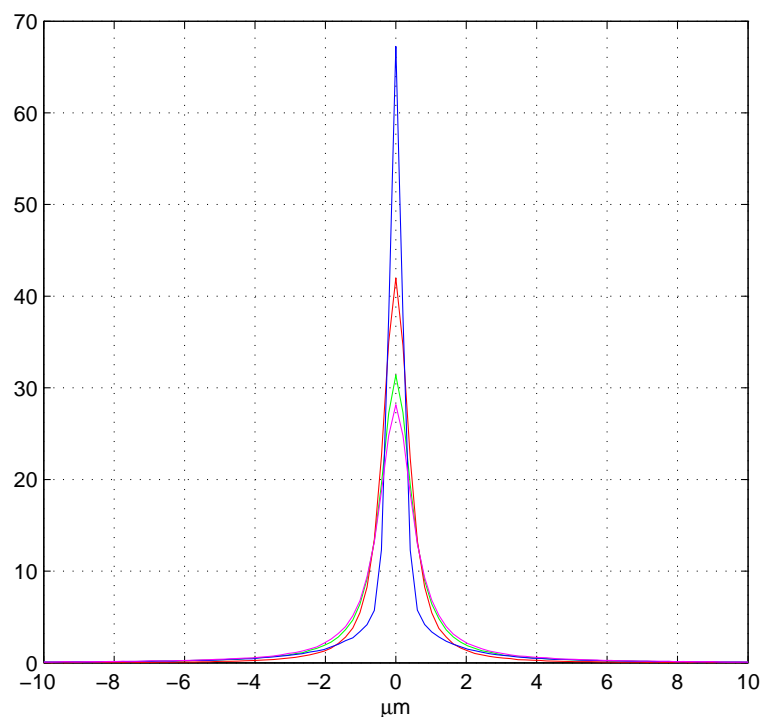

Figure 5: Cauchy error distribution, blue line: original data distribution, green line: residuals (maximum likelihood), red line: differences of the reconstructed positions and the exact ones (maximum likelihood), magenta line, differences of .... (least squares).
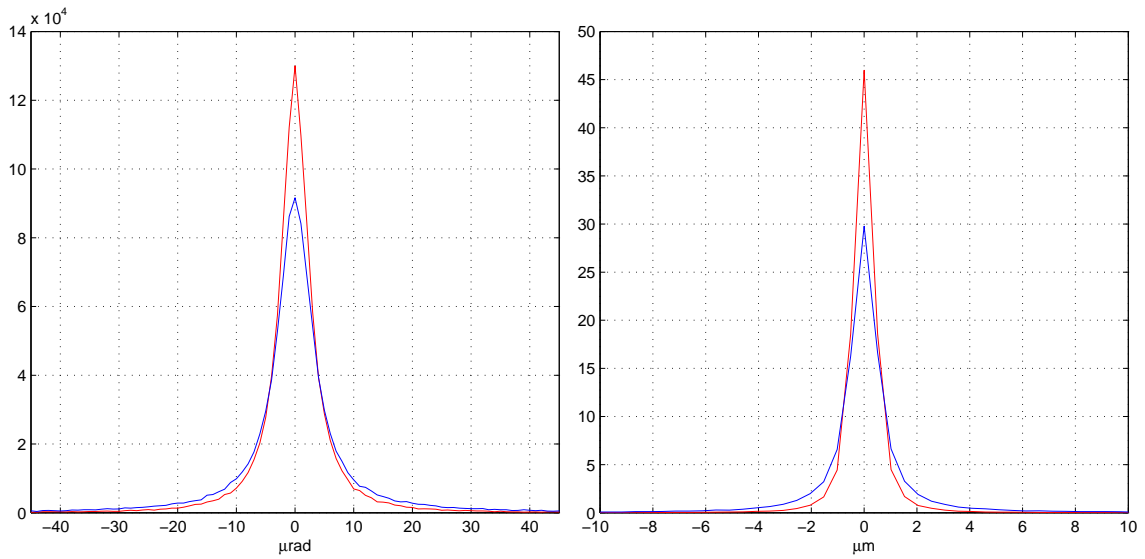
Figure 6: Distribution of γ (left) and β (right). Blue line: reconstructed with the least squares, red line: reconstructed with the maximum likelihood.

Maximum Likelihood for multi-gaussian distribution (linear equation)
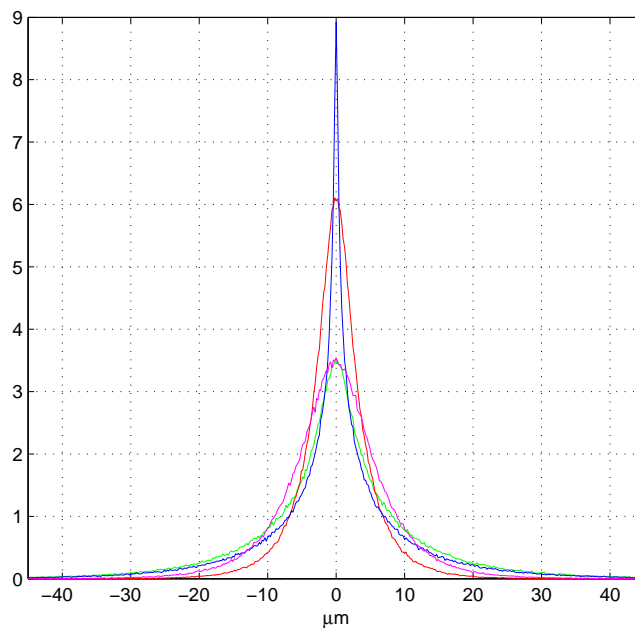

Figure 7: Multi-gaussian error distribution, green line: original data distribution, blue line: residuals (maximum likelihood), red line: differences of the reconstructed positions and the exact ones (maximum likelihood), magenta line, differences of .... (least squares).
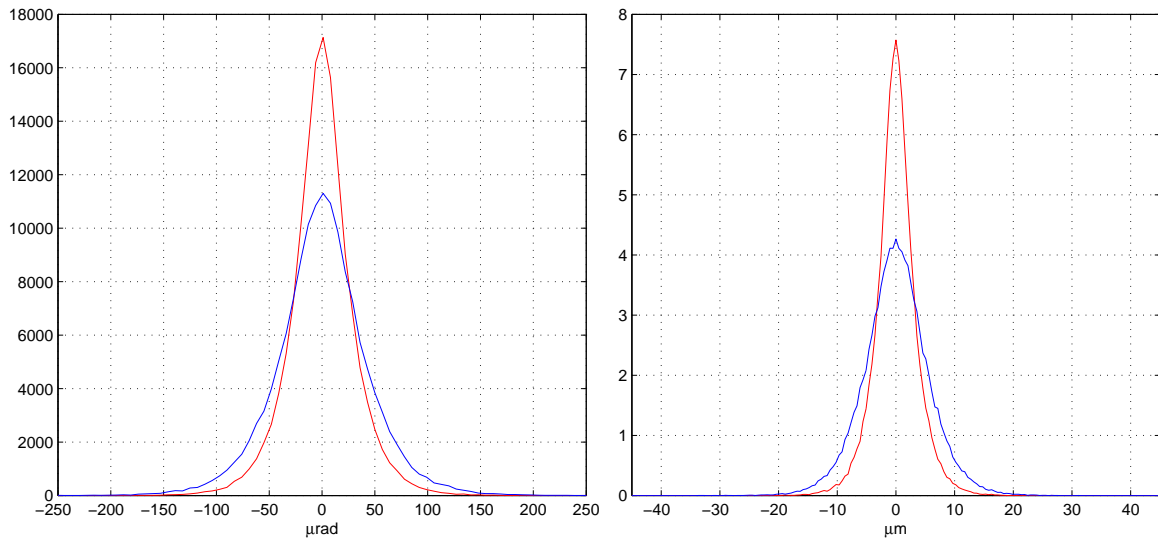
Figure 8: Distribution of γ (left) and β (right). Blue line: reconstructed with the least squares, red line: reconstructed with the maximum likelihood.
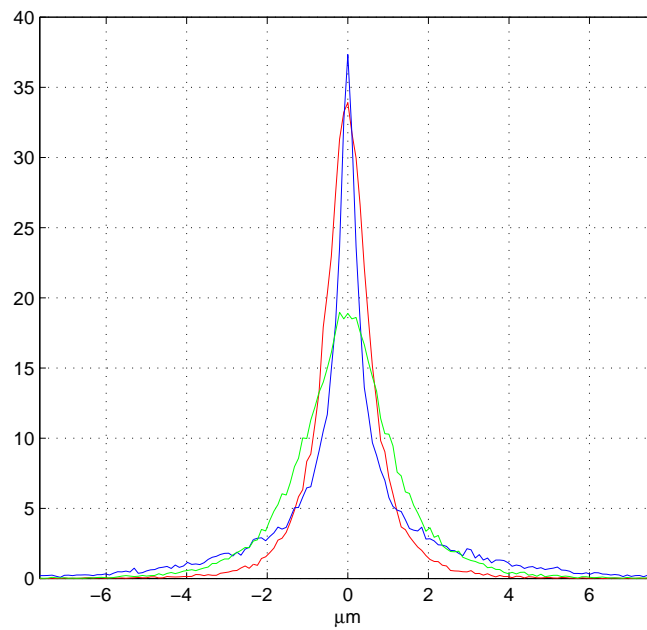

Figure 9: blue line: residuals (maximum likelihood), red line: differences of the reconstructed positions and the exact ones (maximum likelihood), green line: differences of the least squares .
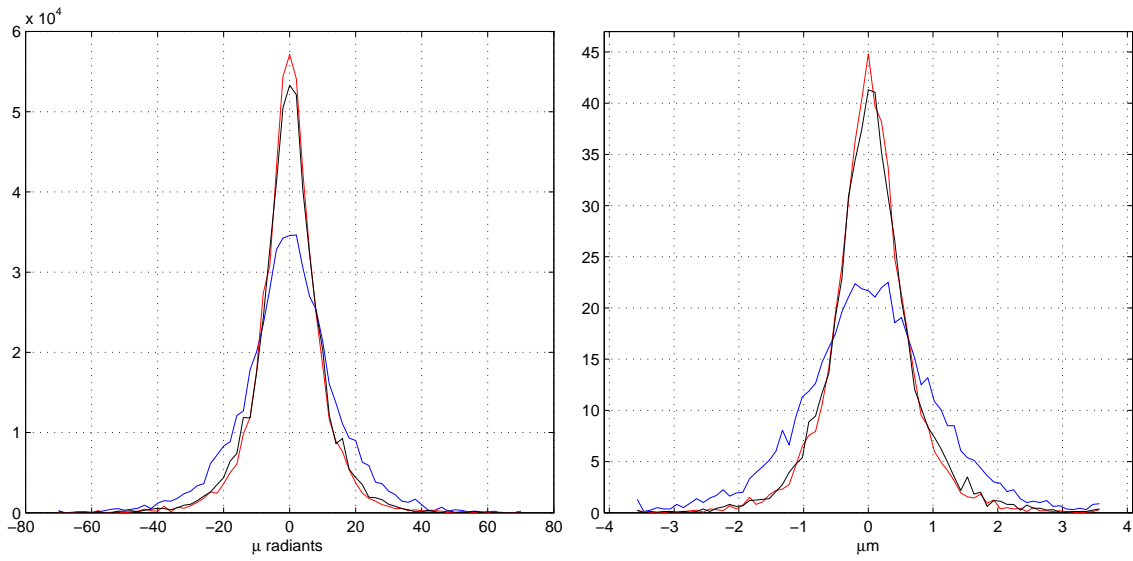
7

Figure 10: Distribution of γ (left) and β (right). Blue line: reconstructed with the least squares, red line: reconstructed with the maximum likelihood.
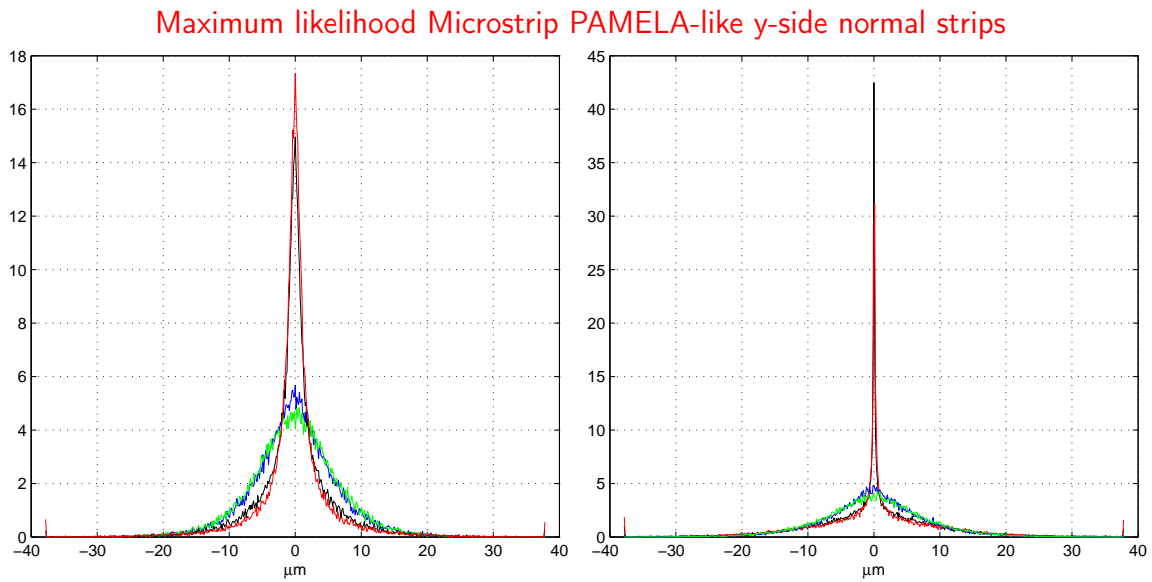

Figure 11: in the left plot: red line differences of the reconstructed positions respect to the exact ones (maximum likelihood), blue line: differences of ...(least squares η_2), green line differences of ...(least squares η_3). To the right, distributions of the residual.
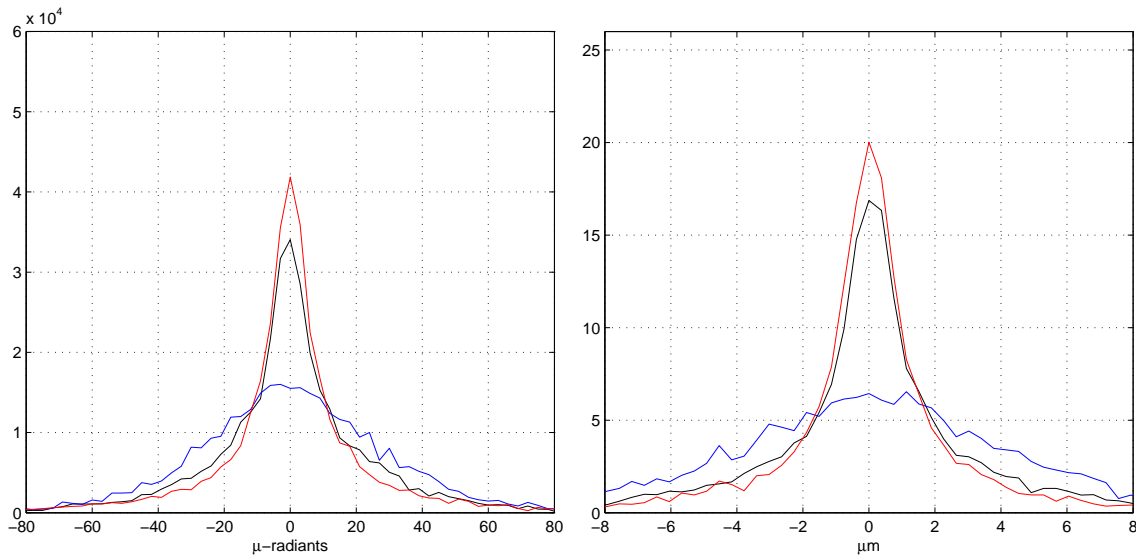
Figure 12: No floating case. Distribution of γ (left) and β (right). Blue line: reconstructed with the least squares, red line: reconstructed with the maximum likelihood.

The maximum likelihood always implies a drastic improvement of the parameter estimation, even if it is very complex. *The determination of the likelihood function implies an impossibly difficult numerical integrations in N-dimensional spaces. (by J. Orear).* This comment describes exactly the problem for the probability calculation.

The probability distributions for the reconstruction algorithms

The following figures try to illustrate the probability distributions to be handled in a microstrip detector. These probabilities are different in different points of the strip.
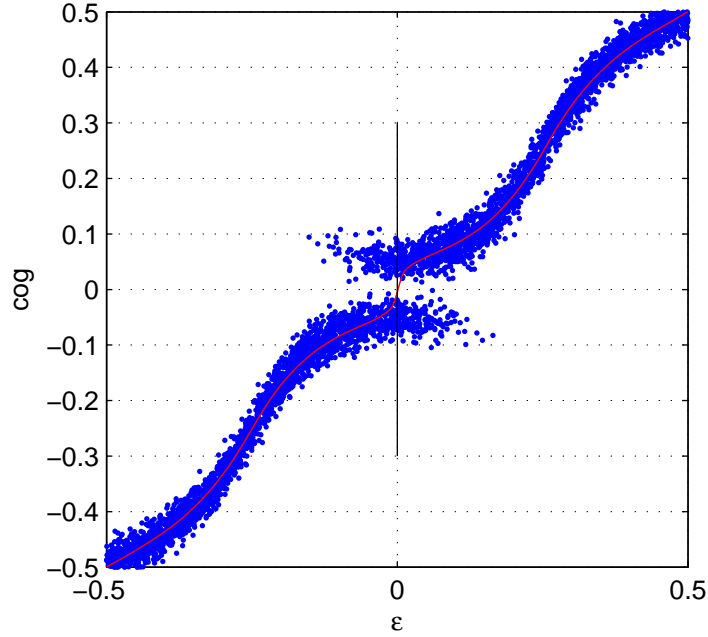
9

Figure 13: Scatter plot of x_{g2} in function of its exact impact point for a two strip algorithm. The red line is $x_{g2}(\eta_2)$ as reconstructed from the data

The starting point of every reconstruction is the center of gravity of the signals collected by the strips. It is a non linear combination of random variables E_i, the energies of the strips:

$$x = \frac{\sum_i E_i R_i}{\sum_i E_i} \qquad (1)$$

The parameters R_i are the relative positions of the strip centers, their reference system is on the strip with the maximum value. We have to consider COG algorithms with two, three, four and five strips with their discontinuities at $x = 0$ for the even strip algorithms and at $x = \pm 1/2$ for odd algorithms.

We assume that the form of the probability distributions for the random variables E_i are gaussian distributions with their averages the energy released by the particle in the strip and their σ the electronic noise of the strip. The

electronic noise of the strips in the absence of signal is well reproduced by a gaussian (with a small distortion by the noisy strips).

Even with these simplifying assumptions the the integrals for the probabilities have no closed form, essential for large numerical use. We resort to various type of approximations the last ones are very realistic, but we are fighting to obtain exact results. Explicit forms of gaussian integrals (not quoted in the tables) have been isolated and so we hope for the last one.

A characteristic of all the probability distributions is their similarity with a Cauchy distribution

$$\lim_{x \to \infty} P(x) = P_{x_{g2}}(x, a_1, a_2, a_3, \sigma_1, \sigma_2, \sigma_3) \approx \frac{constant << 1}{x^2}$$

The quality of the results are illustrated in the following figures that represent the probability distributions along the vertical line of figure 13. Here with Landau fluctuation, we intend the particle path divided in 15 intervals with an energy fraction given by a Landau distribution. The total energy is normalized to 150 ADC counts as above. No effect of this additional fluctuation is seen in the plot.

Two strip distribution for x_{g2} around $\varepsilon \approx 0$
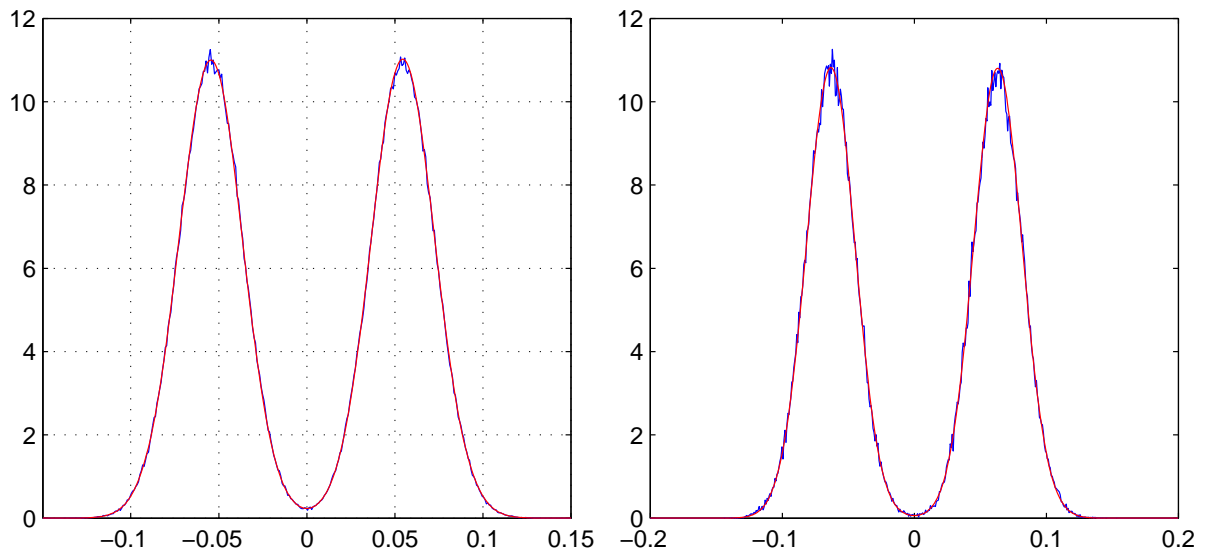
11

Figure 14: to the left: Probability distribution for x_{g2} at fixed $\varepsilon = 0.0001$ for a simulated distribution (blue) (1 million events of fixed noiseless energies) and our calculated probability distribution (red). To the right: a Landau fluctuation is added to the simulated energy distribution (blue line) 200000 events. Red line calculated probability distribution
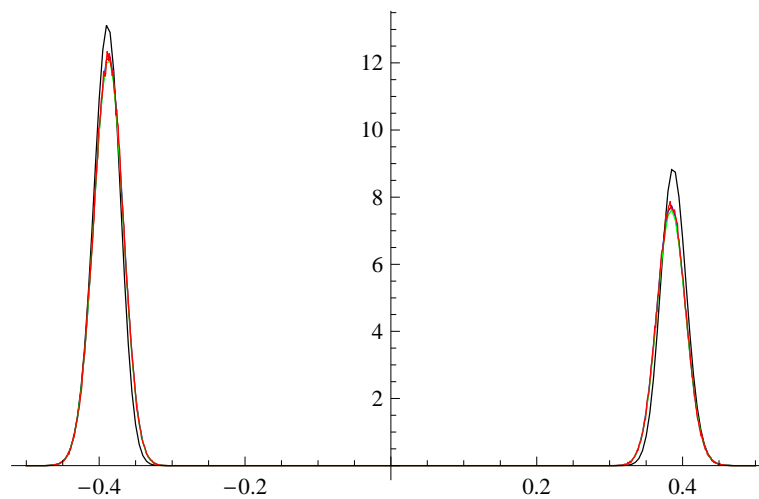

Figure 15: Probability distribution of x_{g2}, $\theta = 20^o$, $\varepsilon = 0$ in for the simulated data (red) and the numerical integral (blue), initial approximation (black), the final approximation (green) 500000 events

Probability distribution for x_{g3} around $\varepsilon \approx 1/2$
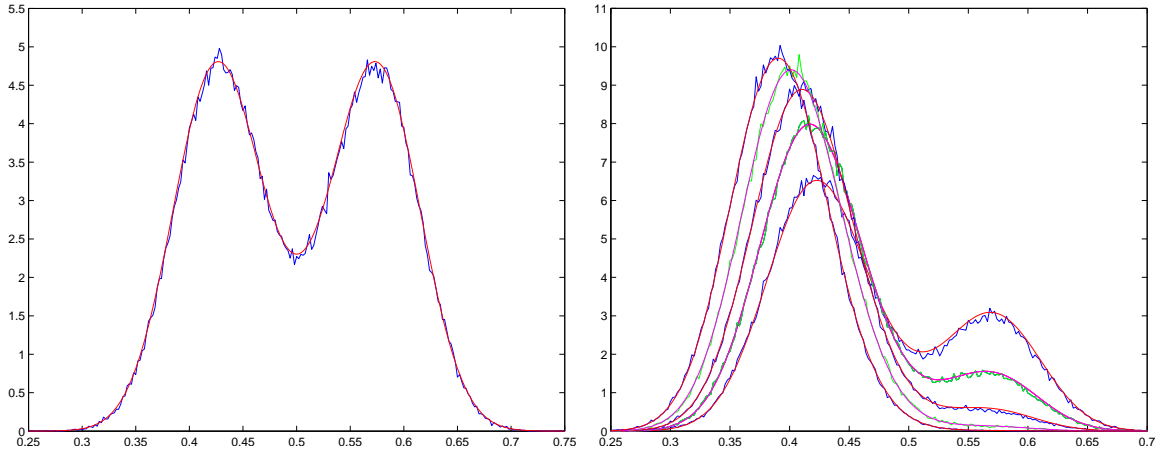

Figure 16: Comparison of $P_d^{xg3}(x)$ with the simulations, floating strip detector, 200000 events. To the left the impact point $\varepsilon = 1/2$ $\sigma_i = 3.9$, and impact direction $\theta = 0^O$. To the right, impact points $\{0.48, 0.46, 0.44, 0.42, 0.40\}$.

Problem:

These probability distributions have a fixed impact point, and they generate the x_g distribution. We need the reverse result: for fixed x_g to generate the ε distribution.

Without a solution of this problem, the probability distributions are useless.

A possible solution

The a possible solution to this critical element is given by the following procedure. We have to remind a property of the signal collected by lossy strips discussed in 1.

For an array of lossy strips, if the signal distribution contains at least a convolution with an interval function with the strip period, the signal collected by the strips is independent from the impact point

13

$$\sum_k h(kT - \varepsilon) = \frac{1}{T} \sum_L P(-\frac{2\pi L}{T})\Phi(-\frac{2\pi L}{T}) \exp[-\frac{2\mathrm{i}\pi L\varepsilon}{T}]$$

$$F.T. \ of \ \Pi(\frac{y}{T}) \Rightarrow T\frac{\sin(\omega T/2)}{\omega T/2}$$

In this case we can apply the reconstruction theorem demonstrated in 1.

The loss: if we delete (periodic) blocks of strips around the strips under study we have the loss.

We could insert a convolution in the signal collected by the strips (survived to the loss) with a sliding window that selects the particles incident in its interior in a test beam. The window slides step by step to cover a period of the surviving strip array. The length of the window must be a period of the array of the active strips. The step size must be sufficient to recover all the details of the response function. The number of collected events at any step must very similar to produce the minimum possible energy fluctuations.

Each active strip collects an energy at any step of the window. A center of gravity $x_g(\varepsilon)$ is calculated. A Fourier series of $x_g(\varepsilon) - \varepsilon$ allows to filter noise with a cut in the number of its elements.

The response function is obtained from:

$$\frac{\mathrm{d}x_g(\varepsilon - T/2)}{\mathrm{d}\varepsilon} = \sum_k f(kT - \varepsilon)$$

We have not such a complex setup so we try to extract this function from the data we have. In the simulations we test this method to be sure of its correctness.

To simulate a sliding window, we spread our data (generally produced on a strip) on many strips.

The convolution is approximated as:

$$h(kT - \varepsilon_f) = \int_{-\infty}^{+\infty} \Pi(\frac{\varepsilon_f - \varepsilon}{T}) f(kT - \varepsilon) \, \mathrm{d}\varepsilon \approx \sum_{\varepsilon_f - T/2}^{\varepsilon_f + T/2} f(kT - \varepsilon_i) \Delta_i$$

The parameters $\{\varepsilon_i\}$ are the impact points of the incident particles, Δ_i their distances. These data are evidently unavailable, we try to substitute the impact points with their approximations $\eta(x_{g2})$ and $\eta(x_{g3})$.

After many delicate corrections studied in the simulations, we reached a sufficient elimination of almost all the artifacts introduced by the substitution of $\eta(x_{g2,3})$ in place of the $\{\varepsilon_j\}$.

Application to the data

We have test-beam data of the PAMELA/ADAMO detector, 5 layers of detectors (Vannuccini file $\theta = 0^o$). To increase the statistics, we consider the detectors identical, and use all the data together.
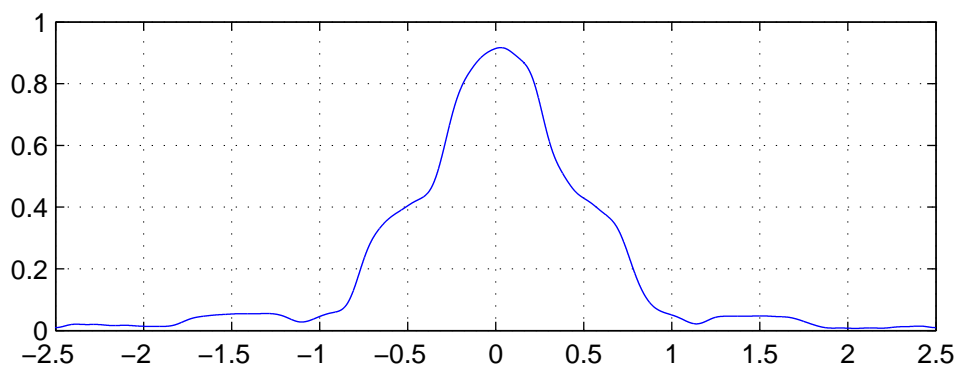

Figure 17: The reconstructed response function (floating strip size or x-side) normalized to 1 over 5 strips
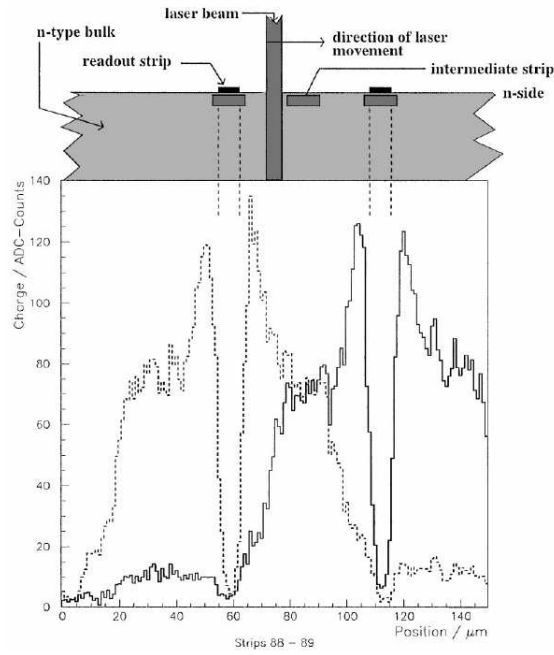
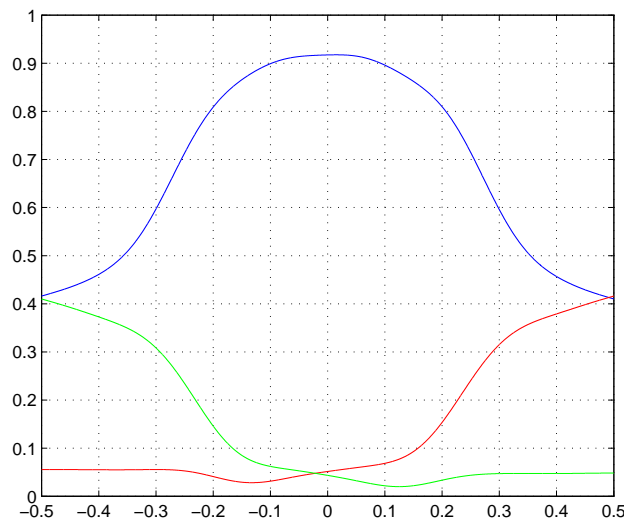Figure 18: The measured response function with a laser beam, even here the structures around ±1.5 are present

Figure 19: The three energies $a1(\varepsilon)$, $a2(\varepsilon)$, $a3(\varepsilon)$)

The pattern of the energy distribution on the strips is obtained cutting the previous response function at the points $\pm 1/2, \pm 3/2$. With

16

$a_1(\varepsilon), a_2(\varepsilon), a_3(\varepsilon)$ we calculate a three strip COG and differentiating it gives the convolution of the average particle-signal with the part of the strip property which deviates from a perfect integrator. For the floating strip side is principally a couple of Dirac δ-functions.
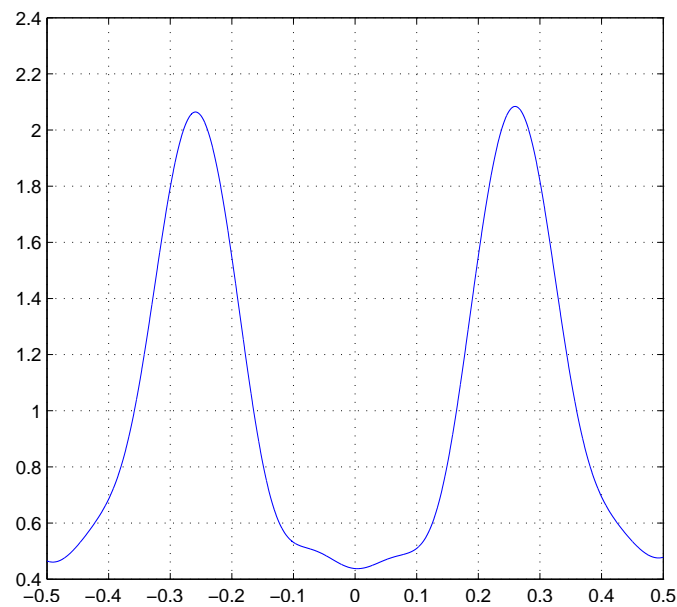
Figure 20:   The reconstructed function obtained differentiating (numerically) the expression $[a_1(\varepsilon) - a_3(\varepsilon)]/[a_1(\varepsilon) + a_2(\varepsilon) + a_3(\varepsilon)]$ in ε

The quality of this reconstruction can be compared with the first one of REF.3.
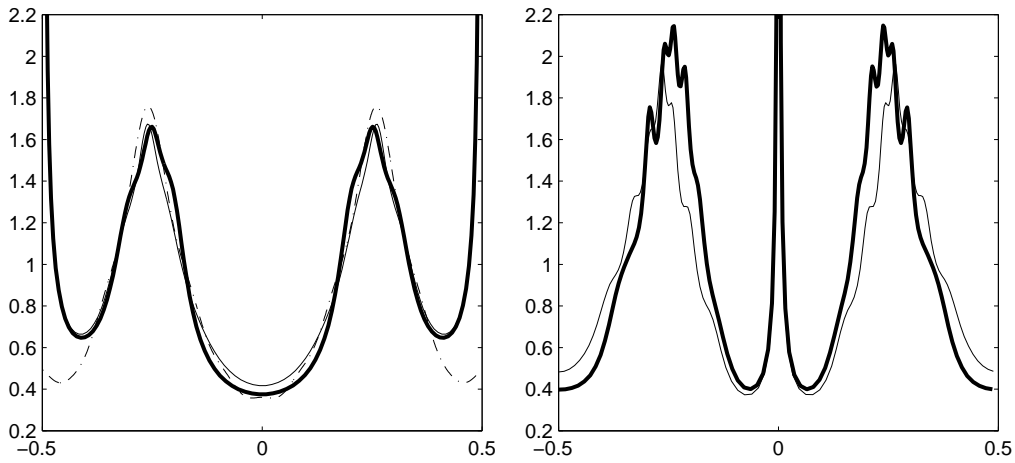
Figure 21: To the left $\varphi(\varepsilon)$ obtained by the x_{g3} algorithm applied to the data (thick line), given by the simulations (thin line). To the right, the x_{g2} is used (From REF.3).

To test the quality of the energy shapes, we compare the reconstructed histograms (maximum energy) for x_{g2} and x_{g3} with the these of the data.
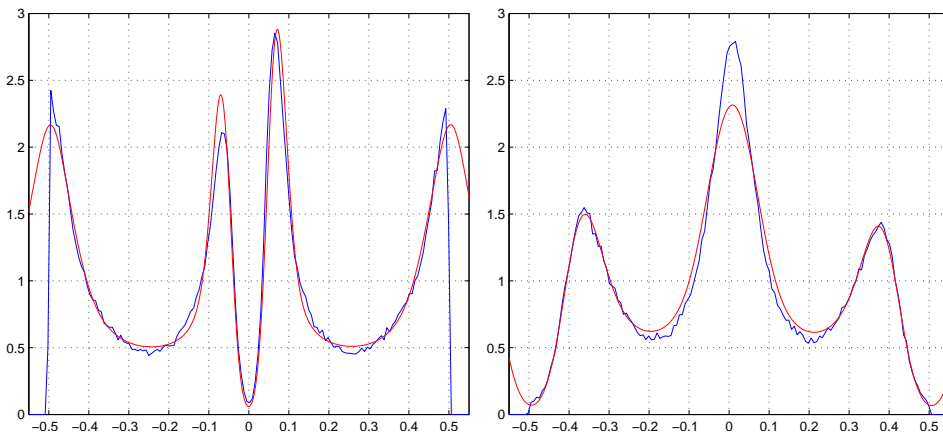

Figure 22: To the left the histogram of x_{g2}, the red line is the reconstructed x_{g2} histogram with the probability distribution and the energies of fig.19 with a fixed normalization energy of 142 ADC-counts. To the right the histogram of x_{g3} and its reconstructed form. The agreement is acceptable given their strong sensitivity to small differences $(1/(\mathrm{d}x_g/\mathrm{d}\varepsilon)$

Generation of a simulated data sample (x-side)

The simulated data are produced in the following way:

1. generate an uniform distribution of points on a strip $(-0.5, +0.5)$
2. for each point calculate the energies a_1, a_2, a_3
3. scale a_1, a_2, a_3 with the distribution of the sum of the signals of three strips (the maximum and the two lateral) of the data.
4. add a gaussian noise of 4 ADC counts (r.m.s.)

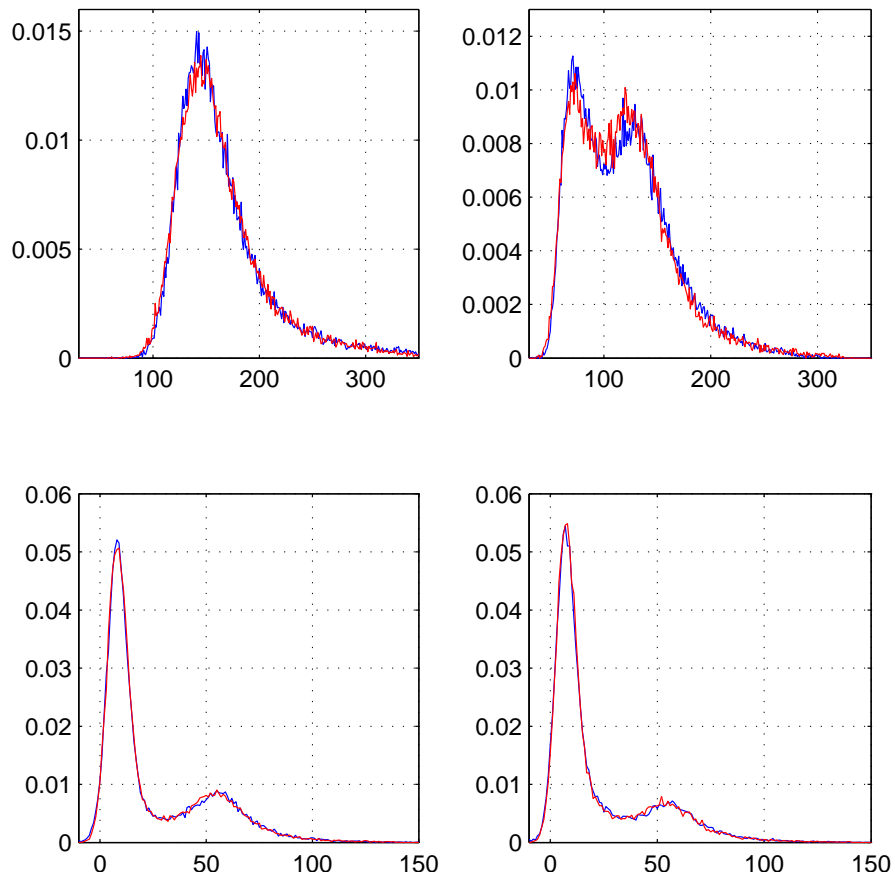The following plot illustrated the resulting distributions:

Figure 23: The blue lines are the data, the red lines are simulations. Left high: sum of the signals collected by three strips. Right high: The energies of the central strip. Left low: Energies of the right strip. Right low: Energies of the left strip.

The energy distributions of the simulated data turn out very similar to the experimental data. The doubling of the noise is unnoticeable. The energies $a_1(\varepsilon), a_2(\varepsilon), \ldots$ are very similar to the previous ones.

Track reconstruction

1. Extract from the simulated data the energies $a_1(\varepsilon), a_2(\varepsilon), a_3(\varepsilon)$
2. generate the tracks assembling five simulated points
3. subtract from the reconstructed positions $(\eta_2(x_{g2}(j)))$ the impact point $\varepsilon(j)$ (tracks parameters $\{0, 0\}$
4. define a good starting point for the minimum search algorithm
5. find the minimum on γ, β of:
   $$\sum_{j=1}^{5} -ln[P^{x2}(x_{g2}(j), \gamma Y_j - \beta + \varepsilon(j))]$$

Illegal approximation

The probability distributions for the x-side are similar to gaussian distributions around $\varepsilon(x_{g2})$, but are Cauchy-like $\rightarrow \infty$ no variance can be defined. In spite of this, a judicious cut on the range of the values gives an effective root mean square of the probability distributions.

These effective σs give a quality factor of the points in the fit, and allow the use of a fast linear equation (least squares).

The track parameters γ and β of this reconstruction is used as staring point to initialize the minimum search routine for the maximum likelihood.

The results are illustrated in the following figures with black lines. The computer time to obtain the effective σ is higher than the maximum likelihood computing. But, without a good initialization, the minimum search routine performs poorly.
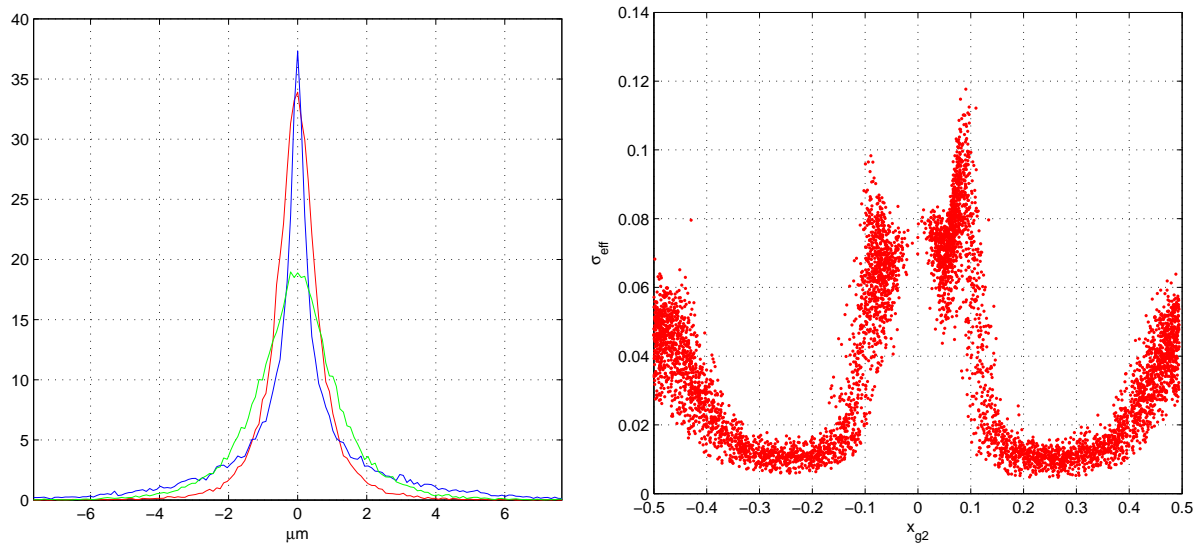
20

Figure 24: Left plot. Blue line: residuals (maximum likelihood), red line: differences of the reconstructed positions and the exact ones (maximum likelihood), green line, differences of .... (least squares). Right plot. Distribution of effective σ (6000 points)
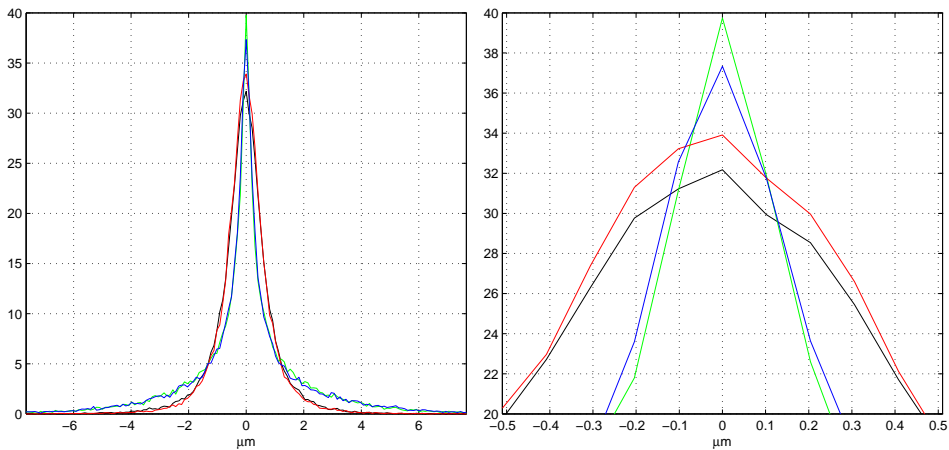

Figure 25: To the left the red line and the blue line are these of fig. 24, the green line is the distribution of the residuals of the least squares with effective σ and the black line is the distribution of the differences respect to the exact impact points. To the right is a particular of the top of the distributions.
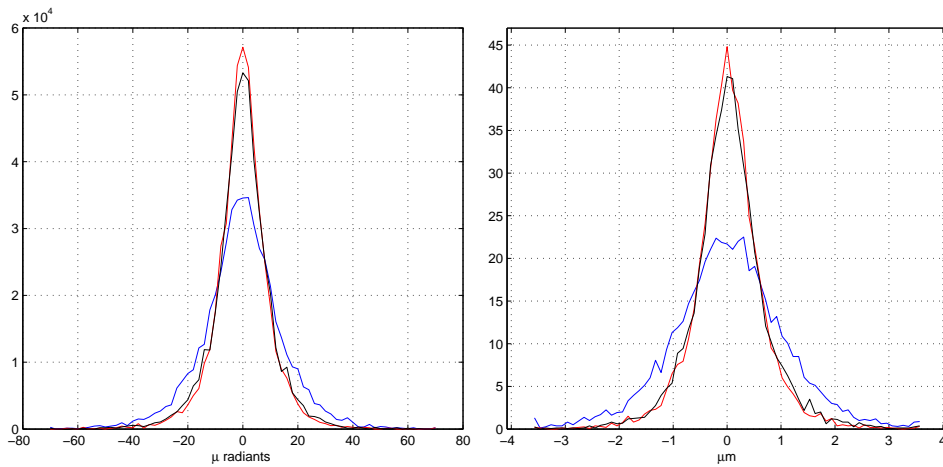
Figure 26: Distribution of γ (left) and β (right). Blue line: reconstructed with the least squares, red line: reconstructed with the maximum likelihood, black line least squares with effective σ.
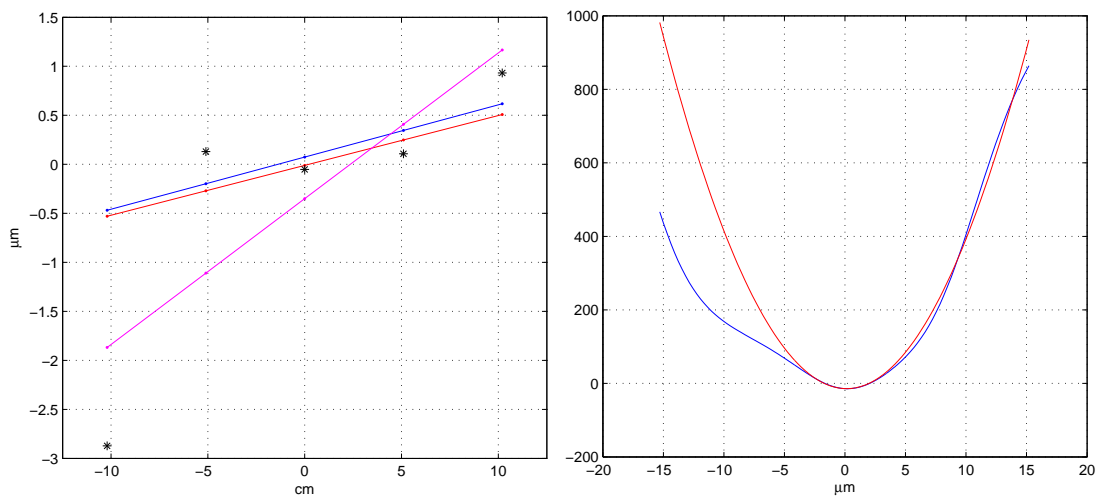
the reconstruction of a track

Figure 27: Left plot. The blue line is the maximum likelihood track, the red line is the track reconstructed with the effective σ, and the magenta line is the least squares track. To the right the $-\ln$ of the likelihood; the blue line for the full probability distributions and the red line for the effective gaussian distributions.

22
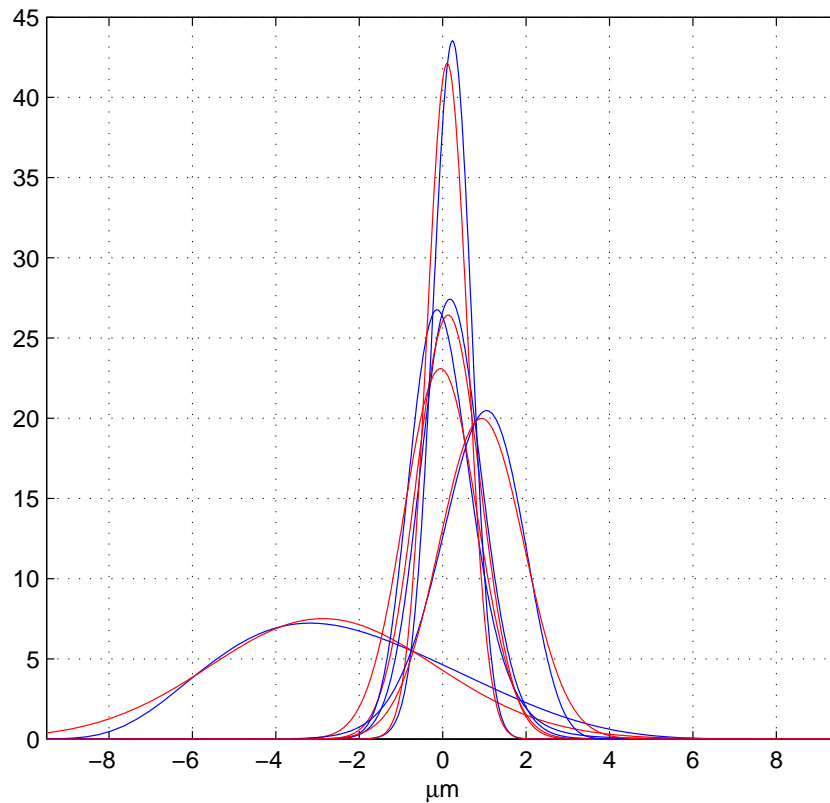
Figure 28: The blue lines are the correct probability distributions, the almost overlapping red lines are the gaussian distributions whose mean squares are extracted by the blue-line distributions.

Y-side no-floating strips

Here the noise is higher than the other side: 8 ADC counts r.m.s.

The steps of this reconstruction are similar to the previous one. The absence of the floating strips reduces the charge spreading over the other strips. Thus, the energies $a_1(\varepsilon), a_2(\varepsilon) \ldots$ have a different form, and the noise amplifies a slight artefact in a region of very low energy.
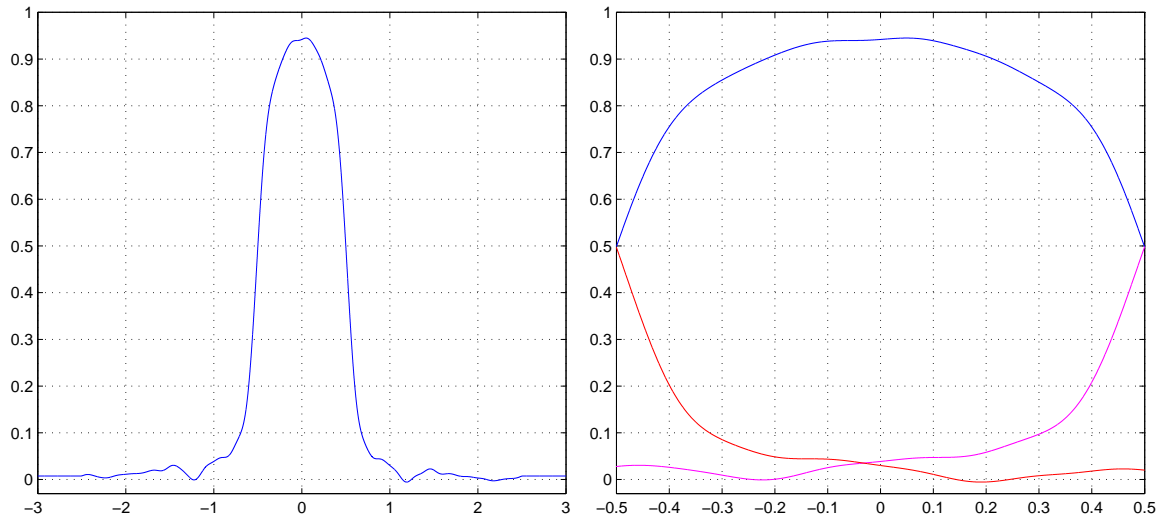
Figure 29: To the left the response function from the data of the y-side. To the right the fractional energies y-side.

The histogram of x_{g3} shows the typical slight defect we discussed.
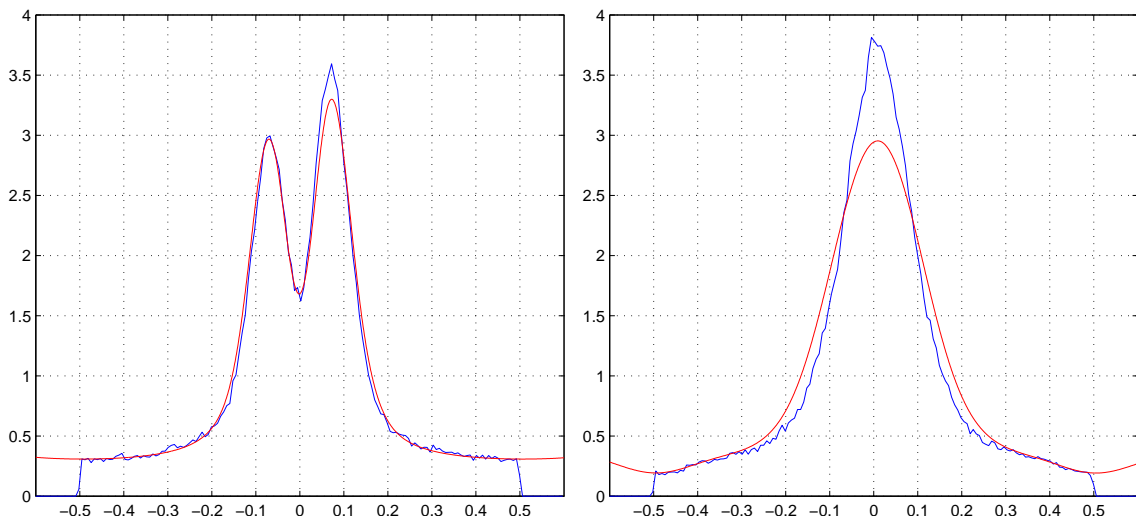

Figure 30: To the left the x_{g2} histogram, the red line is our calculated probability distribution at a fixed energy. To the right the x_{g3} histogram and the probability distribution at a fixed energy.

One noticeable difference is in the necessity to eliminate part of the noise-

widening in the distribution of sum of the experimental energies. The double addition of the noise modifies in an unacceptable way the reconstructed energy distributions. We have to compress and shift the experimental distribution before its use.
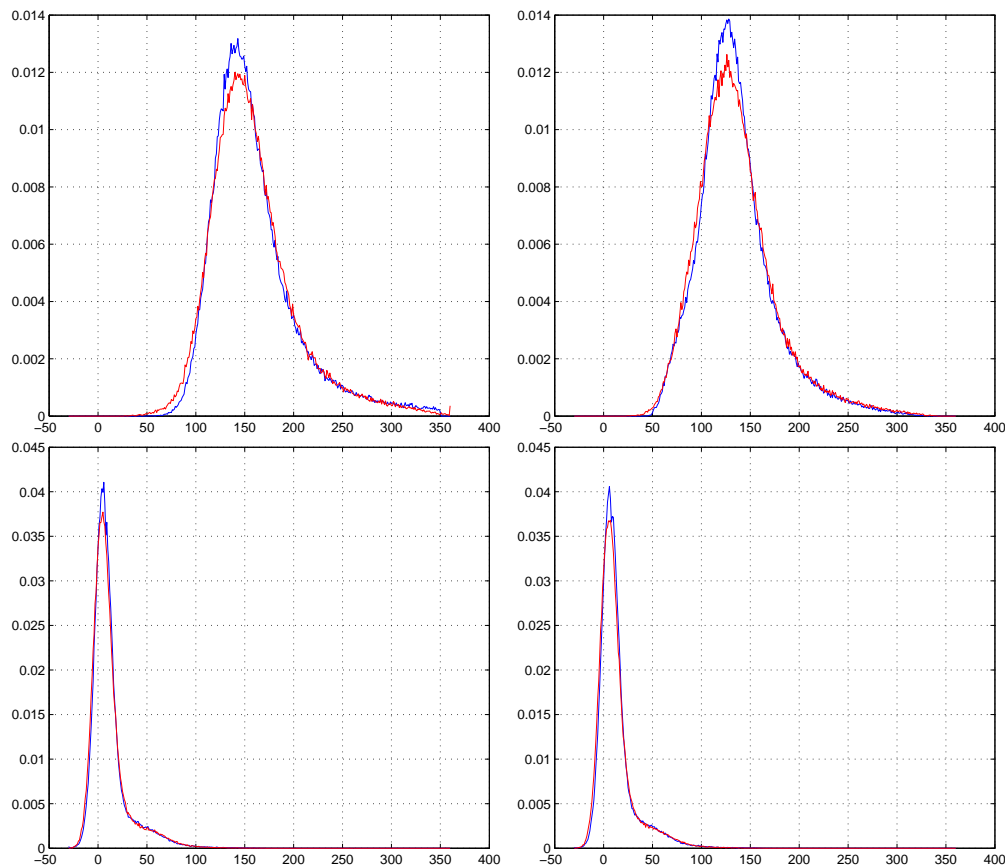
Figure 31: High left: the histogram of the energy collected by three strips, the blue line are the data and the red line is our simulation. High right: the histograms of the energy distribution the central strip. Low left: the histograms for the left strip. Low right: the histograms of the right strip.

The track reconstruction now goes as above. The differences consist in a different energy function and a larger error.

Even now the effective σs are extracted from the probability distribution for each point, and are used to initialize the minimum search routine.

25

RESULTS

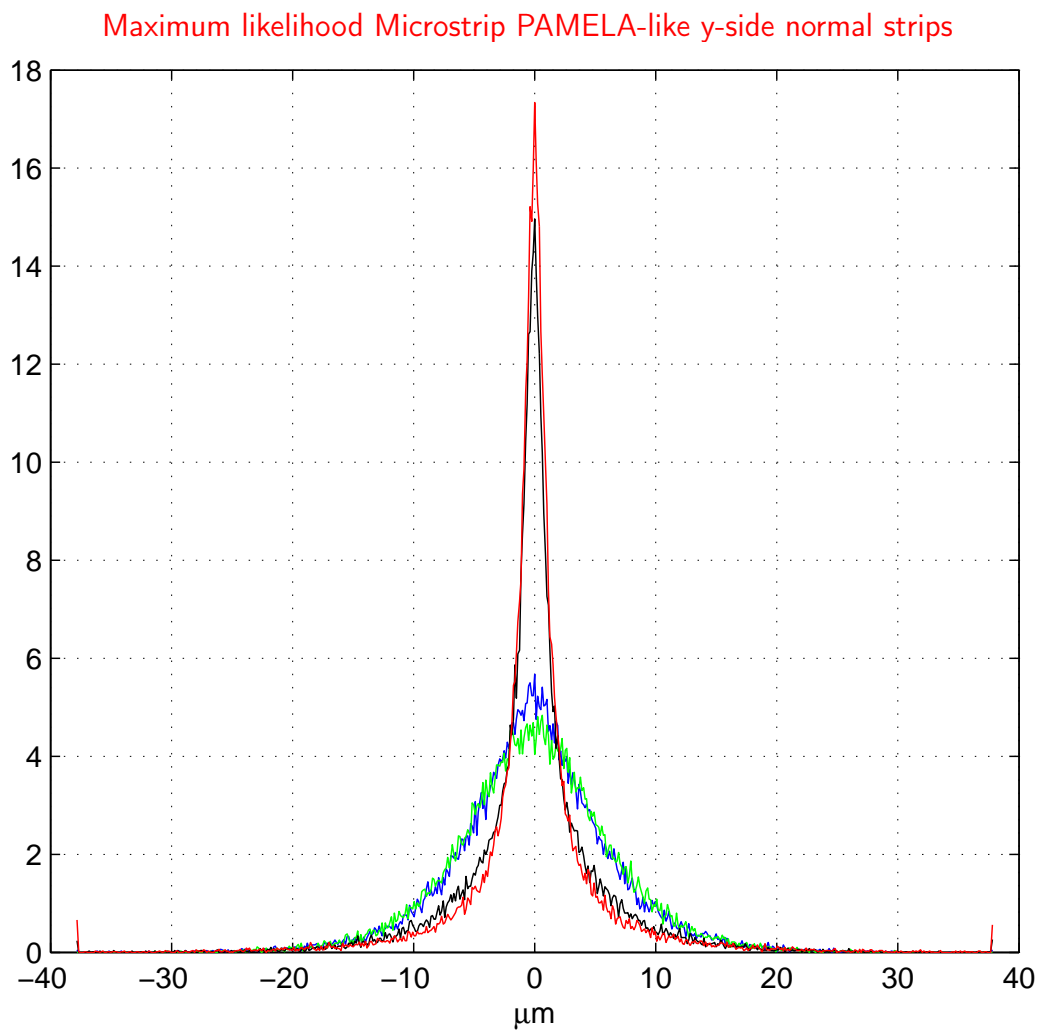Maximum likelihood Microstrip PAMELA-like y-side normal strips

Figure 32: Red line differences of the reconstructed positions respect to the exact ones (maximum likelihood), blue line: differences of ...(least squares η_2), green line differences of ...(least squares η_3).
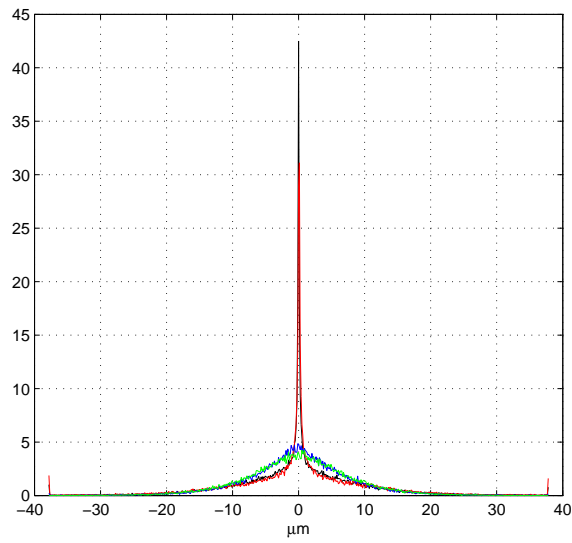
Figure 33: Red line residuals of maximum likelihood, black line: residuals for effective σ. blue line residuals of least squares from η_2, green line residuals of least squares η_3.
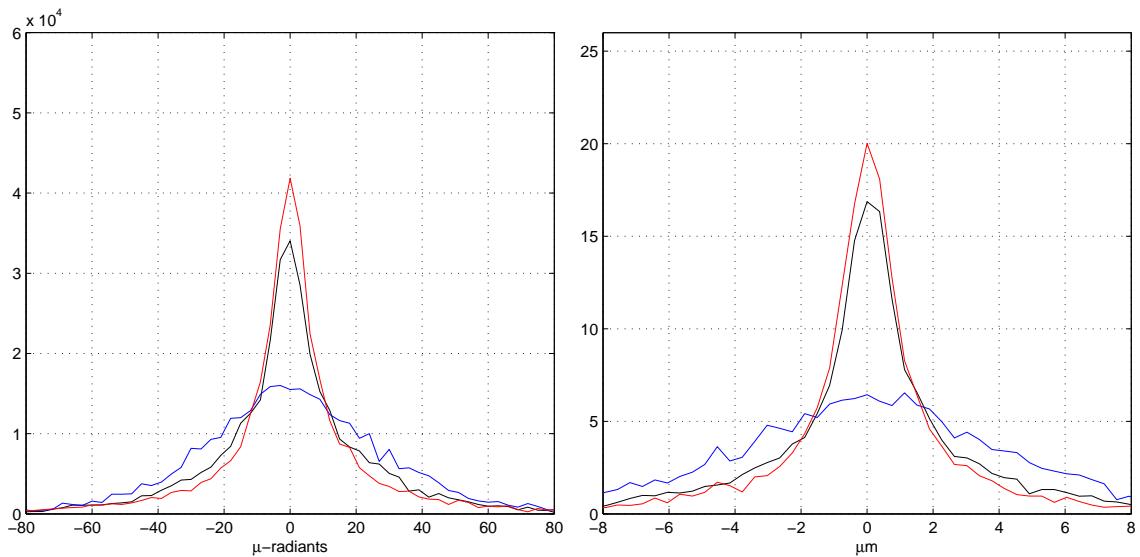

Figure 34: No floating case. Distribution of γ (left) and β (right). Blue line: reconstructed with the least squares, red line: reconstructed with the maximum likelihood.
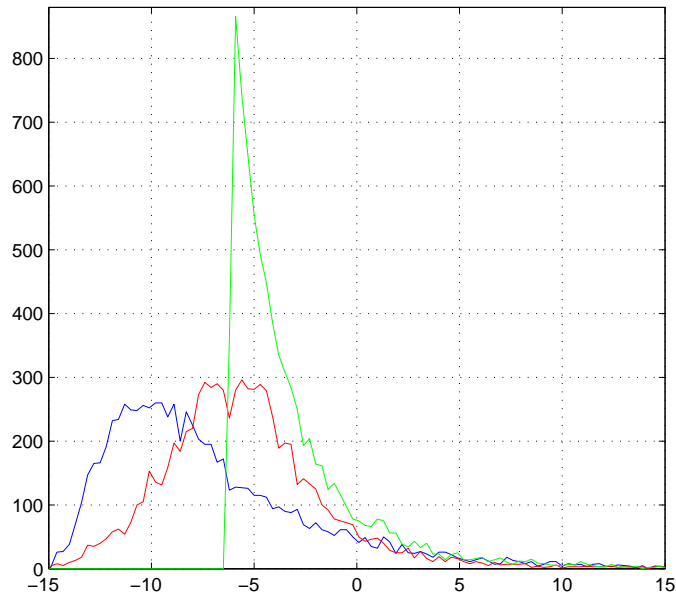
Figure 35: The histogram of the likelihood for the reconstructed tracks, the blue line is the -log of the probability of the full method, the red line for our approximate gaussians, and the green line all the gaussian distributions are identical (least squares) with the width extracted from the resolution.
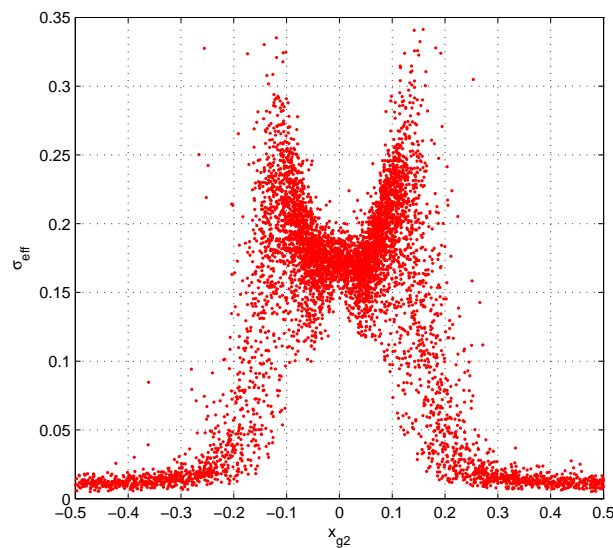

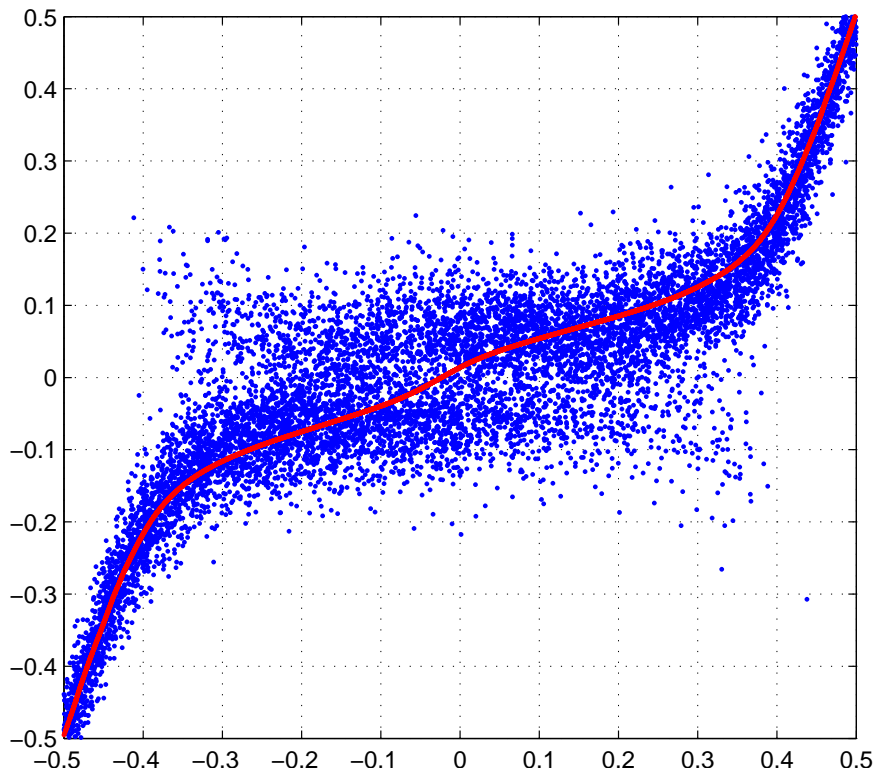Figure 36: Plot of (x_{g2}, σ_{eff}) (6000 points)

Figure 37: Scatter plot of (ε, x_{g2}) (blue points), red points $x_{g2}(\eta_2)$ (10000 points)
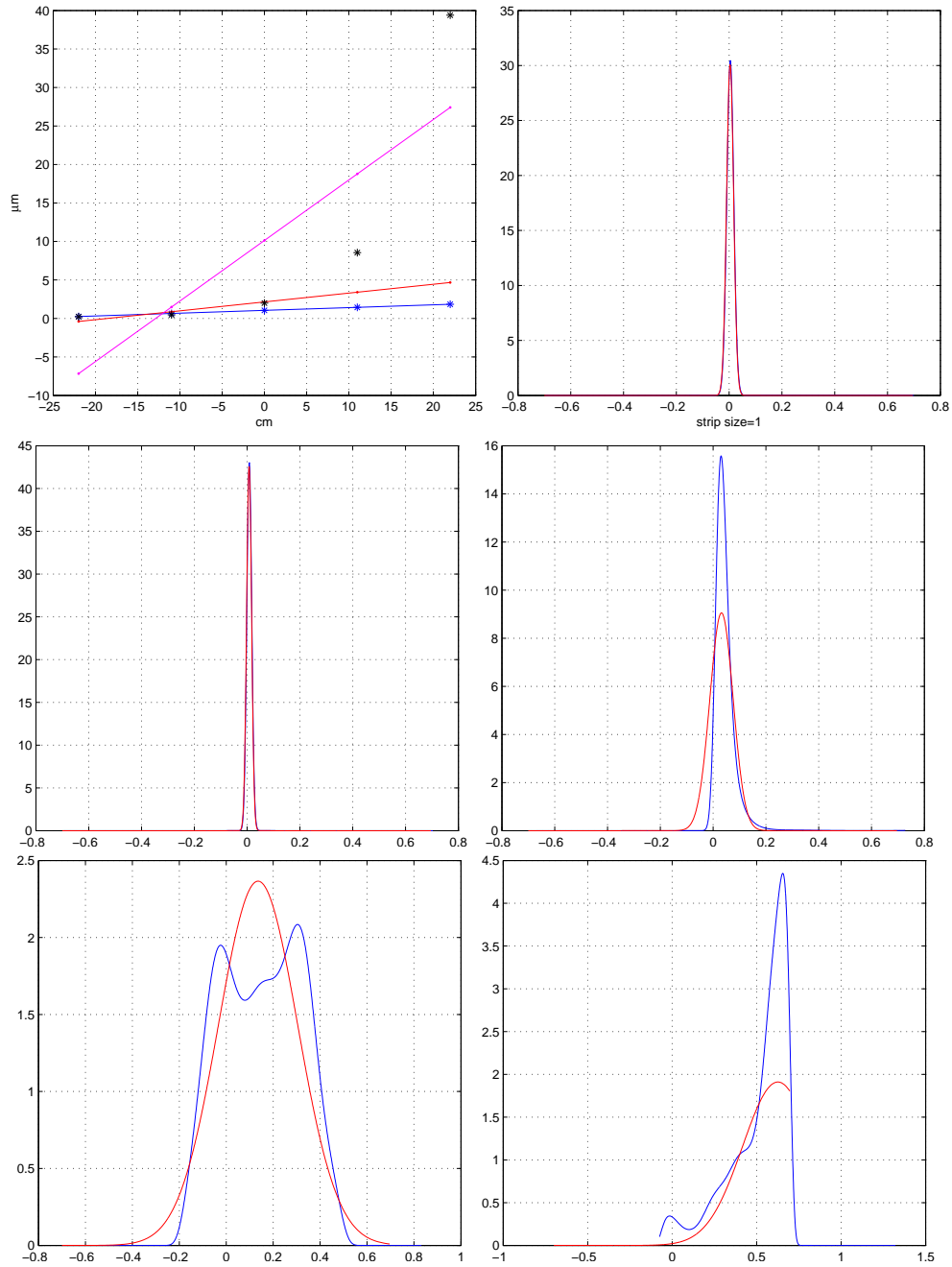
Figure 38: An excellent event (tracks number=1613). The first figure reports the reconstructed tracks: the blue line is the the result the minimum search routine, the magenta line is the least squares with identical σ, red line is with effective σs. The other figures are the probability distributions, the blue line is the probability distribution, the red line the effective gaussian.
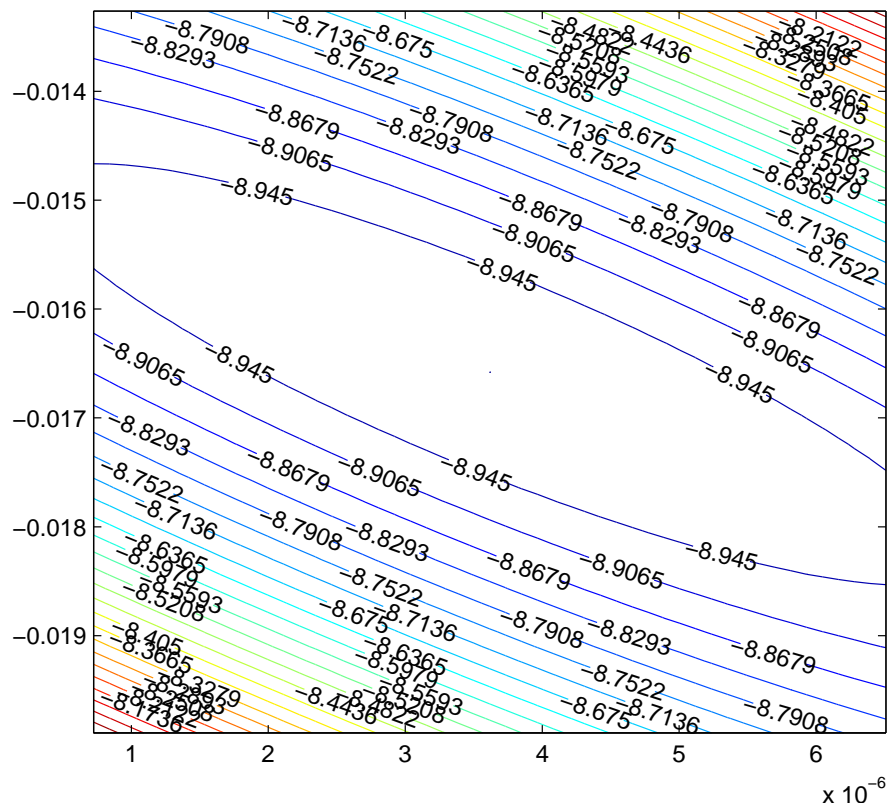
Figure 39: The form of the surface of the -log of the likelihood for the minimum search routine a good minimum is evident.
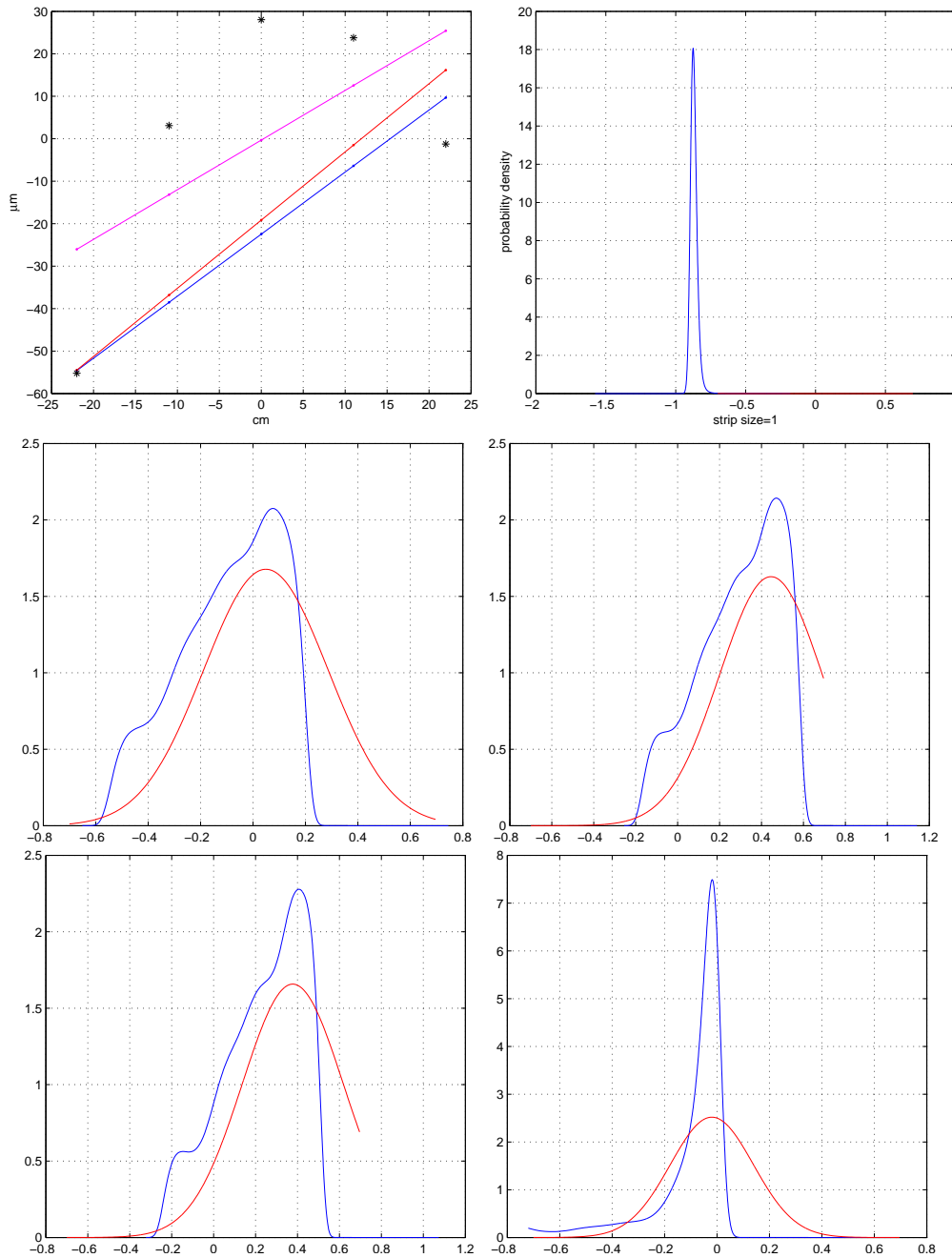
Figure 40: A worst event (tracks number=1046). The first figure reports the reconstructed tracks: the blue line is the the result of the first stop of the minimum search routine, the magenta line is the least squares with identical σ, red line is with effective σs. The other figures are the probability distributions, the blue line is the probability distribution used in the maximum likelihood, the red line the effective gaussian.
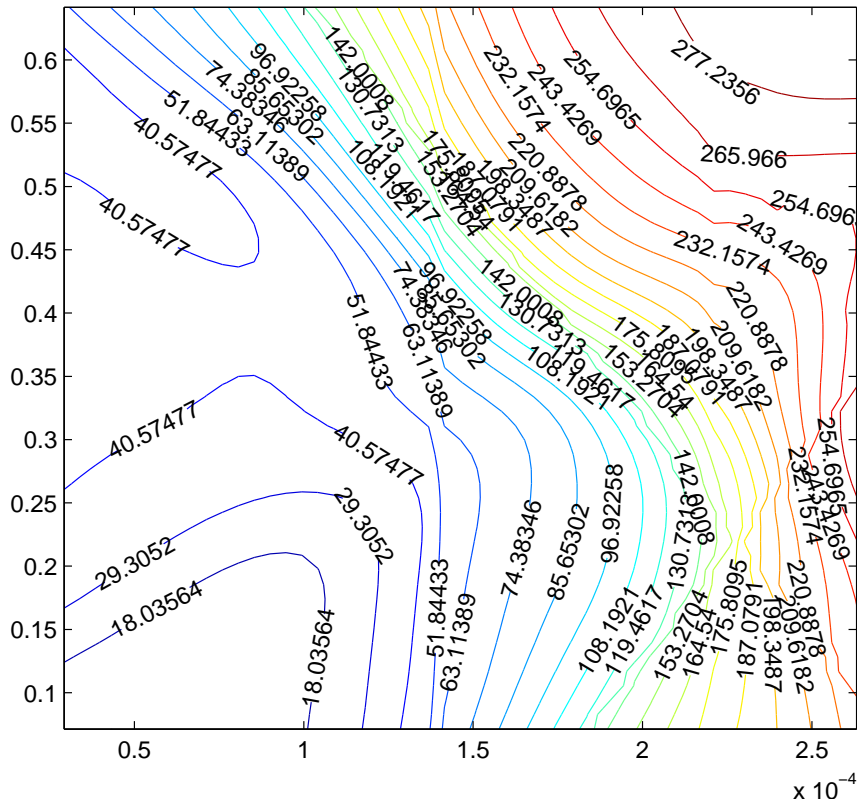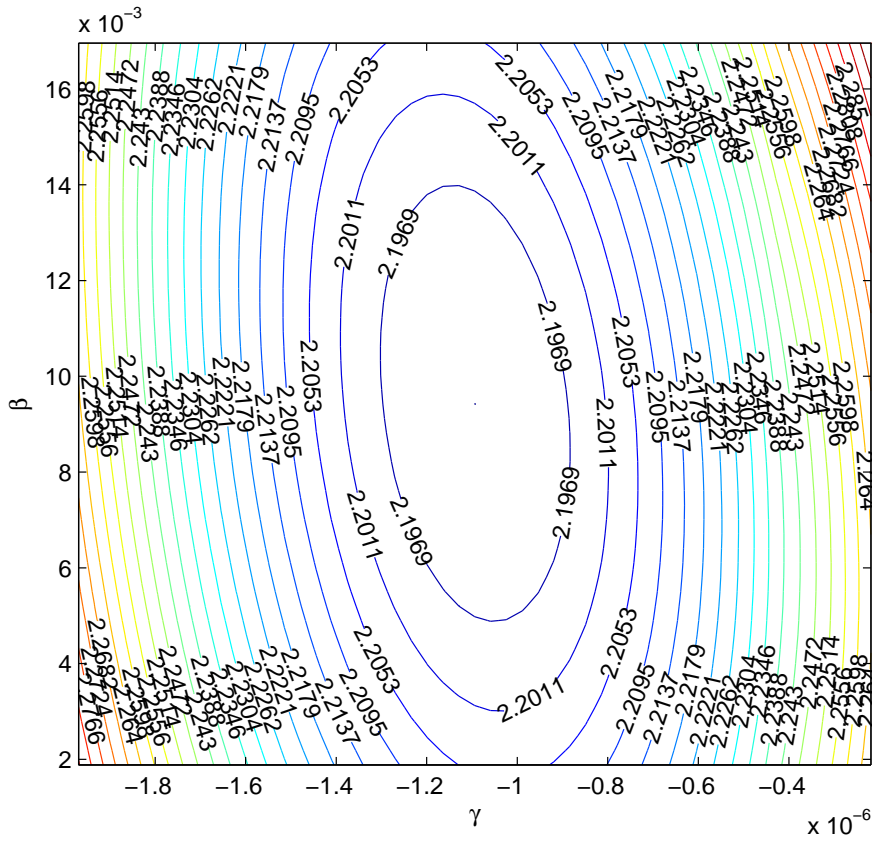
Figure 41: The form of the surface of the -log of the likelihood at the first stop of the minimum search routine with no minimum present.

This track contains the lowest point to the right of figure 37, this is the worst point of all the simulated tracks. It has an impact point $\varepsilon \approx 0.45$ and an $x_{g2} \approx -0.3$ and it is shifted by its position of ≈ -0.75-strip length. The probability distribution has a narrow region that gives a small effective σ (the first plot to right in figure 40). An excellent point for the gaussian approximation, but a bad initialization point for the minimum search routine. No minimum is found at the first run.

33

Figure 42: Second run of the minimum search routine starting from the previous values of γ and β, an evident minimum is present around $\{0, 0\}$.
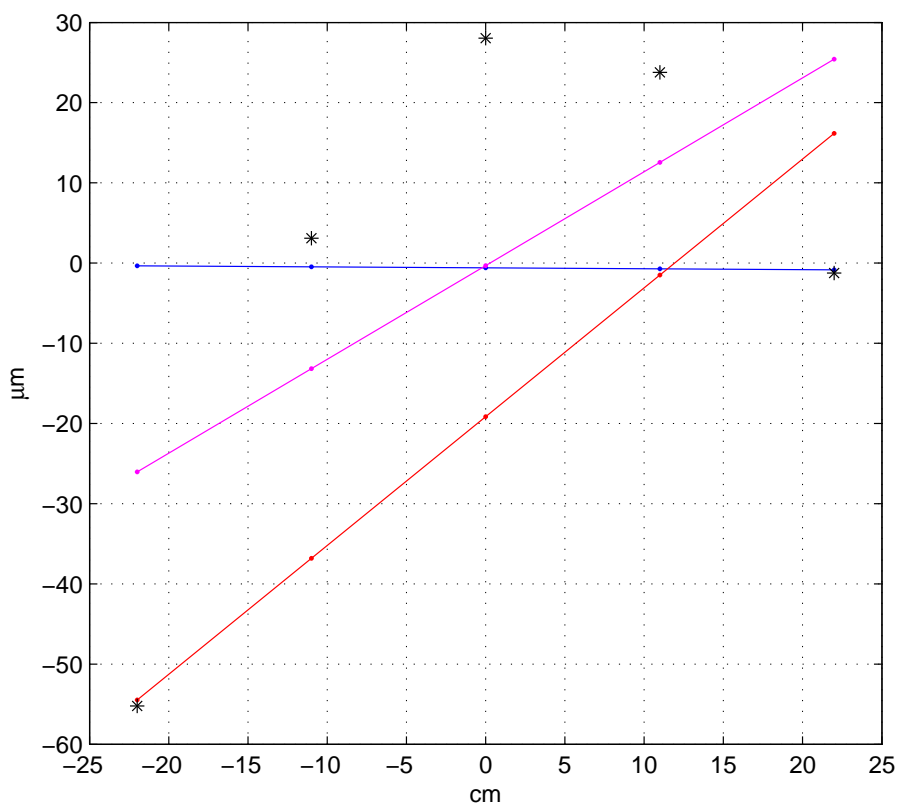
Figure 43: The maximum likelihood reconstructs an excellent track avoiding the dangerous point. The linear gaussian method and the least squares reconstruct worst tracks.
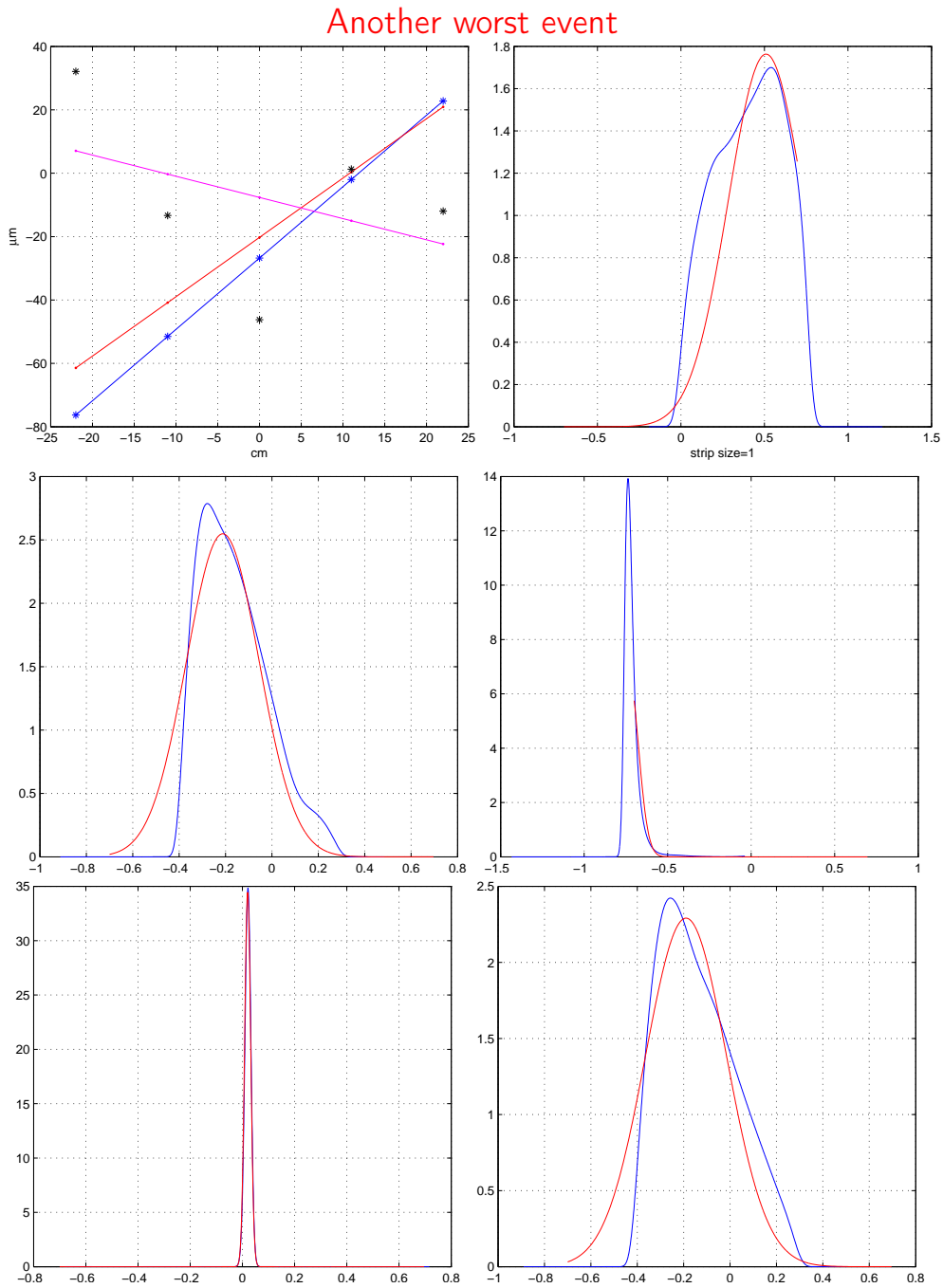
Figure 44: Another worst event (tracks number=5). The first figure reports the reconstructed tracks: the blue line is the the result of the first stop of the minimum search routine, the magenta line is the least squares with identical σ, red line is with effective σs. The other figures are the probability distributions, the blue line is the probability distribution used in the maximum likelihood, the red line the effective gaussian.
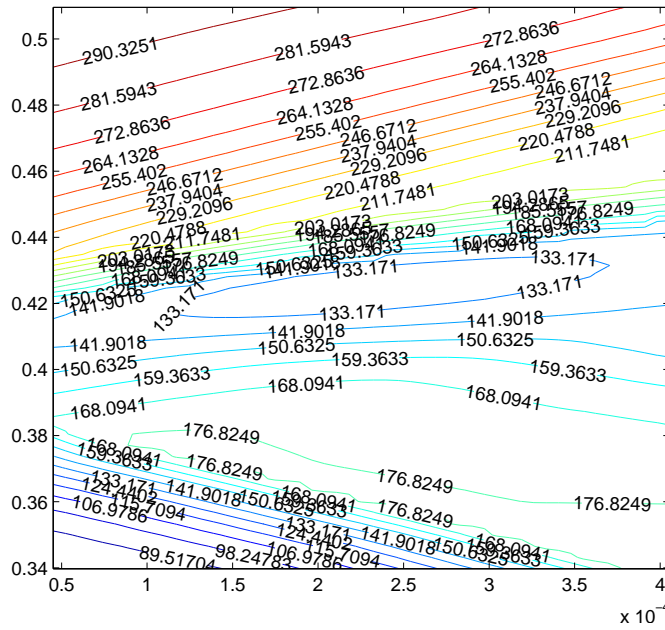
Figure 45: The form of the surface of the -log of the likelihood at the first stop of the minimum search routine with a trap in local minimum.
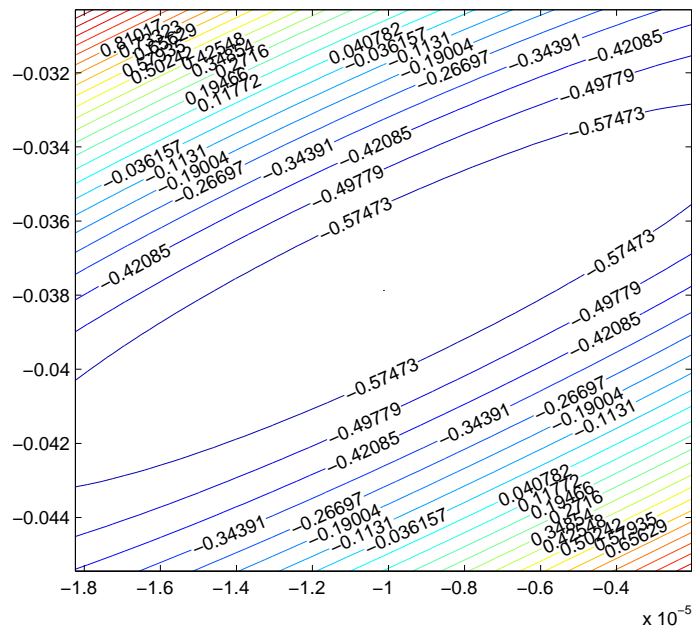

Figure 46: The second iteration starting from the a value 0.2 the first of γ and β, an evident minimum around $\{0, 0\}$ is present.
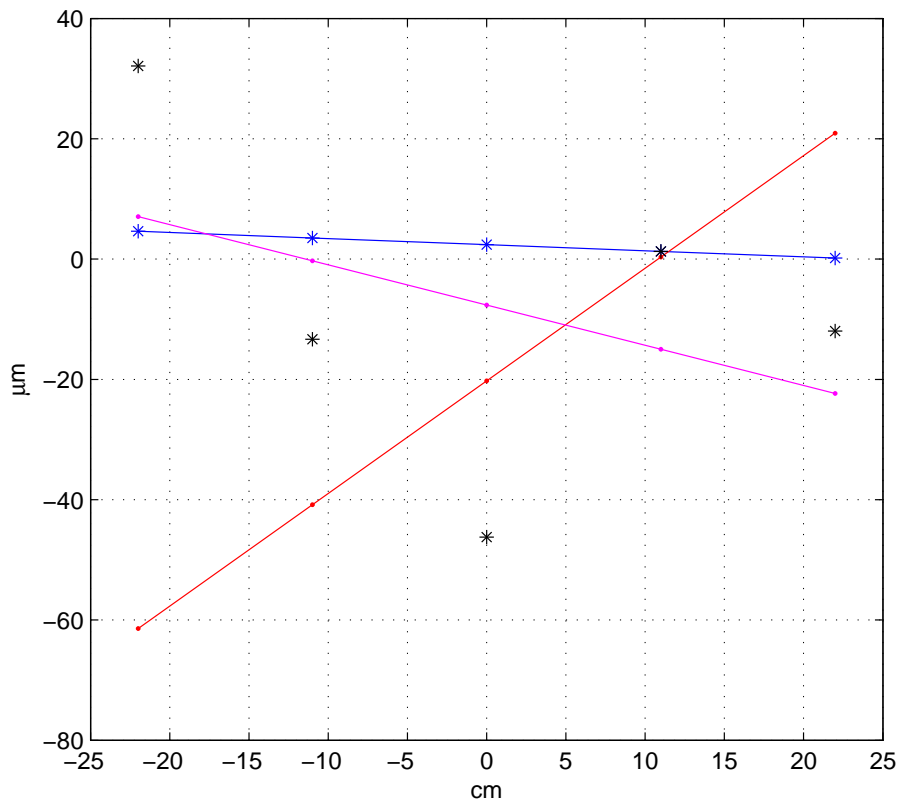
37

The SECOND FINAL result

Figure 47: The maximum likelihood reconstructs a good track that avoids the dangerous point.

CONCLUSIONS

The last two figures show the effectiveness of the maximum likelihood with the proper probability distributions. Few other events similar to these illustrated can be isolated in the simulation. The minima are stable with a reasonable modification of the energy distributions (figure 29).

The minimum search routine is not optimized for this problem.