A NEW ROBUST FITTING ALGORITHM FOR VERTEX RECONSTRUCTION IN THE CERES EXPERIMENT

G. Agakichiev [a], O. Barannikova [b], F. Ceretto [c], U. Faschingbauer [c], P. Glässel [d], E. Kolganova [a], G. Ososkov [a], Yu. Panebratsev [a], J. Rak [c], N. Saveljic [e], I. Tserruya [f], Th. Ullrich [d], J.P. Wurm [c]

[a] *JINR, 141980 Dubna, Russia*

[b] *JINR and Ivanovo State University, Ivanovio, Russia*

[c] *Max-Planck-Institut für Kernphysik, 69117 Heidelberg*

[d] *Universität Heidelberg, 69120 Heidelberg, Germany*

[e] *JINR and University of Montenegro, Podgorica, Yugoslavia*

[f] *Weizmann Institute, Rehovot, Israel*

Submitted to Nucl. Instr. and Meth.

Abstract: In this paper we report on the application of a *robust* approach based on a reweighted least squares method for the vertex reconstruction of the CERES/NA45 experiment at the CERN SPS for the case of 160 GeV/u Pb+Au collisions and compare its performance with the presently used algorithm based on the minimization of summed Gaussian weights.

In the CERES/NA45 experiment a doublet of cylindrical silicon drift detectors (SDD) is used for vertex reconstruction and tracking. The vertex reconstruction algorithm has to cope with a background-contaminated environment. Furthermore, in the case of ultrarelativistic heavy-ion collisions where high event multiplicities occur ($dn_{ch}/d\eta \simeq 500$), the probability of confusing hits from close or intersecting tracks is high. The algorithm needs to provide the precise vertex location even at high occupancies, at a reasonable speed. Our findings are that the robust iterative procedure presented here is about an order of magnitude faster than approaches using standard minimization packages, at a comparable or slightly better accuracy.

1. Introduction

The CERES experiment studies the production of low-mass electron pairs in proton-proton, proton-nucleus, and nucleus-nucleus collisions at the CERN SPS. Vertex and track reconstruction in the experiment is based on the information of two silicon drift detectors SDD-1 and SDD-2[1] situated about 10 cm behind the extended, segmented target (see Fig. 1). They cover the full spectrometer acceptance of 8° to 15° with 2π azimuthal coverage for all target disks.

The specific target used for the 160 GeV/u Pb beam is segmented into 8 individual disks of 600 μm diameter and 25 μm thickness, equidistantly spaced along the beam direction by 3 mm each. This target design allows a larger interaction rate while keeping the photon conversion probability within the spectrometer acceptance low ($X/X_0 = 0.37\%$).

We are dealing here with two sets of hits from each detector. The target and SDD doublet are located in a practically field-free region and the particle trajectories are straight lines connecting the corresponding hits in SDD-1 and SDD-2.

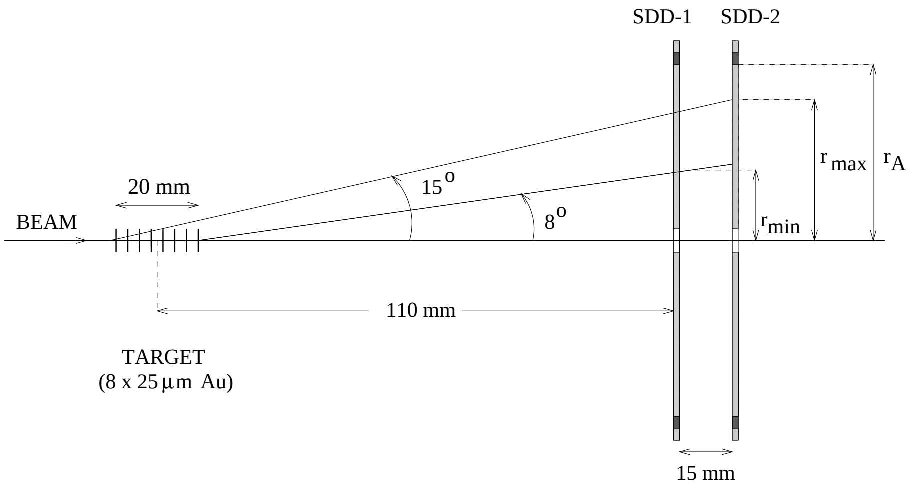

Fig. 1. Geometry of the SDD doublet and segmented target. The acceptance of the CERES spectrometer is indicated.

2. Least squares formulation of the problem

Let (x_{i1}, y_{i1}), $i_1 = 1, \ldots, n_1$ and (x_{i2}, y_{i2}), $i_2 = 1, \ldots, n_2$ be the measured points from SDD-1 and SDD-2, respectively, with some number of background points among them. In this case the conventional least-square method (LSM) for estimating the vertex coordinates x_v, y_v and z_v can be based on minimizing the functions

$$L'(x_v, y_v, z_v) = \sum w_i e_i'^2 \tag{1}$$

$$L''(x_v, y_v, z_v) = \sum w_i e_i''^2 \tag{2}$$

where e_i' and e_i'' are residuals and w_i are the weights assigned. The value e_i' is a measure for the deviation of a SDD-1 hit with coordinates x_{i1}, y_{i1}, z_{i1} from the line which passes through the vertex (x_v, y_v, z_v) and its corresponding SDD-2 hit with coordinates x_{i2}, y_{i2}, z_{i2}, at the position of SDD-1:

$$\begin{aligned} e_i' = & \left(\left\{ x_{i1} - x_{i2} + \frac{z_{i2} - z_{i1}}{z_{i2} - z_v}(x_{i2} - x_v) \right\}^2 + \right. \\ & \left. \left\{ y_{i1} - y_{i2} + \frac{z_{i2} - z_{i1}}{z_{i2} - z_v}(y_{i2} - y_v) \right\}^2 \right)^{\frac{1}{2}} . \end{aligned} \tag{3}$$

Likewise, the value e_i'' is a measure for the deviation of a vertex point x_v, y_v, z_v from the straight line, given by the corresponding hits from SDD-1 and SDD-2 in the z-position z_v of the vertex

$$\begin{aligned} e_i'' = & \left(\left\{ x_v - x_{i1} - \frac{z_v - z_{i1}}{z_{i2} - z_{i1}}(x_{i2} - x_{i1}) \right\}^2 + \right. \\ & \left. \left\{ y_v - y_{i1} - \frac{z_v - z_{i1}}{z_{i2} - z_{i1}}(y_{i2} - y_{i1}) \right\}^2 \right)^{\frac{1}{2}} . \end{aligned} \tag{4}$$

The fundamental LSM assumption is that the residuals, or the deviations from the measured point, are normally distributed. However, this is true only in the case of a clean sample which is not contaminated with background. Such background hits are due e.g., to closely spaced tracks, or to secondary particles not originating from the target. The distribution of residuals including a background fraction ϵ can be approximated as in the gross-error model invented by J. W. Tukey:

$$f(e) = (1 - \epsilon)\phi(e) + \epsilon h(e) \tag{5}$$

with a normal distribution $\phi(e) = \exp(-e^2/2\sigma^2)/\sigma\sqrt{2\pi}$ and a background $h(e)$, which is assumed to be uniform ($h(e) = h_0$ in some interval of the width $\sigma \ll 1/h_0$). The background level ϵ varies considerably depending on the experimental environment. It is evident that in this case the weight of distant background points in the LSM functions (1) and (2) is inappropriate and leads to unnecessarily large errors in the estimated parameters x_v, y_v and z_v. A possibility to cut-off large residuals is to introduce a new parameter and take only residuals smaller than this parameter into account. However, in this case the obtained result would be strongly influenced by the initial values of x_v, y_v, and z_v and the fit possibly ill-behaved due to the fact that the derivatives with respect to the parameters are discontinuous when residuals cross the cut-off between iterations.

3. Summed Gaussian Weights

To avoid the problems mentioned above, the LSM function can be replaced by another one which introduces a smooth cut-off for distant hits. In particular, in [2] it was proposed to minimize the following function

$$L(x_v, y_v, z_v) = -\sum exp[-{e_i}^2/2\sigma^2] \tag{6}$$

which has been used up to now for determination of the vertex for CERES data. A suitably chosen σ is assumed to be constant for all data points. The reason for the choice of function (6) can be illustrated by the expansion of L, which shows the similarity between this estimation and an unweighted least squares method with $w_i = (2\sigma^2)^{-1}$ for small e_i,

$$L(x_v, y_v, z_v) = -n + \sum \frac{{e_i}^2}{2\sigma^2} - \frac{1}{2}\sum \frac{{e_i}^4}{4\sigma^4} + \ldots, \tag{7}$$

while for larger e_i the corresponding summands L_i decrease exponentially, suppressing the influence of strongly deviating hits. As one can see, the second term in the expansion corresponds to the function which follows from the unweighted least square method. Since the residuals for the hits that truly belong to the particle track are normal-distributed the obtained minimum can, to some extent, be interpreted on the basis of a χ^2 distribution. Looking for the minimum of function (6) and differentiating it with respect to each coordinate of the vertex one obtains a system of non-linear equations. The function itself can not be linearized without loosing its properties. Therefore, a traditional function minimizing package had to be used. The initial values for x_v and y_v are set to 0, corresponding to the position of the center of the target disks in the xy-plane. In order to obtain a starting value z_v^0 for the extended target, prior to fitting, a scan is performed by stepping in a $\pm$ 2 cm region (in z) around the center of the target. The found minimum of L defines the starting value for z.

4. Robust Method for Vertex Reconstruction

As was mentioned before we are dealing with a contaminated data set of points in a sense that some points lie far from the track to be reconstructed. Such stray points can spoil the estimates of vertex coordinates x_v, y_v and z_v if their weights w_i are compatible with the weights of useful points. For this case we propose the robust estimation of x_v, y_v and z_v based on the iterative reweighted least square estimation of x_v, y_v and z_v.

4.1. Optimal choice of the weight function in the weighted least square

Since the residuals e_i are non-Gaussian distributed we use a more general approach, i.e. the maximum likelihood (ML) method. An analogous approach was successfully used by P.

Huber [3] and leads to the so-called M-estimates of the parameters in question. However, we carry out our approach in a different way. Keeping in mind that the corresponding ML-functional is strongly non-linear (leading to considerable computing difficulties), we transform the functional partial derivatives in a way which allows to reduce the problem to the optimal choice of the weight function in the weighted least-square sum. The logarithmic likelihood function for measured deviations e_i distributed according to equation (5) is

$$ln \prod_{i=0}^{n} f(e_i) = \sum_{i=0}^{n} ln\Big(\frac{(1-\epsilon)}{\sigma\sqrt{2\pi}} \exp(-\frac{{e_i}^2}{2\sigma^2}) + \epsilon h_0\Big). \tag{8}$$

After differentiating eq. (8) with respect to x_v, y_v and z_v and dividing by

$$(1 - \epsilon)\sigma^{-1}(2\pi)^{-\frac{1}{2}} \exp(-{e_i}^2/2\sigma^2),$$

we obtain a conventional LSM system of normal equations

$$\begin{cases} \sum w_i \frac{1}{\sigma^2} e_i \frac{\partial e_i}{\partial x_v} = 0 \\ \sum w_i \frac{1}{\sigma^2} e_i \frac{\partial e_i}{\partial y_v} = 0 \\ \sum w_i \frac{1}{\sigma^2} e_i \frac{\partial e_i}{\partial z_v} = 0 \end{cases} \tag{9}$$

with the optimal weight function

$$w_{opt}(t) = \frac{1+c}{1 + c \cdot \exp\left(\frac{t^2}{2}\right)}, \tag{10}$$

where t denotes the ratio e/σ. The only parameter

$$c = (1 - \epsilon)^{-1} \epsilon \sigma h_0 \sqrt{2\pi}$$

is the ratio of the mean number of noise hits to the mean number of useful hits within an interval $\sigma\sqrt{2\pi}$. Thus it is determined by the contamination of data – not in the whole range of the sample but within its essential part where all useful observations are concentrated. As in our case, the value of c is often roughly known in experimental models. The factor $(1 + c)$ is introduced to fulfill $w(0) = 1$. The approximation of $w_{opt}(t)$ by a fourth order polynomial leads to the famous Tukey's bi-squared weights [4]

$$w(t) = \begin{cases} \left(1 - (t/c_T)^2\right)^2 \text{ if } |t| < c_T \\ \qquad 0 \quad \text{otherwise} \end{cases} \tag{11}$$

Here, with c_T we denote Tukey's constant.

4.2. Calculation of vertex coordinates with robust approach

To find the vertex coordinates x_v, y_v and z_v we need to solve the LSM system (9) for either of the residuals e'_i or e''_i given by eq. (3) and (4), respectively.

Dealing with function $L'(x_v, y_v, z_v)$, i.e., with e'_i residuals, we get a system of equations which is not simple to solve because they contain a term $(z_{i2} - z_v)$ in the denominator. To avoid this problem, function $L'(x_v, y_v, z_v)$ can be linearized by multiplying it with an approximately constant term $(z_{i2} - z_v)$, so we can deal with the function

$$\widetilde{L'}(x_v, y_v, z_v) = \sum w_i \widetilde{e'}_i^2 \tag{12}$$

where $\widetilde{e'}_i = (z_{i2} - z_v)e'_i$.

In this case, evaluating e_i from equation system (9) with $\widetilde{e'}_i$, one has to solve the corresponding system of linear equations:

$$\begin{cases} A'x_v + B'z_v & = C' \\ A'y_v + E'z_v & = F' \\ B'x_v + E'y_v + G'z_v & = H', \end{cases} \tag{13}$$

where we denote

$$\begin{aligned} A' &= \sum w_i(z_{i2} - z_{i1})^2; \; B' = \sum w_i(x_{i1} - x_{i2})(z_{i2} - z_{i1}) \\ C' &= \sum w_i[z_{i2}(x_{i1} - x_{i2})(z_{i2} - z_{i1}) + x_{i2}(z_{i2} - z_{i1})^2] \\ E' &= \sum w_i(y_{i1} - y_{i2})(z_{i2} - z_{i1}) \\ F' &= \sum w_i[z_{i2}(y_{i1} - y_{i2})(z_{i2} - z_{i1}) + y_{i2}(z_{i2} - z_{i1})^2] \\ G' &= \sum w_i[(x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2] \\ H' &= \sum w_i[z_{i2}(x_{i1} - x_{i2})^2 + x_{i2}(z_{i2} - z_{i1})(x_{i1} - x_{i2}) + z_{i2}(y_{i1} - y_{i2})^2 + \\ &\quad y_{i2}(z_{i2} - z_{i1})(y_{i1} - y_{i2}). \end{aligned}$$

From the equation system (13) we have

$$\begin{cases} x_v = (C' - B'z_v)/A'; \\ y_v = (F' - E'z_v)/A'; \end{cases} \tag{14}$$

where

$$z_v = (A'H' - E'F' - B'C')/(A'G' - B'^2 - E'^2).$$

Dealing with e''_i residuals we go straight to the solution. In this case one has to solve the following system of linear equations

$$\begin{cases} A''x_v + B''z_v = P'' \\ A''y_v + D''z_v = Q'' \\ B''x_v + D''y_v + F''z_v = R'', \end{cases} \tag{15}$$

where we denote

$$A'' = \sum w_i;\ B'' = \sum w_i \frac{x_{i1} - x_{i2}}{z_{i2} - z_1};\ C'' = \sum w_i x_{i1};$$

$$D'' = \sum w_i \frac{y_{i1} - y_{i2}}{z_{i2} - z_1};\ E'' = \sum w_i y_{i1};$$

$$F'' = \sum w_i \left[\left(\frac{x_{i2} - x_{i1}}{z_{i2} - z_1} \right)^2 + \left(\frac{y_{i2} - y_{i1}}{z_{i2} - z_1} \right)^2 \right];$$

$$G'' = \sum w_i \left[x_{i1} \left(\frac{x_{i1} - x_{i2}}{z_{i2} - z_1} \right) + y_{i1} \left(\frac{y_{i1} - y_{i2}}{z_{i2} - z_1} \right) \right];$$

$$P'' = z_1 B'' + C'';\quad Q'' = z_1 D'' + E'';\quad R'' = z_1 F'' + G''.$$

With z_1 we denote z-position of SDD-1.

From the equation system (15) we have

$$\begin{cases} x_v = [B''(z_1 - z_v) + C'']/A''; \\ y_v = [D''(z_1 - z_v) + E'']/A''; \end{cases} \tag{16}$$

where

$$z_v = z_1 + (A''G'' - D''E'' - B''C'')/(A''F'' - B''^2 - D''^2).$$

The weights in the above expressions are computed iteratively using Tukey's weight formula

$$w_i^k = \begin{cases} (1 - (e_i^{(k)}/(c_T\hat{\sigma}^{(k-1)}))^2)^2 \text{ if } |e_i^{(k)}| \le c_T\hat{\sigma}^{(k-1)} \\ 0 \quad \text{otherwise} \end{cases}$$

where e_i^k is the residual of either of the deviations e_i' or $\widetilde{e''}_i$ obtained in iteration k, and $\hat{\sigma}^{(k-1)}$ is the estimate of variance which is evaluated as [5]

$$\hat{\sigma}^{(k)2} = \sum w_i^{(k)} (e_i^{(k)})^2 / \sum w_i^{(k)}.$$

For our calculations we have varied the constant c_T and obtained the best resolution of the vertex coordinates for $c_T \simeq 3$. Instead of scanning the 8 targets to determine the initial parameters, as it was done for the Gaussian summed weights (SGW), the center of the target region with $x_v{}^{(0)} = y_v{}^{(0)} = 0$ and $z_v{}^{(0)}$ was used as the starting value for the first iteration, saving considerable CPU time.

5. Calculation results

In this section we compare the results for the vertex reconstruction obtained with the Summed Gaussian Weights (SGW) approach and the robust approach. The underlying sample consists of 4000 Pb + Au events. In the following, the results for the SGW method were obtained by using e_i' residuals in function (6), the results for the robust approach were obtained by using e_i'' residuals in equation system (9). It should be noted that the usage of $\widetilde{e_i'}$ residuals leads essentially to the same results.

Fig. 2 and 3 show the vertex z-coordinate distribution for the case of the SGW and the robust method, respectively. As one can see from the resulting histograms, both distributions nicely reflect the target region. Each of the disks is clearly seen as a peak in the distribution. All peaks have Gaussian form, which is illustrated by the fitted Gaussians.

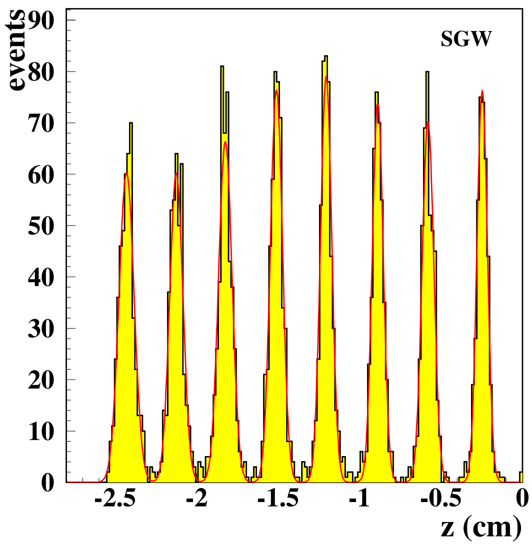

Fig. 2. Reconstructed z-coordinates of the vertices fitted with eight Gaussians corresponding to eight target disks obtained with the SGW approach

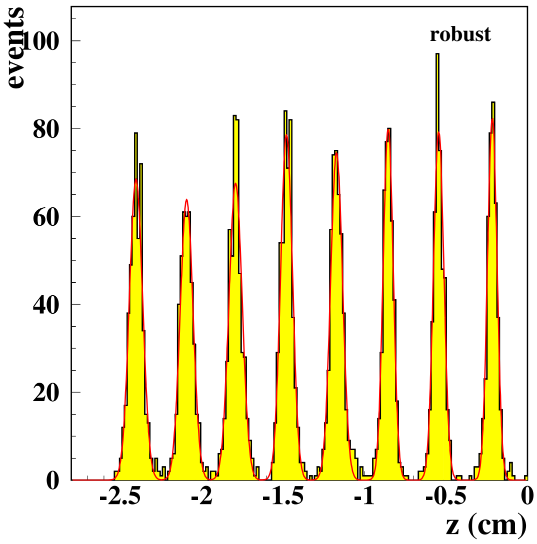

Fig. 3. Reconstructed z-coordinate of the vertices fitted with eight Gaussians corresponding to eight target disks obtained with the robust approach

The resolutions of the target disks obtained by fitting each of the peaks individually are shown in Table 1 for both cases. The robust approach gives a slightly better resolution for each disk. This was confirmed by tests of both methods with Monte-Carlo generated data, which also resulted in a better resolution of all vertex coordinates.

The distributions of the x and y-coordinates of the vertex shown in Fig. 4 basically reflect the lateral size of the beam spot on the target. Here, the results of the two methods are practically indistinguishable both concerning the mean and the width of the obtained distributions.

In practice, a reasonable accuracy of the geometric position of the vertex, obtained by the robust approach, is already achieved after several iterations. Fig. 5 shows the distribution of number of iterations per event for the robust weights approach. As one can see, 5 iterations on average are enough to find the minimum. Together with the uncritical choice

Table 1
z-resolutions obtained with SGW and robust fit

disk No.	rms (SGW) [μm]	rms (robust) [μm]
1	307	264
2	322	287
3	361	289
4	327	329
5	386	344
6	396	337
7	409	372
8	461	395

of the starting values, this leads to a gain in speed by about an order of magnitude for the robust method as compared to standard general purpose packages for minimization (for example MINUIT[6]).

The local track accuracy of the silicon drift chambers, i.e. the radial and azimuthal residuals in SDD-1 for tracks defined by the interaction vertex and a hit in SDD-2, are shown in Fig. 6 for both algorithms. As it can be seen from the figure, both approaches lead to almost the same distributions of the Δr and $\Delta\varphi$ residuals ($\sigma_\varphi \simeq 6$ mrad, $\sigma_r \simeq 100$ μm). The track resolution results from the combined effect of the intrinsic resolution of the chambers, the vertex resolution and multiple scattering.

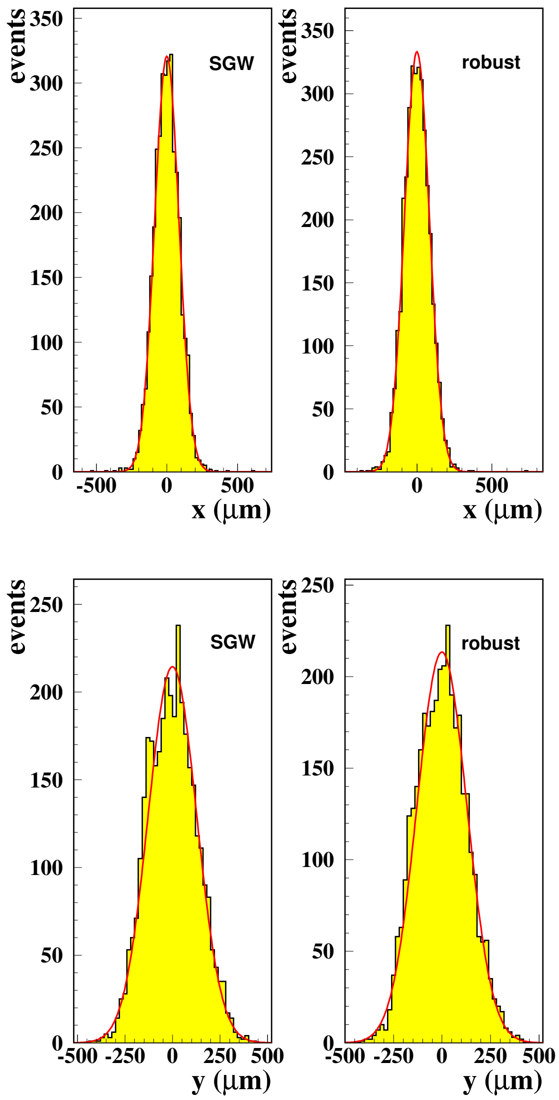

Fig. 4. Reconstructed x and y-coordinates of the vertices obtained with the SGW and the robust approach, fitted by Gaussians

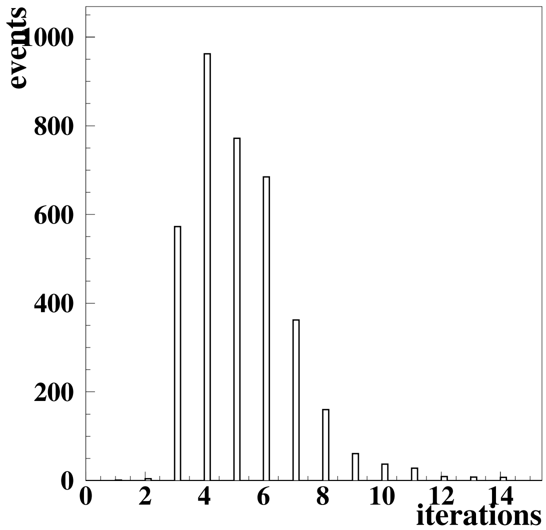

Fig. 5. Number of iterations per event for the robust approach

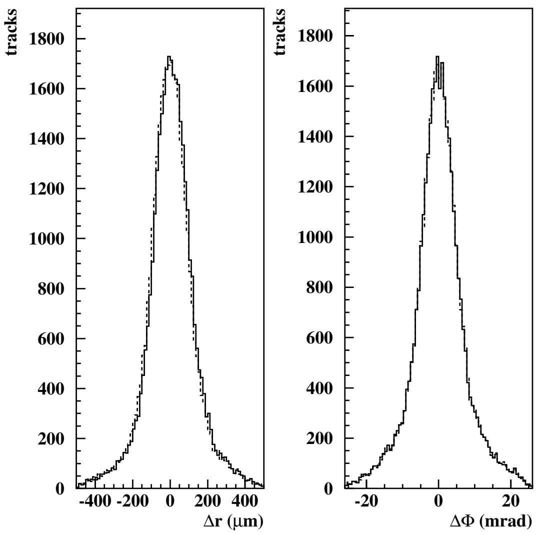

Fig. 6. Local track accuracy of the silicon drift detectors: radial and azimuthal residuals in SDD-1 for tracks defined by the vertex and a hit in SDD-2 (SGW: solid line, robust: dashed line)

6. Conclusions

We have presented results on vertex reconstruction for CERES data obtained with the SGW approach and the robust approach. Both algorithms give good results, clearly resolving the target region profile. The vertex xy-coordinate resolutions obtained by both methods are almost the same, the z-resolution is somewhat better for the robust method.

The advantage of the robust approach, as an iterative method, is its insensitivity on the choice of initial values for the parameters in question. Starting from the center of the segmented target the algorithm converges to the right position after $\sim$5 iterations. The robust fitting approach allows to reconstruct the vertex coordinates without using standard general purpose packages for minimization. All in all, this results in a considerable increase in speed, a very important factor for the time consuming mass-production stage of the analysis of huge data samples.

References

[1] U. Faschingbauer et al., Nucl. Instr. and Meth. A 377 (1996) 362.

[2] G. Agakichiev et al., Nucl. Instr. and Meth. A 371 (1996) 243.

[3] P. Huber, Robust Statistics, J. Willey & Sons, NY (1981).

[4] F. Mosteller, W. Tukey, Data analysis and regression: a second course in statistics, Addison - Wesley (1987).

[5] G. A. Ososkov, Proc. Second International Tampere Conference in Statistics (Tampere, Finland 1987).

[6] MINUIT - Function Minimization and Error Analysis, CERN Program Library entry D506