CLAS–NOTE 92–006

Pattern Recognition in CLAS using Minimum Spanning Tree Method

C. L. Tam

Supercomputer Computation Research Institution, Tallahassee, Florida 32306

Abstract

Particle track pattern recognition in inhomogeneous magnetic field is a difficult task in designing the CLAS software. And an efficient solution of this problem is required for all experiments using this spectrometer. We found that an algorithm based on the minimum spanning tree method can be especially suitable for this task. First, a simple tree algorithm is used to find the clusters of drift chamber hits, resolve the left-right ambiguity, and reconstruct track segments in each of the superlayers. Track segments from different regions of the magnetic field are then linked to form complete particle tracks using a modified minimum spanning tree algorithm, while track segments having uncorrelated directions (Θ, ϕ) can be quickly eliminated. The rigidity of the linked track can then be estimated based on the deflection angle of the track segment in the outer region of the detector. This algorithm has been tested using data from test run of prototype drift chamber and from Monte Carlo simulation. It is found to be efficient in finding tracks in the CLAS detector geometry.

1. Introduction

The CEBAF Large Acceptance Spectrometer (CLAS) consist of a toroidal magnet with large volume of drift chambers situated in the inhomogeneous magnetic field (Fig. 1a). The presence of the inhomogeneous magnetic field is a challenge to pattern recognition and track reconstruction with this detector.

By taking advantage of the CLAS detector design, where the drift chambers are arranged in 6 narrow superlayers in 3 regions of the detector detector volume (Fig. 1b), the reconstruction of track segments in each superlayer can be carried out independent of the global non-uniformity of the magnetic field. Thereby reducing the task of global particle track pattern recognition in the non-uniform **B** field to the linking of track segments from the six superlayers.

The presence of the inhomogeneous **B** field results in complex particle trajectories which are not easily described by simple quadratic functions. One traditional approach in recognizing this class of patterns is the Template Matching Method[1]. By compiling a large dictionary of all the possible tracks patterns ("templates"), candidates of possible tracks can be compared to the "templates" in the dictionary, and a track pattern is "recognized" when a "matching template" is found. This method has been used successfully in several high energy experiments[2][3].

The speed of the search for a "matching template" depends critically on the number of the templates in the dictionary, therefore, a suitable dictionary must contain only the minimum number of templates needed to represent all the anticipated track patterns. Due to the inhomogeneity of the magnetic field in CLAS, coupled with the large dynamic range of particles it will accept, a very large number of templates ($10^5 - 10^6$) will be needed. Hence the search for the matching template will be very time consuming. And the large amount of storage required further complicates the efficient implementation of this method

In this paper, a different approach to the CLAS track pattern recognition based on the principle of the minimum spanning tree method is presented. It does not require the large dictionary necessary in the Template Matching Method. The decision making in general will be much faster. It is therefore particularly suited for on-line data analysis.

2. The Application of Minimum Spanning Tree Algorithm in CLAS

2.1 Track Segment Reconstruction within Superlayers

Figure 2a shows a typical particle track traversing a superlayer, and a few uncorrelated background hits in the area. The major tasks for the track segment reconstruction within superlayers are:

1. Finding clusters of drift chamber hits corresponding to track segments,
2. Resolving the left-right ambiguity of the hits,
3. Determining the angle dependent electron drift velocities for each hit and performing χ^2 fits

Resolving of left–right ambiguities is a common problem in operation of drift chambers, and many algorithms has been developed for this task. Figure 2b shows the pairs of mirror hits associated with each of the sense wire fired, and a few examples of the different left-right combinations. For a drift chamber superlayer consisting N layers of sense wires, there are 2^N permutation of combinations. Due to the relatively high momentum of the particles of interest, their trajectories can be safely approximated by straight lines within the narrow span of a superlayer. So, in the most simplistic approach, the track segment within the superlayer can be found by fitting all 2^N permutations to straight line in order to find the combination having the smallest χ^2. But a much faster method using the minimum spanning tree approach can be used for resolving the left-right ambiguity.

This approach was first used in the TASSO detector at PETRA[4]. The method was modified and expanded for use with a CLAS geometry prototype drift chamber. As shown in the example in figure 2b, each of the two mirror hits in a sense wire is linked to other hits in the neighboring layers by straight **branches**. The **branch weight** associated with branch i is defined as the angle of the branch, θ_i. The periodical nature of the CLAS drift chamber cell arrangement enables a very fast calculation of the branch weight: $\theta_i = \frac{1}{h}(s + (-1)^{\pm 1} d_i + (-1)^{\pm 1} d_{i+1})$. A combination of branches forming a path traversing the height of the superlayer is called a **full tree**. There are 2^N full trees, corresponding to the 2^N permutations of left-right combinations. It is apparent from Fig. 2b that a straight line track segment corresponds to a combination of connecting branches forming a nearly straight line (a **minimum spanning tree**). *Therefore, trees consisting of branches having the smallest differences in their branch weights will be accepted, while combinations of branches having widely varying branch weights can be eliminated.*

We also found that a simplified application of this algorithm can be well suited for the task of eliminating background hits from the clusters of hits produced by track segments. In the first pass, drift times of the hits are not considered. Only the sense wire positions are linked to form trees and to calculate branch weights ($\theta_i = \frac{s}{h}$). As illustrated in Fig. 2a, uncorrelated background hits can then be easily rejected due to their anomalous branch weights. Correlated background hits, however, will be more efficiently identified and eliminated during the later fitting stage by their large χ^2s.

This algorithm has been used on the data from a test run carried out at Brookhaven National Laboratory using the prototype drift chamber operating in **B** field of up to 1.5 Tesla. The algorithm is found to be successful in resolving the left-right ambiguity for over 85% of the hits, while hits very close (less than 1mm) to the sense wire are found to be better resolved by χ^2 fitting.

In addition, the average branch weight of a tree (Θ_{ave}) can be used as a good estimate of the track segment angle Θ_{trk} (Fig. 3). This angle is especially important in determining the accurate electron drift velocity in the high magnetic field, due to the strong angular dependency of eletron drift velocity in B field[5]. Therefore, to reconstruct the track segment, only a small fraction ($\sim 15\%$) of the 2^N permutations need to be fitted to a straight line by least square fits. This is found to be significantly faster than the naive method which fits all of the 2^N permutations, and without any loss in track resolution. In our prototype chamber operating in magnetic a field of up to 1.5 Telsa, resolutions of better than $200\mu m$ were achieved.

2.2 Global Track Finding in Inhomogeneous Magnetic Field

One important charateristic of a toroidal magnet, its ϕ preserving property, can be incorporated to greatly ease the problem of global track pattern recognition. Since only track segments having similar polar angles ϕ can be possible candidates of a complete particle track. Selection of candidates based on their ϕ angle drastically reduces the track segment density.

After track segments in all superlayers are reconstructed, principle of the minimum spanning tree method can be extended to the global track finding. Figure 4 is an example of multiple tracks traversing one sector of the detector. Fig. 4a shows the track segments detected in the drift chamber, including some background track segments, and Fig. 4b shows some examples of the **full trees** formed by the linking of arbitrary combinations of track segments from each superlayers. In this global picture, both the track segments (solid lines) and their connecting links (dashes) are considered **branches**, and a **branch weight** is the deflection ($\delta\Theta_i = \Theta_i - \Theta_{i-1}$) each branch suffered due to the magnetic field (Fig. 4b).

As one example of the role of the branch weight $\delta\Theta$, we observe that all branch weights $\delta\Theta$ of an (undecayed) particle track in zero magnetic field are zero! It is also apparent from figure 5a that since the cumulative effect of the magnetic field is to deflect the particle further and further away from the radial direction as the particle travels toward the outer regions of the detector, *trees having branch weights $\delta\Theta$ of alternating signs can be rejected.*

In a uniform magnetic field, it can easily be shown that for tracks of high energy particles (where multiple scattering effects can be neglected), the branch weights of a track segment i obey the following relationship: $\delta\Theta_i = \delta\Theta_{i-1}$. Extensive simulation of trajectories in the CLAS toroidal **B** field found this relationship remains a good approximation despite the inhomogeneity of the **B** field. Fig. 5 shows the distribution of $\delta\Theta_i - \delta\Theta_{i-1}$ for various $\Theta_1 > 30^0$ and particles with momentum $p > 0.5GeV$. Very low momentum particles and particles at extreme forward angles, have a slightly broader distribution. *By accepting only branches whose branch weights $\delta\Theta$ falls within $\pm 10^0$ of this relationship (*"**weighted" minimum spanning trees***), we are able to correctly identify track segments from trajectories of particles having momentum greater than 0.5 GeV/c and Θ_1 beyond* 30^0. Since the track segment density expected in any given ϕ range is typically fairly low, the application of the tree algorithm is found to be very efficient in rejecting uncorrelated track segments. As an example, the track segments in Fig. 4a were successfully linked into 4 complete tracks as shown in Fig. 5b, while background track segments can be safely eliminated.

Furthermore, for track segments in the outer regions of the drift chamber, their angle of deflection from the radial direction $(\Theta_i - \Theta_1)$ (Fig. 5a) is strongly correlated to the rigidity of the trajectory (Fig. 7). These correlation relationships can be parametized as functions of momentum, Θ and ϕ etc. Rough estimate of the track parameters $(p, \Theta, \phi...)$ can then be obtained using these correlation functions.

3. Summary

We demonstrated an approach for fast track recognition in the CLAS inhomogeneous **B** field using the minimum spanning tree method.

1. When reconstructing track segments within superlayers, we found that:
   (a) Using only hit wire positions, the **Minimum Spanning Trees** formed can be used to identify clusters of hits.
   (b) By including drift times in calculating branch weights, the left-right ambiguity can be resolved by the selection of the **Minimum Spanning Trees** for more than 85% of the hits.
   (c) The average branch weight Θ_{ave} can serve as estimate of Θ_{trk}. This allows a more accurate determination of the angle dependent electron drift velocity. Final χ^2 fits provides for a resolution of the remaining left-right ambiguities and a rapid elimination of correlated backgrounds.
2. For global pattern recognition of complete tracks, we can:
   (a) Take advantage of the ϕ preserving property of the CLAS toroidal **B** field to seperate track segments into different ϕ ranges. Within each ϕ region, only trees whose branch weights $\delta\Theta$ have the same sign will be accepted.

(b) Select the "**Weighted**" **Minimum Spanning Trees** to link candidate track segments together to form complete tracks. Background track segments having uncorrelated branch weights can be rejected.

(c) Obtain an estimate of the rigidity of the completed track from the deflection angles of track segments in the outer regions of the detector.

The minimum spanning tree algorithm is very efficient in background rejection and fast pattern recognition for the anticipated track density of the CLAS detector. Certain elements of this approach have been incorporated into the 2nd level (hardware) triggering algorithm. The algorithm is also quite suitable for a high-level software trigger. For on-line or off-line data analysis, it can be combined with other sophisticated fitting or interpolation algorithms to achieve the desired accuracy in the reconstruction of complete particle tracks in the CLAS detector[5][6].

References

[1] Grote *Rep. Prog. Phys.* 50 473 (1987)

[2] Behrend *et al. Comput. Phys. Commun.* **22** 365-74 (1986)

[3] Georgiopoulos et. al.*et al., Nucl. Instrum. Methods* **A249** 451-4 (1986)

[4] Cassel et. al.*et al., Nucl. Instrum. Methods* **A185** 235-51 (1981)

[5] Mestayer et. al. Submited to *IEEE Tranc. in Nucl. Science*

[6] B.B. Niczyporuk, *CLAS-NOTE 91-001*

[7] TAM et. al. to be submited as CLAS-NOTE

Figure Captions

Fig. 1 a) CLAS toroidal magnetic spectrometer. b) Mid-plane view of CLAS.

Fig. 2 a) Typical track segment in a superlayer in region 2 of the drift chamber. The distance between sense wires of the same layer is $s = 1.732cm$, the spacing between layer is $h = 1.5cm$. b) Example of some random left-right combinations. Also the definition of **braches** and **branch weights**. Note the solid line is a **minimum spanning tree**.

Fig. 3 Resolution of $\Theta_{ave} - \Theta_{trk}$

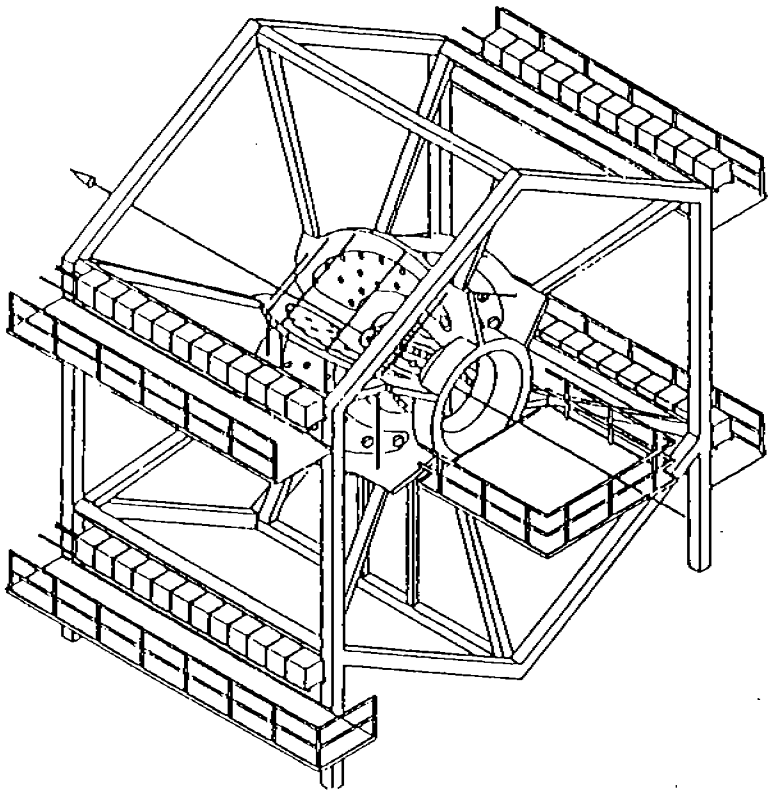
Fig. 4 a) (Top) Example of multiple track segments in one section of the detector. b) (Bottom) Some randomly formed **trees**. In the global picture, both track segments (solid) and their connecting links (dash) are considered **branches**. Note examples of their **branch weights**.

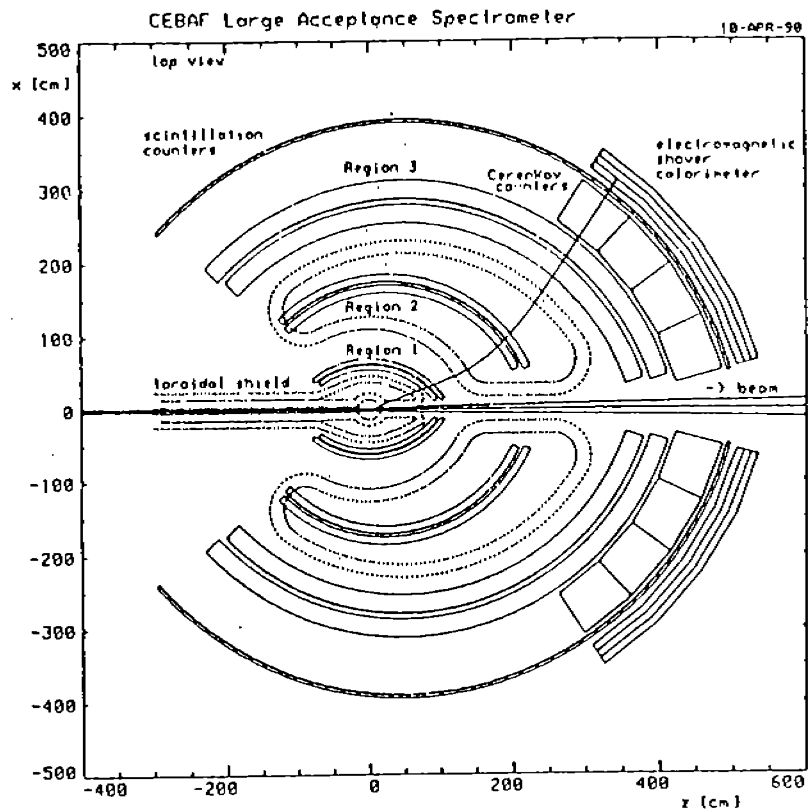
Fig. 5 a) (Top) Typical deflection of tracks in magnetic field. b) (Bottom) Complete tracks reconstructed from the track segments in figure 4a.

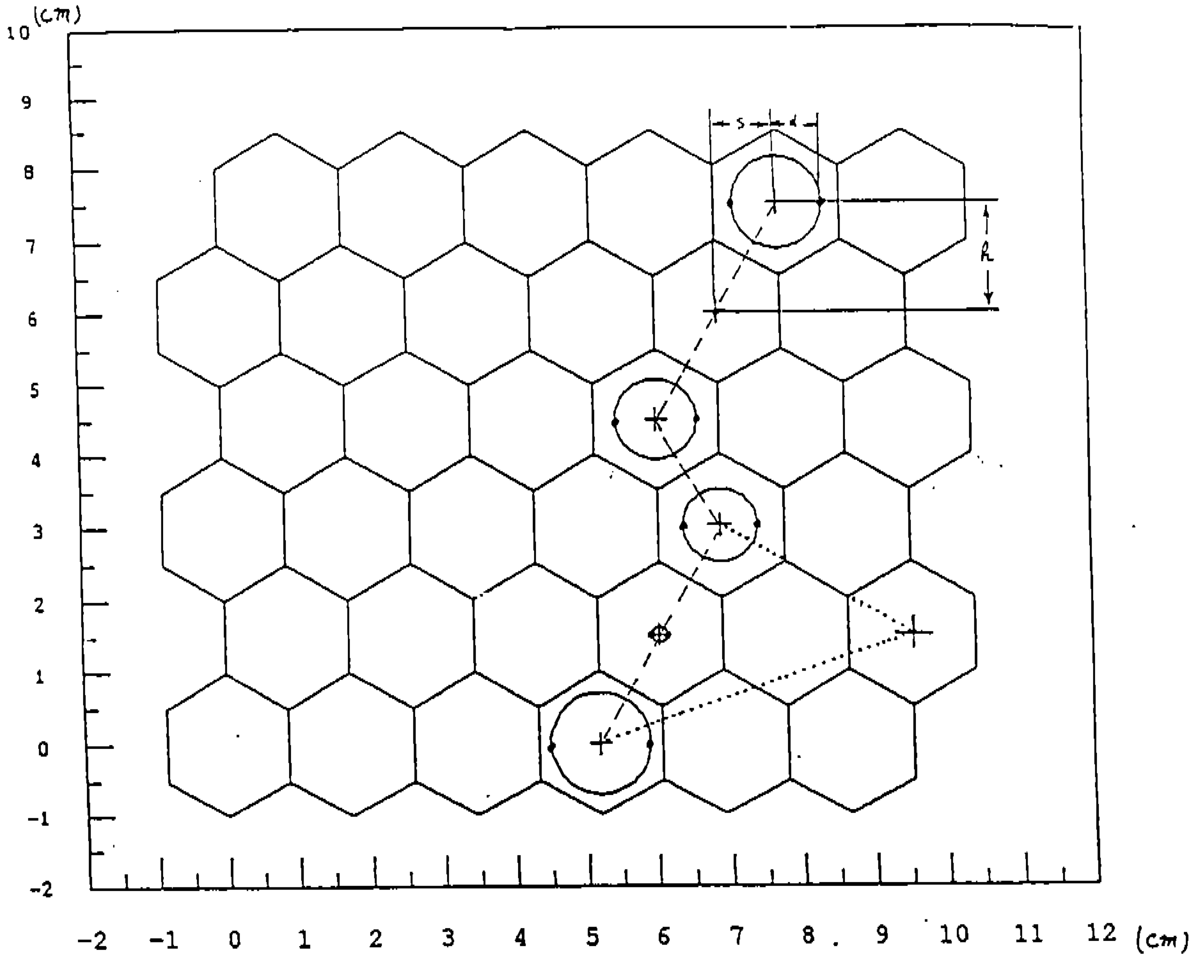
Fig. 6 Distribution of $\delta\Theta_i - \delta\Theta_{i-1}$ for trajectories of $0.5 < p < 4$ and $30 < \Theta_1 < 140$.

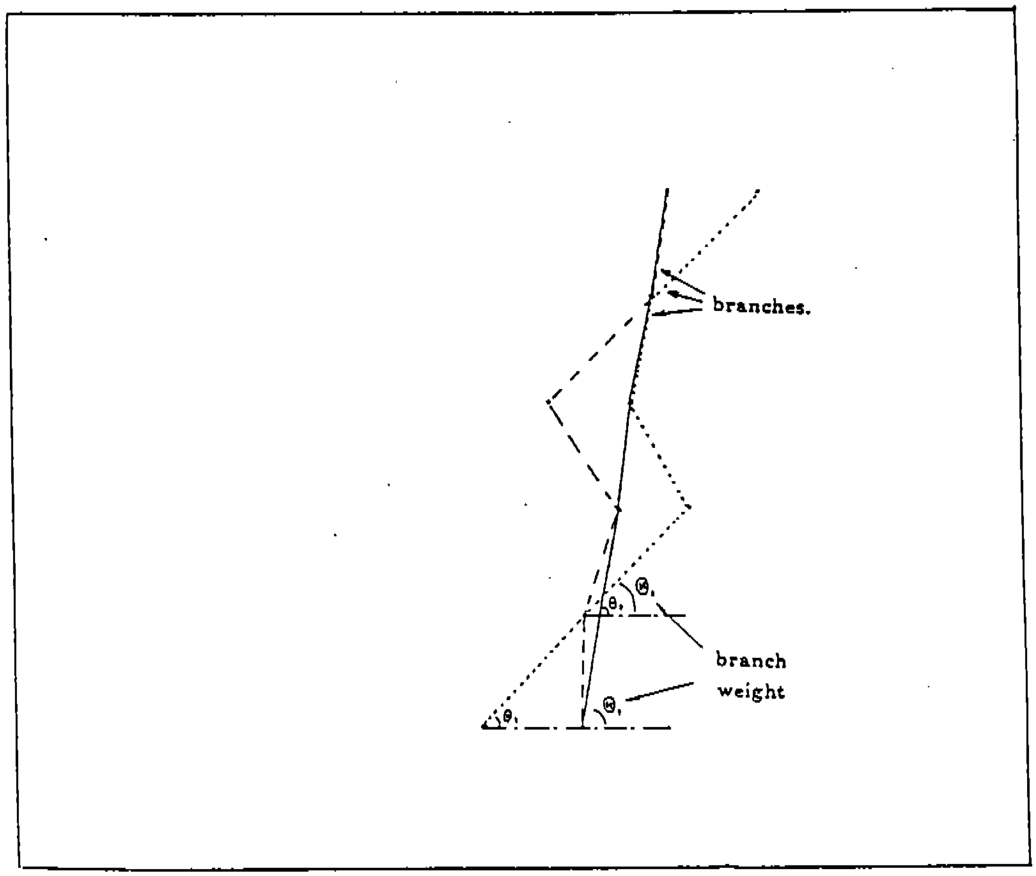
Fig. 7 Deflection angle of track segments in a) region 2, b) region 3 of drift chamber as a function of momentum at different Θ_1 for a 1GeV/c eletron.

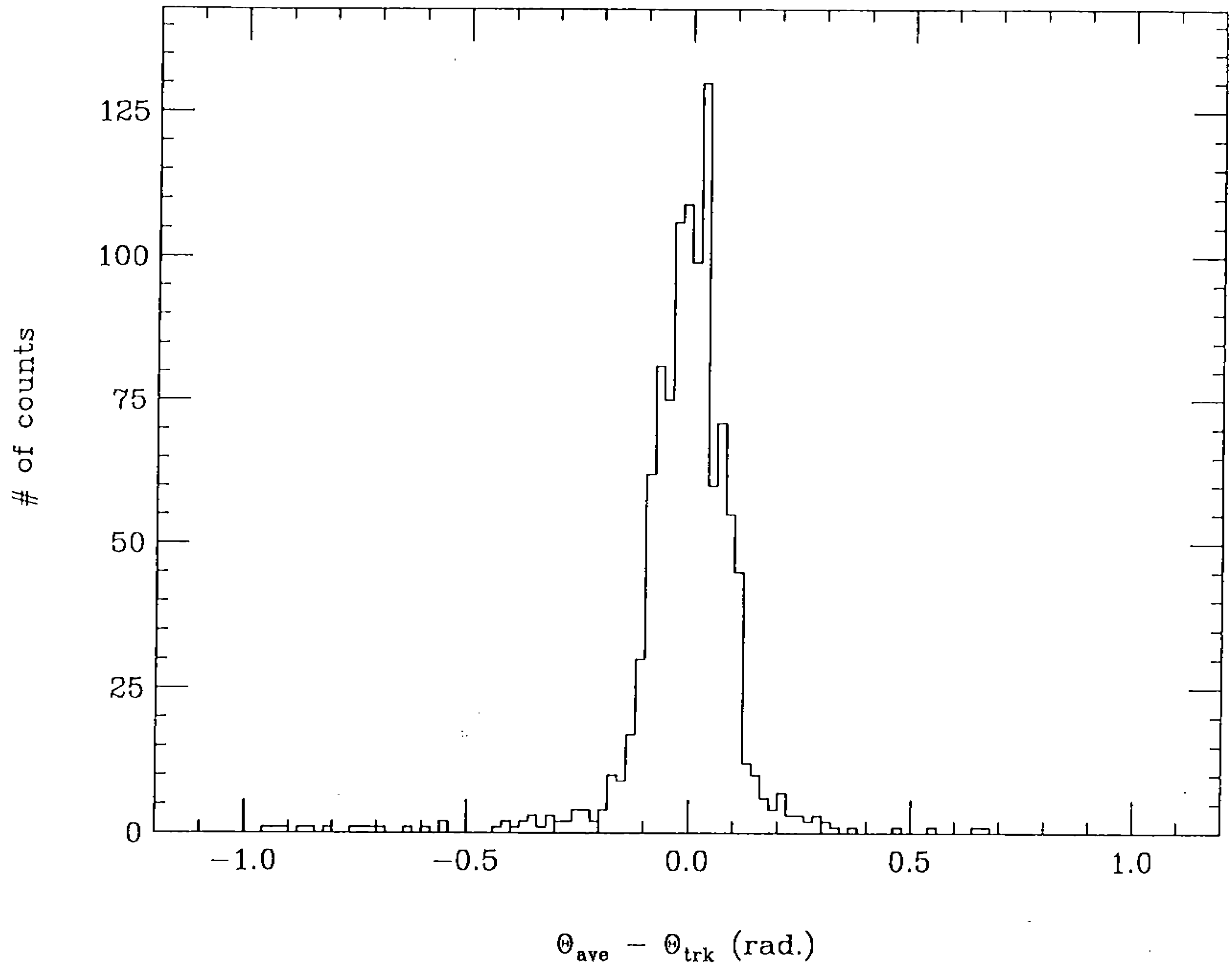
CLAS Magnet and Support Structure





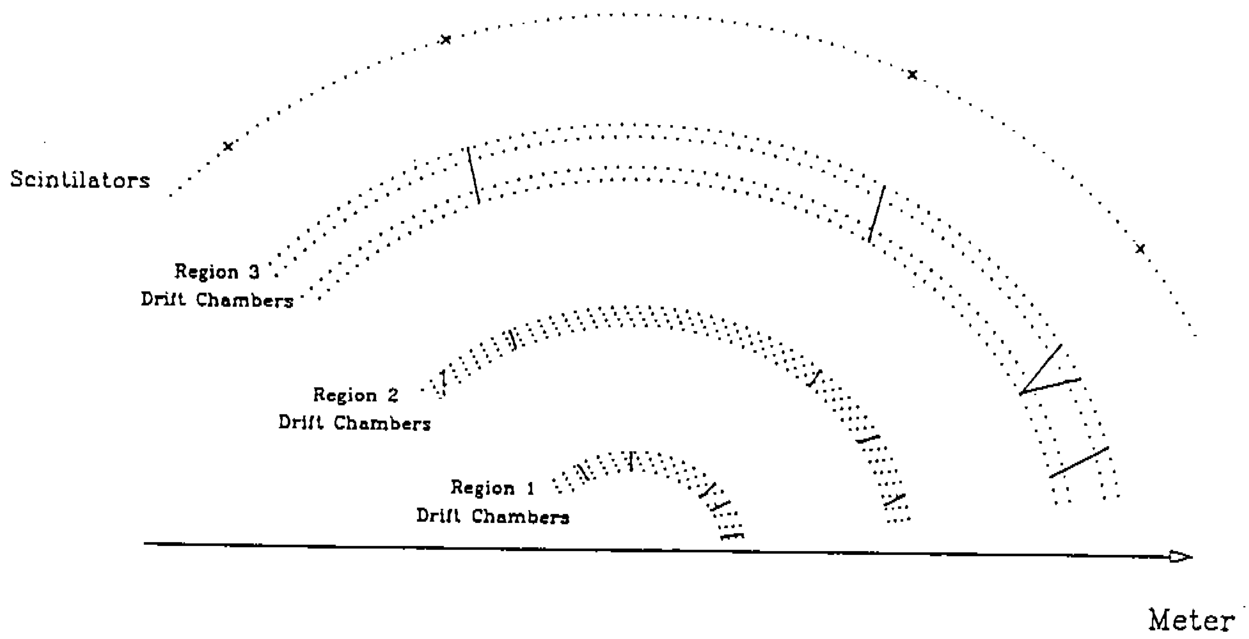

10 (cm)
9
8
7
6
5
4
3
2
1
0
-1
-2
s
s
h
-2
-1
0
1
2
3
4
5
6
7
8
9
10
11
12 (cm)


branches.
branch
weight

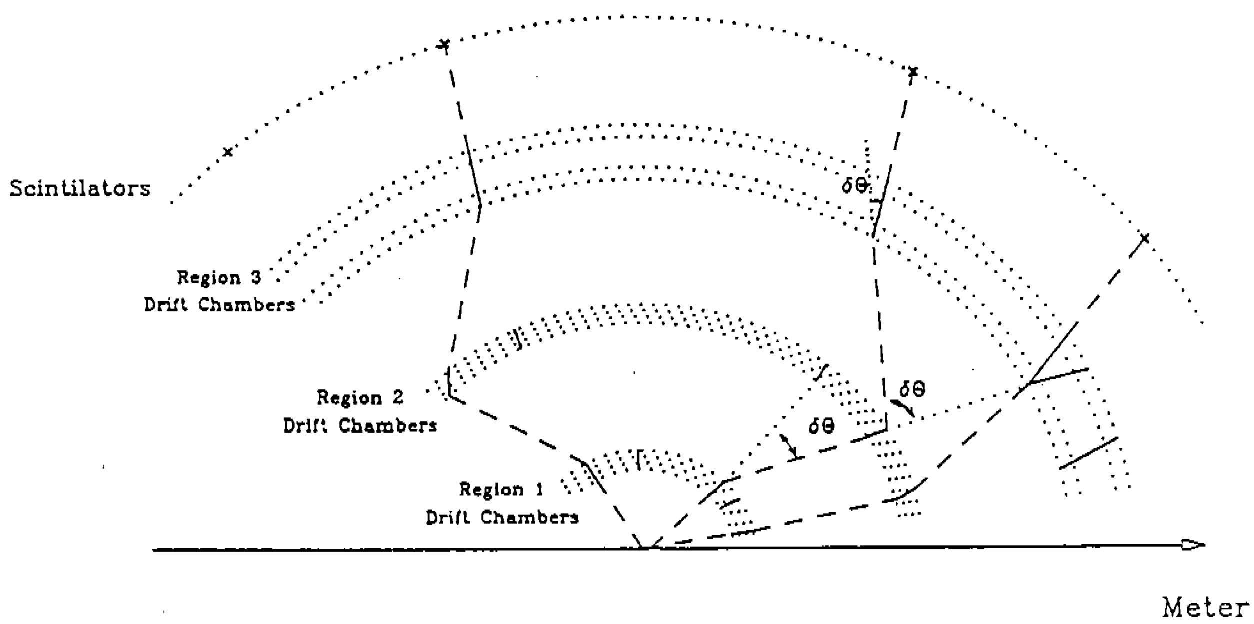


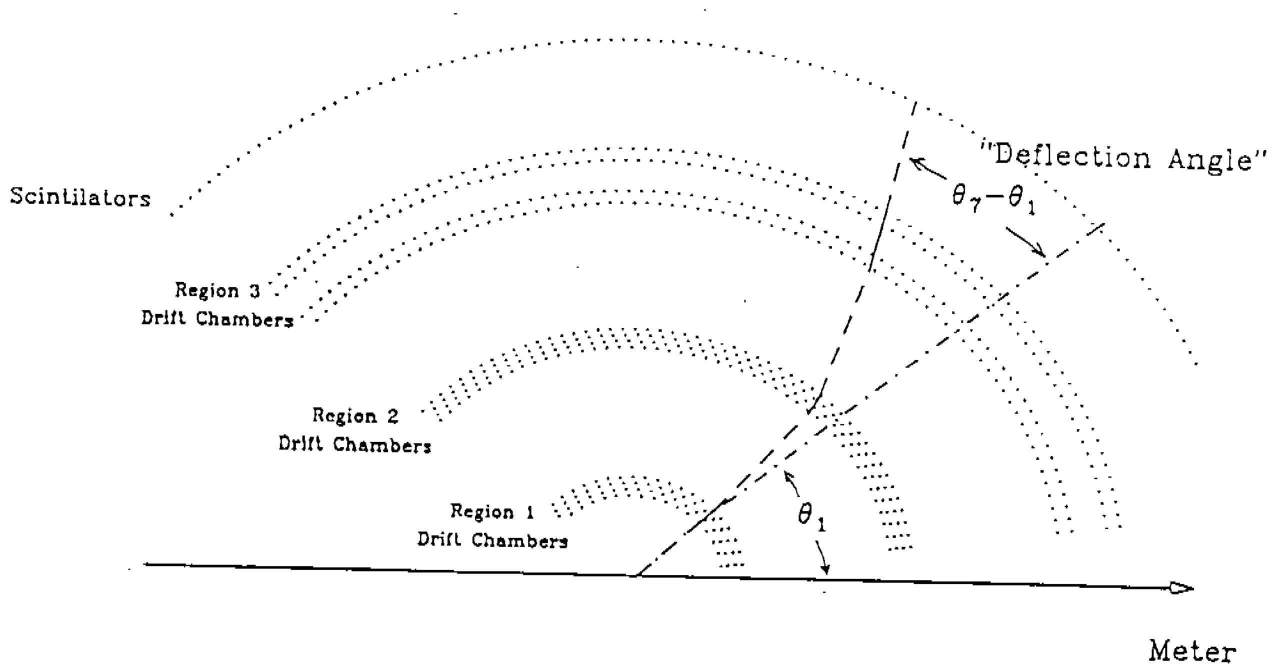
125
100
75
50
25
0
−1.0
−0.5
0.0
0.5
1.0
# of counts
$\Theta_{ave} - \Theta_{trk}$ (rad.)

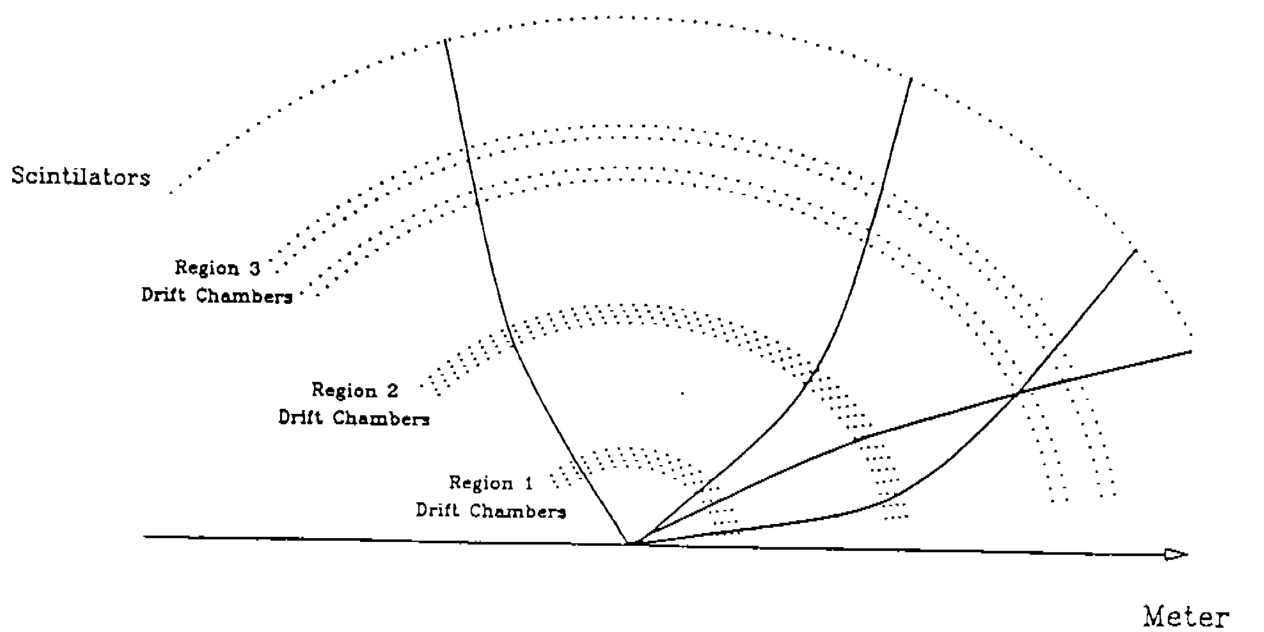
Track segments in superlayers

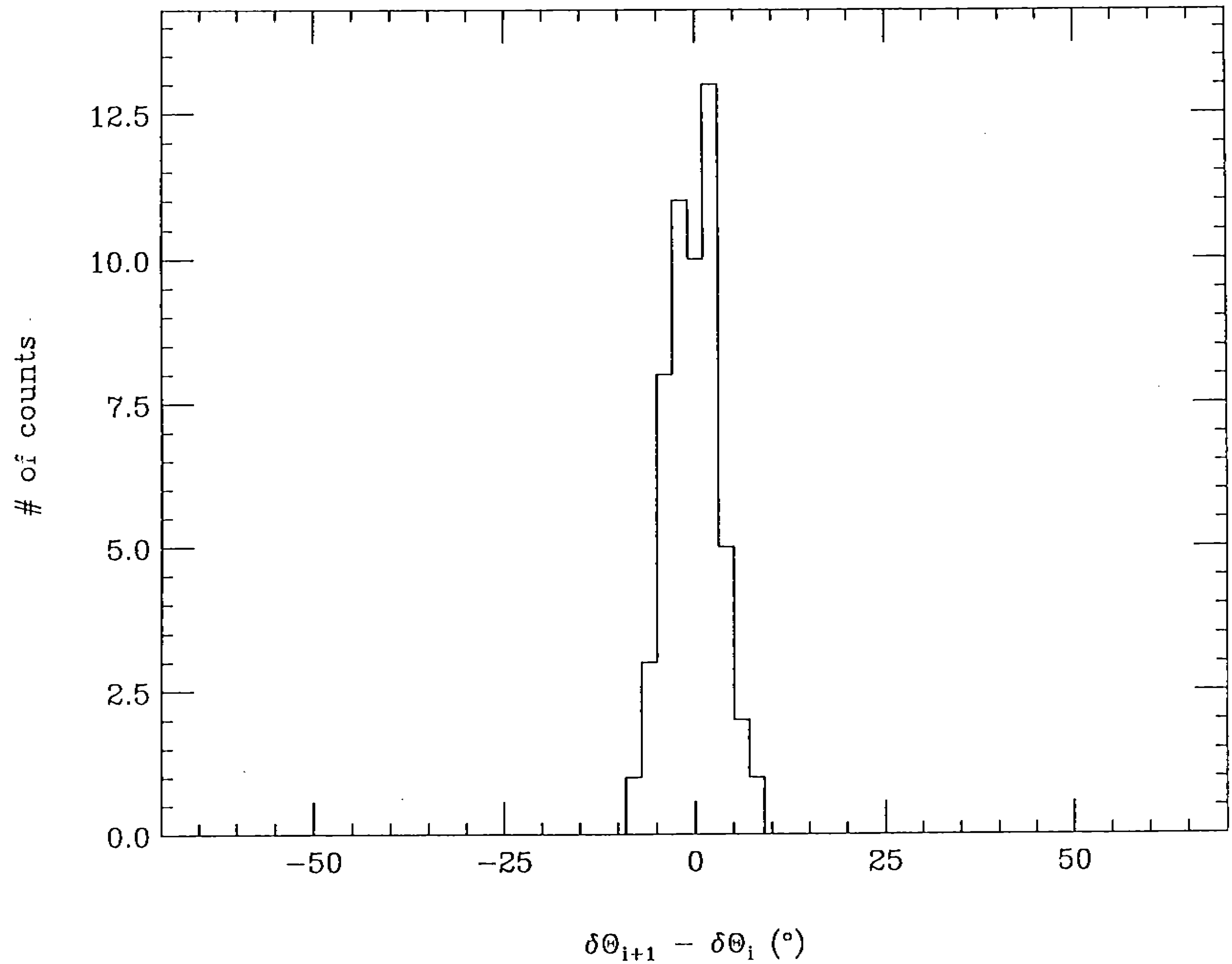


Trees and Branches

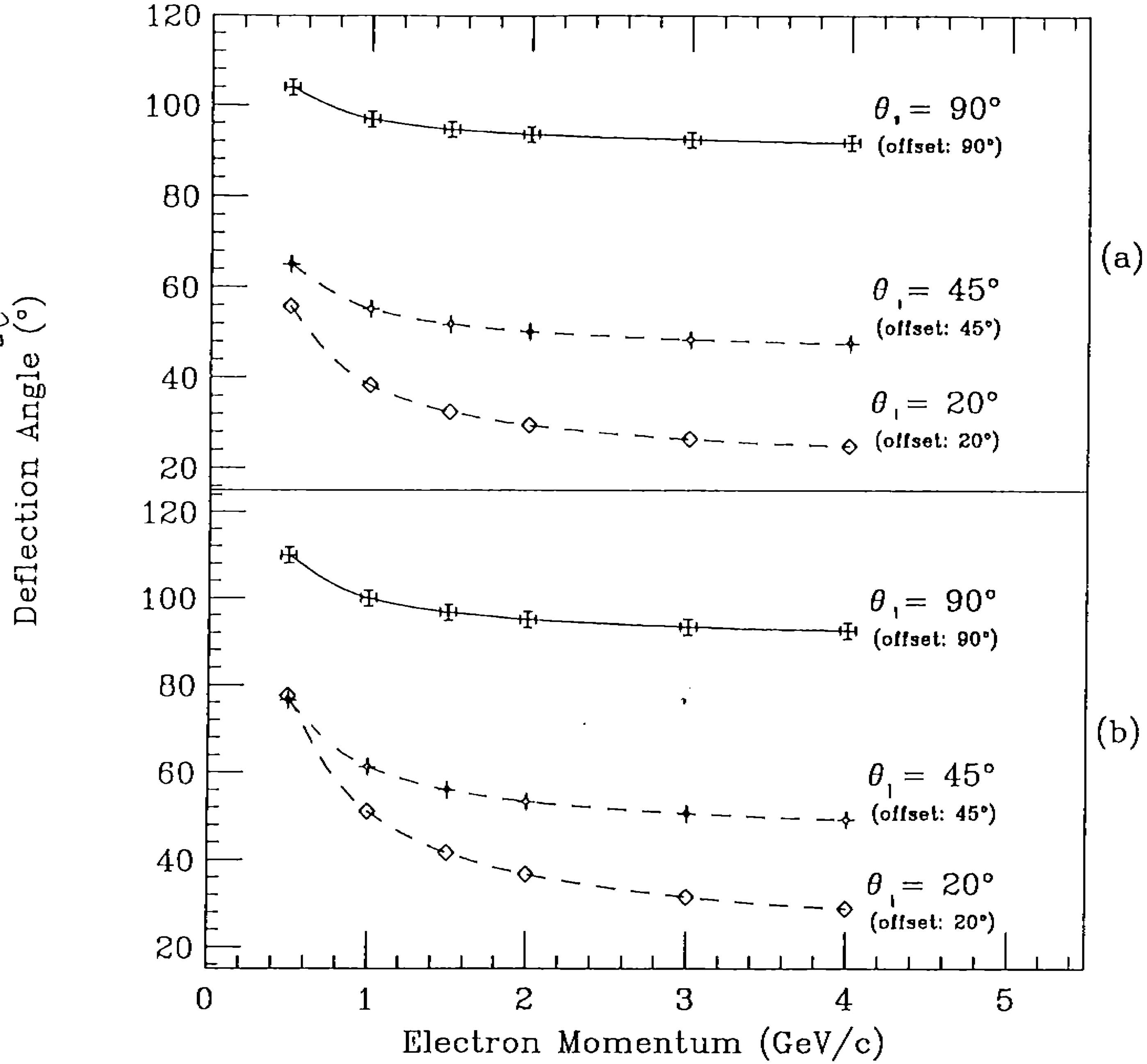



"Deflection Angle"
$\theta_7-\theta_1$
Scintilators
Region 3
Drift Chambers
Region 2
Drift Chambers
Region 1
Drift Chambers
θ_1
Meter


Scintilators
Region 3
Drift Chambers
Region 2
Drift Chambers
Region 1
Drift Chambers
Meter



12.5
10.0
7.5
5.0
2.5
0.0
# of counts
−50
−25
0
25
50
δΘi+1 − δΘi (°)



Deflection Angle (°)
Electron Momentum (GeV/c)
θ1 = 90° (offset: 90°)
θ1 = 45° (offset: 45°)
θ1 = 20° (offset: 20°)
(a)
(b)