GEM TN-93-468

Pattern Recognition in the GEM Central Tracker at Luminosity of 10^{33} $cm^{-2}s^{-1}$

W.W. Kinnison
M. L. Brooks
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

September 28, 1993

Abstract:

A GEANT based pattern recognition algorithm has been developed for simulations of the GEM central tracker. We describe the pattern recognition algorithm and present the results of studies of the track finding efficiency for single isolated tracks and for all tracks present in a Higgs event with minimum bias background at luminosity $10^{33}cm^{-2}s^{-1}$.

GEM TN-93-468

Pattern Recognition in the GEM Central Tracker at Luminosity of 10^{33} $cm^{-2}s^{-1}$

M. L. Brooks, W. W. Kinnison
Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Abstract A GEANT based pattern recognition algorithm has been developed for simulations of the GEM central tracker. We describe the pattern recognition algorithm and present the results of studies of the track finding efficiency for single isolated tracks and for all tracks present in a Higgs event with minimum bias background at luminosity $10^{33}cm^{-2}s^{-1}$.

Pattern Recognition in the GEM Central Tracker at Luminosity of 10^{33} $cm^{-2}s^{-1}$

Introduction A pattern recognition algorithm has been developed for simulations of the GEM central tracker. The algorithm was developed with the goal to reconstruct all tracks in physics events (with minimum bias background levels expected at the SSC) that are above approximately 1 GeV/c over the entire eta and phi range of the detector, using only coordinates from the tracker. This paper will describe the algorithm, and present some of the results of the pattern recognition when it was used for physics event simulations of the GEM tracker.

Pattern Recognition The pattern recognition algorithm first finds all potential vectors (two coordinate points within the tracker that are joined, with kappa, phi, and theta values estimated by assuming a third point at the origin of the tracker) that are consistent with a "stiff" track originating approximately from the origin of the tracker. The layers of the tracker are then searched, and all possible trees of vectors are built which originate on the first layer, out to the N-3rd layer, where N is the total number of layers in the detector. For a given tree, a chi-square value is calculated for each possible path of vectors on the tree, and the best chi-square value that passes a chi-square cut is recorded as a track. Each set of coordinates from the found tracks is then fit to a helical track and the helical parameters are stored.

Vectors are formed separately between layers in the silicon barrel, between silicon barrel layers and forward silicon layers, between forward silicon layers, between silicon and pad layers, and between layers in the pad detectors. In a given region, all possible vectors between adjacent layers are formed first. The vector search is first constrained by only searching for matching points in a limited phi region of the detector, where the phi region is determined by the phi of the inner coordinate. For all potential vectors, an estimate of kappa is calculated by assuming a third point at the origin. Kappa is then given by:

$$K = (d\phi/dr)/\mathrm{sqrt}(1 + (R*d\phi/dr)^2) \qquad (1)$$

where

$$d\phi/dr = (\phi_2 - \phi_1)/(r_2 - r_1)$$
$$\phi_2 = \tan^{-1}(y_2 / x_2)$$
$$\phi_1 = \tan^{-1}(y_1 / x_1)$$
$$R = (r_1 + r_2)/2$$

and the subscripts 1 and 2 refer to the first coordinate of the vector, and the second coordinate, respectively. All vectors that do not pass a kappa cut (which is an input variable to the program) are discarded. The values for the kappa cut were estimated from κ=(0.3*FIELD)/(2*p) where FIELD is the magnetic field in the GEM central tracker, and p was taken to be approximately 300 MeV/c (the value for p was different for the different detector regions, and was increased at higher luminosities).The theta values of the two coordinates are then calculated, and any two coordinates which have theta values that differ by more than a theta cut (also an input variable) are also discarded. These two cuts are necessary to keep the number of vectors and sizes of trees to a manageable level at high luminosities and we do, in general, make the cuts tighter with increasing luminosity so that the cpu time per event can be kept relatively small while keeping as many of the true vectors as possible.

After the adjacent layer search, if a vector was not formed for a particular coordinate that passed the above cuts, then the next layer is searched for a matching coordinate. Up to two layers are skipped to form a vector. For all coordinates in the barrel region that are consistent with a track that originates approximately from the origin and passes from the barrel region to the forward

region, the forward layers are also searched for a matching coordinate to form a vector. Similar kappa and theta cuts are used for these vectors as well as for vectors that are formed between layers in the forward region.

For each vector that is formed that passes the kappa and theta cuts, six parameters are calculated and stored: 1) the estimated kappa value that was given above, 2) an estimate of ϕ_0, which is the ϕ value at the origin of the track consistent with the estimated curvature of the track, 3) $\cos(\theta)$, where θ is the angle, relative to the z axis, between the two coordinates, 4) the estimated error of kappa, which is calculated from the expected coordinate resolutions, 5) the estimated error of ϕ_0, and 6) the estimated error of $\cos(\theta)$. The parameters, other than kappa, are given by:

$$\phi_0 = (\phi_2 + \phi_1)/2 - \sin^{-1}(K*R) \tag{2}$$

$$\cos(\theta) = \cos[\tan^{-1}((r_2 - r_1)/(z_2 - z_1))] \tag{3}$$

$$\sigma_K = \mathrm{sqrt}(\sigma_A{}^2 + \sigma_B{}^2) \text{ where} \tag{4}$$

$$\sigma_A{}^2 = \sigma^2(d\phi/dr)/(1 + (R*d\phi/dr)^2)^3$$

$$\sigma_B{}^2 = \sigma^2(R)*(R*K^3)^2$$

$$\sigma_{\phi 0} = \mathrm{sqrt}\{\sigma^2[(\phi_2 + \phi_1)/2] + [(\sigma_K*R)^2 + (\sigma_R*K)^2] / [1 - (K*R)^2]\} \tag{5}$$

$$\sigma_{\cos(\theta)} = \mathrm{sqrt}[\sigma^2{}_{\Delta z}*(\Delta r)^4 + \sigma^2{}_{\Delta r}*(\Delta r*\Delta z)^2]/(\Delta r^2 + \Delta z^2)^{1.5} \tag{6}$$

$$\Delta r = (r_2 - r_1)$$

$$\Delta z = (z_2 - z_1).$$

The σ's that are dependent upon r, z, and ϕ are calculated from the expected resolutions of the space points which come from the detector and alignment resolutions. These resolutions are different for the silicon and pad layers, and are different for the barrel and forward regions of silicon and pads.

Trees of vectors are then built for each coordinate on the first layer through the N-3rd layer. These trees can have vectors which cross layers within a detector region, which pass from the silicon or pad barrel region to the silicon or pad forward region, or which pass from silicon detectors to pad detectors. The work that will be presented here built trees separately in the silicon and pad detector regions, and then joined tracks in the two regions by looking for good chi-square matches of the track parameters. Currently, though, the pattern recognition that is used for the GEM tracker builds trees that extend all the way from the inner silicon to the outer pad layers, and complete tracks are found in one pass through the tree search.

For each tree, weighted chi-square values of the three vector parameters (κ, ϕ_0 and $\cos(\theta)$) that were given above are calculated for each possible path through the tree. The coordinates from the path that has the best chi-square value and passes a chi-square cutoff for each tree are stored as a cluster. In the work that will be presented here, the chi-square cutoff has been set at a fairly large value of 30. The chi-square distribution for single muons can be seen in Figure 1. Because of the chi-square cutoff, it is possible that no cluster will be stored for a given tree. It is also possible that the smallest chi-square value will be formed for a cluster which does not include all of the true coordinates from a track (e.g. a track that produces six coordinates may have a smaller chi-square value if only the first five are included in the cluster rather than all six coordinates). Because of this, care has been taken to try to reduce the number of clusters found that have thrown out a "good" coordinate because adding it increases the chi-square value of the cluster. If there are two chi-square values which are small (less than 2), or are greater than 2 but close in value to each other, and one path has more coordinates than the other, then the path with the largest number of coordinates is stored, as long as the chi-square value of the path with the larger number of

coordinates passes the chi-square cutoff. This helps reduce the number of coordinates that are thrown out because they would add a minor increase to the chi-square value.

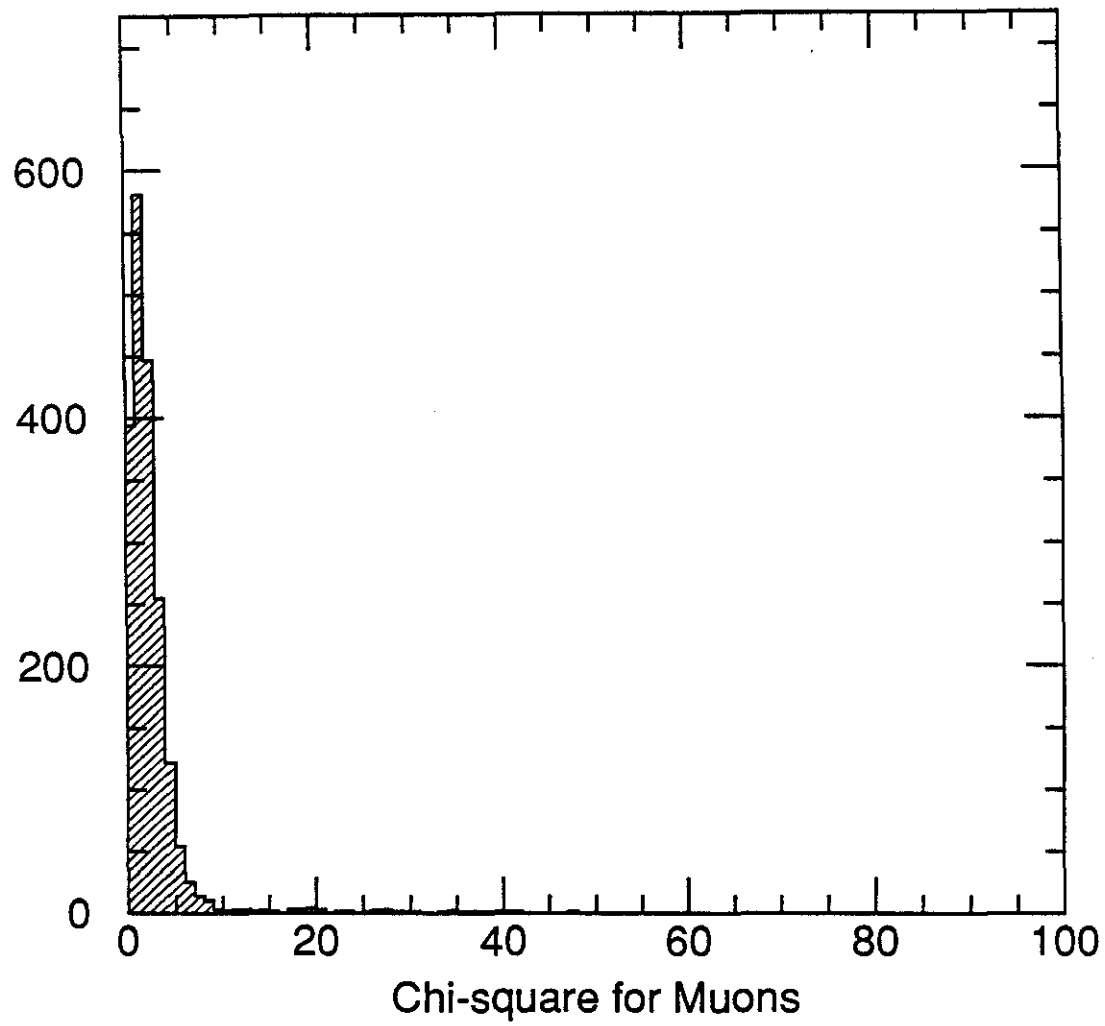


Figure 1. The chi-square distribution for potential tracks when pattern recognition is performed on single muon tracks.

Track Finding Efficiencies The track finding efficiencies of the pattern recognition algorithm have been calculated for the GEM central tracker for Higgs events at luminosity 10^{33} $cm^{-2}s^{-1}$. The GEM tracker consists of an inner tracker made of silicon wafers and an outer tracker made of interpolating pad chambers (IPC's). There are two regions of both silicon and IPC's: central regions where the detectors are arranged in cylinders that are concentric to the beam axis, and forward regions where the detectors are arranged in disks that are perpendicular to the beam axis. In the central regions there are six layers of stereo pairs of silicon extending ±18 cm and four layers if IPC stereo pairs which extend to ±100 cm. The forward silicon disks cover z = ±20 cm

to z = ±87 cm and the forward IPC disks cover z = ±113 cm to z = ±168 cm. A picture of the GEM tracker geometry that is in the GEANT code can be seen in Figure 2.

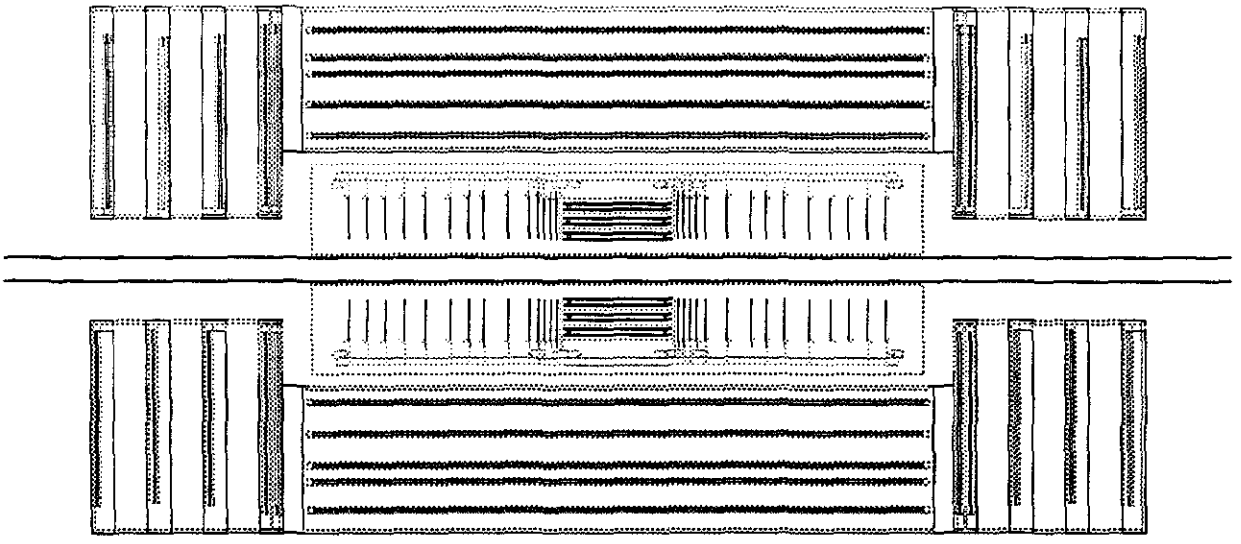


Figure 2. The layout of the GEM central tracker as produced by GEANT, showing both the silicon portion as well as the IPC portion.

For our studies of pattern recognition the track finding efficiency in the GEM tracker has been defined as (in the silicon portion)

$$\text{efficiency} = \frac{\text{\# of found tracks with 4 or more correct points}}{\text{\# of tracks that passed through 4 or more silicon layers}} .$$

For the efficiency in silicon plus pads, n in the numerator was 8 and the denominator required 4 silicon layers and 4 pad layers to have been passed through. This definition removes most of the geometric efficiency from the track finding efficiency calculation. However, as can be seen in Figure 3, where the number of tracker layers hit vs. eta is shown for all stiff tracks in a physics event, virtually all tracks pass the geometric criterion that four or more silicon and four or more pad layers are hit. There is a small area, at the boundary between the barrel and forward regions (η=(1.0,1.5)) where the number of hits drops, but it still generally stays above eight hits. (Note, there are secondaries in this histogram which show hits below eight because these particles were produced in the middle of the tracker.) The decreased geometric coverage in certain areas of the detector makes it more difficult for the pattern recognition algorithm to find tracks, but does not decrease the efficiency significantly as will be shown.

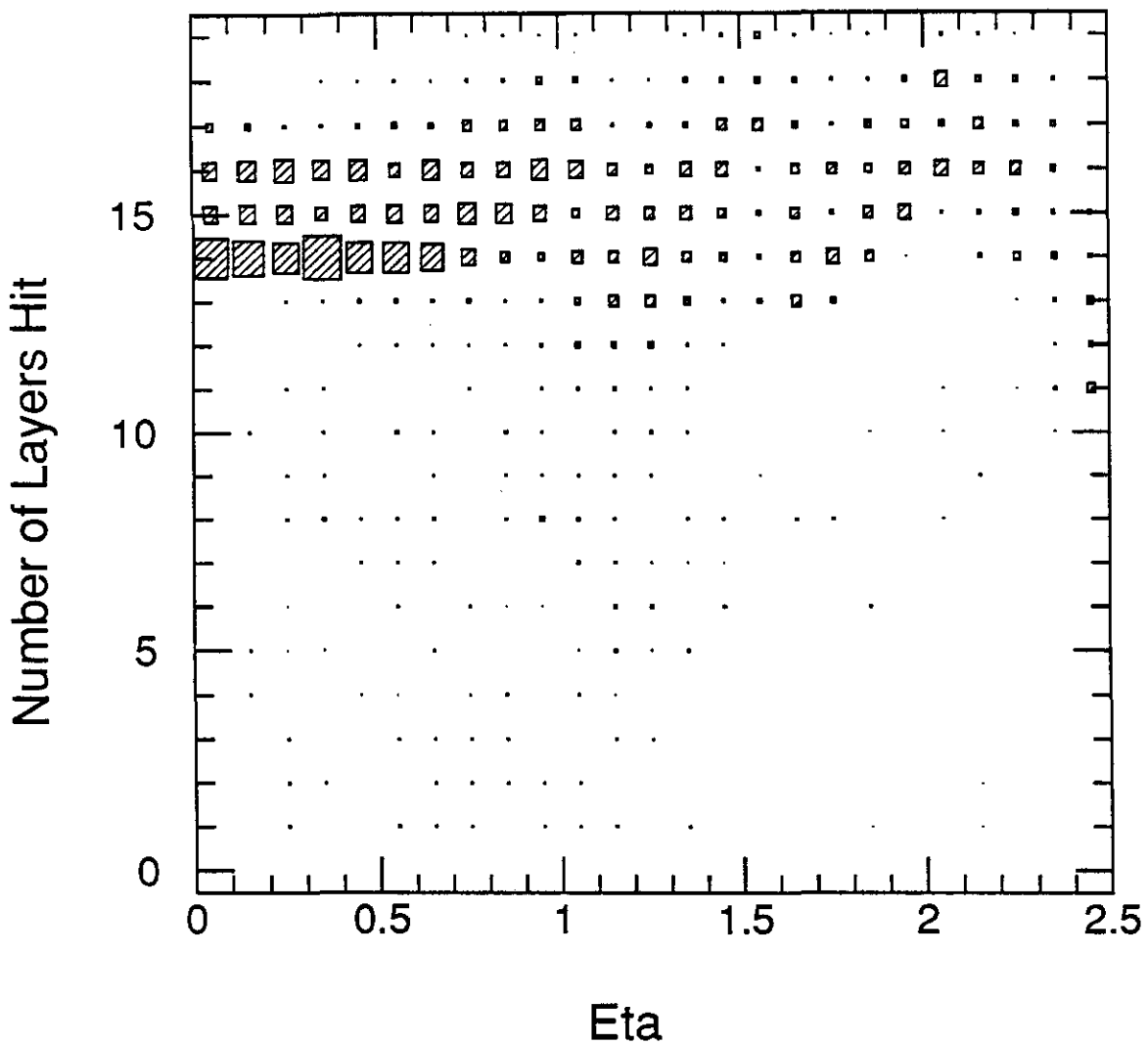


Figure 3. The number of tracker layers that stiff tracks in a full physics event, with vertex smearing, hit when passing through the tracker vs. eta.

When we calculate the track finding efficiency for single muon tracks with momentum above 5 GeV/c passing through the tracker, we find almost 100% efficiency in the silicon portion and in the combined silicon plus pad tracker. This can be seen in Figure 4 where we show the efficiency for finding tracks above 5 GeV/c vs. η when single muons of momentum=(1,500) GeV/c were run through the tracker. This high efficiency implies that our coordinate forming inefficiency is reasonably low, our vector search is sufficient to find almost all of the vectors that come from true tracks, and our chi-square cuts are not throwing out good tracks.

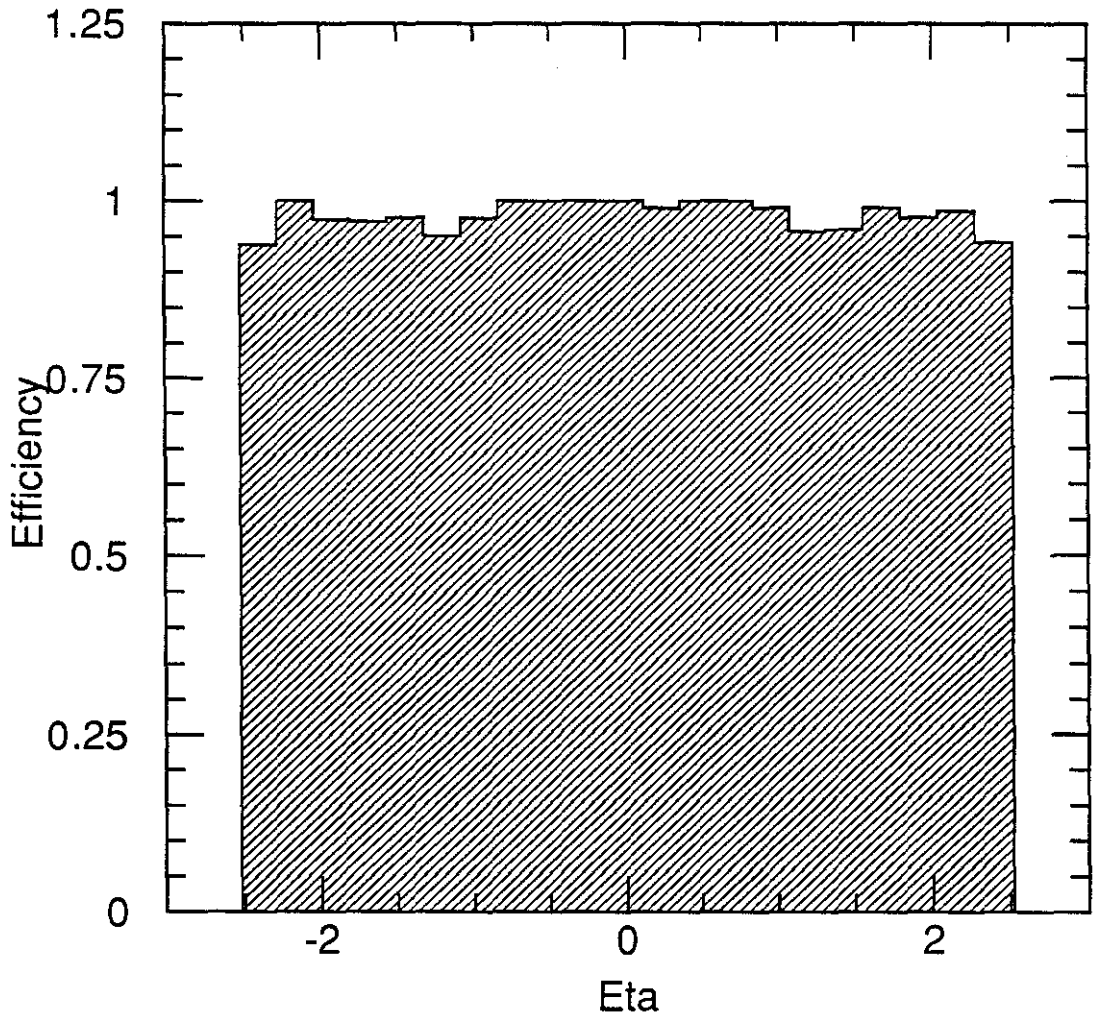
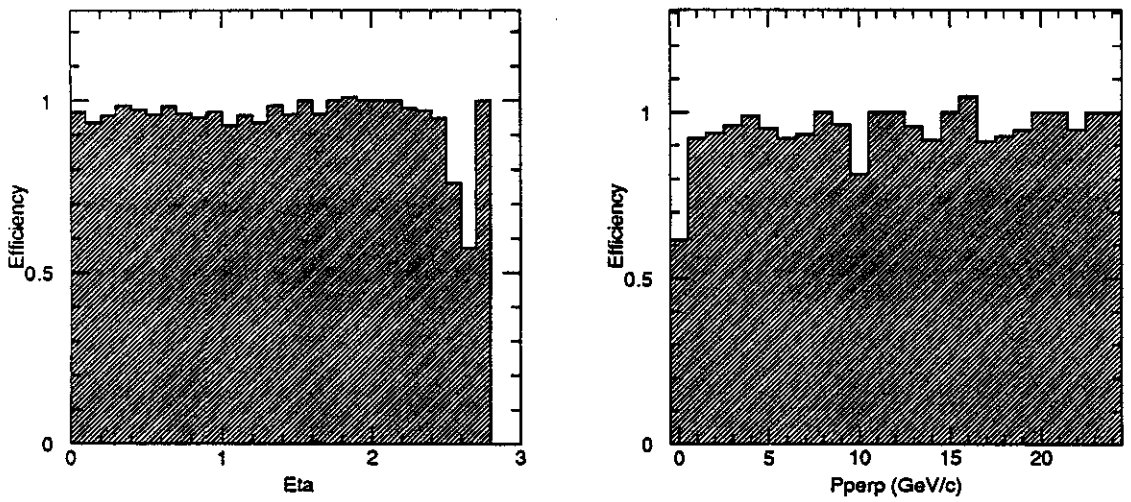


Figure 4. The fraction of muon tracks vs. eta with momentum above 5 GeV/c that are found with the pattern recognition algorithm when single muon tracks with momenta from below 1 GeV/c to 500 GeV/c are run through the tracker.

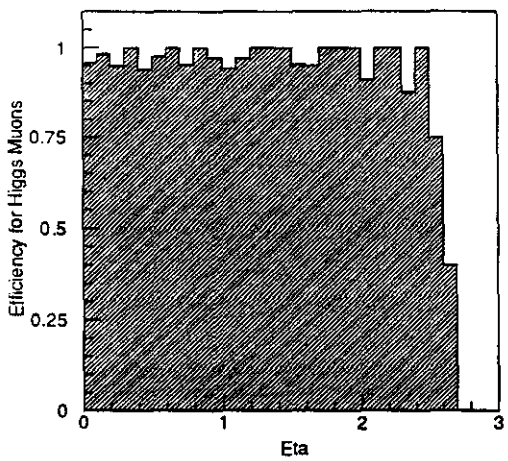
To fully test our track finding algorithm, we have looked at track finding efficiencies in a typical physics event, which we have defined to be a Higgs event (m_{Higgs} = 300 GeV, $H^o \rightarrow l^+l^- l^+l^-$) with a random number of minimum bias events (Poisson distribution with mean 1.6) as background. When we calculate the efficiency for finding all tracks above 5 GeV/c, we find that approximately 97% of all tracks are found (Figure 5) if a track is defined to be "found" if 4 or more correct points are found in the silicon tracker and 4 or more are found in the pads as was described above. Note that the efficiency in Figure 5 is approximately flat up to about eta=2.4 and then drops off, and the apparent rise in the last bin of the histogram is not real, but is just caused by poor statistics. The efficiency is also fairly flat vs. the p_T of the particle, down to approximately 1 GeV/c as can be seen in Figure 6. Again, the statistics are somewhat poor in this histogram, which

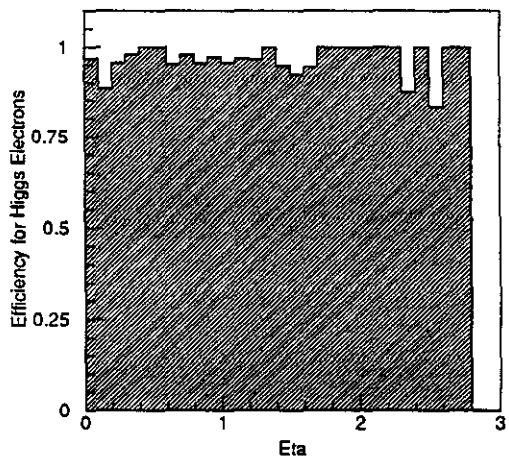
causes an apparent dip at approximately 10 GeV/c, but the real efficiency loss can be seen in the first histogram bin.



Figures 5 and 6. The fraction of tracks above 5 GeV/c that are found vs. eta and the fraction of all tracks found vs. p_T by the pattern recognition algorithm when a Higgs event with 10^{33} minimum bias background events are run using the GEANT code.

The efficiency for finding the Higgs leptons has been recorded separately and it was also found that approximately 97% of the electrons are found and approximately 97% of the muons are found (see Figures 7 and 8). This shows that the pattern recognition can find electrons equally as well as muons, despite the fact that the electrons in general produce many more extra hits for the pattern recognition (because of interactions with the material of the tracker) than muons.





Figures 7 and 8. The fraction of muons from Higgs decays that are found vs. eta and the fraction of electrons from Higgs decays that are found vs. eta by the pattern recognition when Higgs events with 10^33 minimum bias background events are run. Both integrated efficiencies are approximately 97%.

The number of fake tracks that were found by the pattern recognition algorithm was also investigated. The fake tracks were defined to be all tracks that were found with 8 or more coordinates that had more than 2 incorrect coordinates from the silicon or more than two incorrect coordinates from the pads. When 100 Higgs events with minimum bias background were run, 591 good tracks above 5 GeV/c were found with eight or more coordinates, and 2 fake tracks above 5 GeV/c were found; 5362 good tracks were found counting all momentum tracks, and 26 fake tracks were found counting all momentum tracks (see Figure 9).

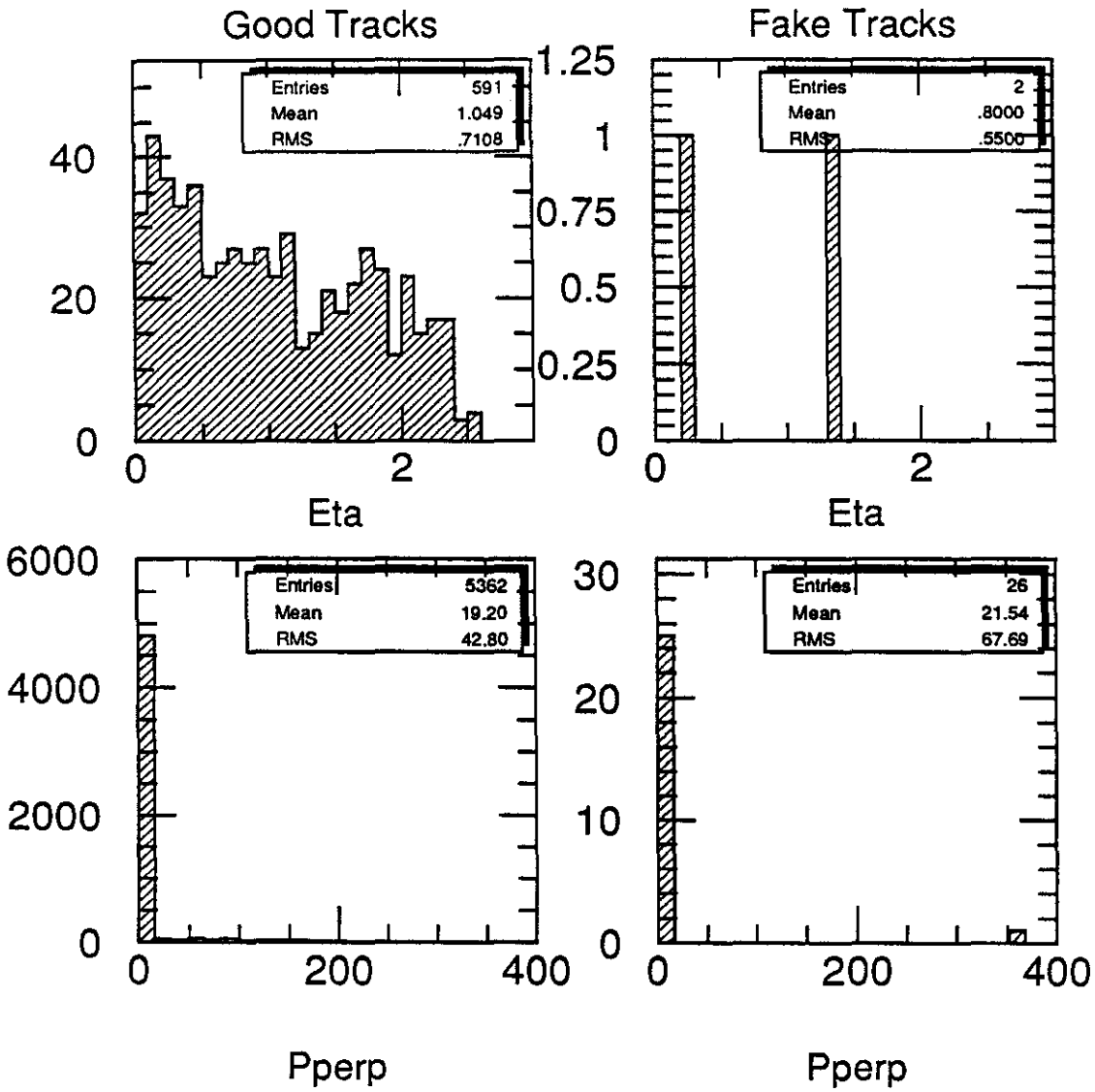


Figure 9. The distribution of good tracks and fake tracks that are found with the pattern recognition in a Higgs event with 10^{33} minimum bias background 1) that are above 5 GeV/c p_T vs. eta and 2) that are of all momentum vs. p_T of the track.

Conclusion: The pattern recognition algorithm that has been developed can easily separate high p_T tracks from minimum bias background at the 97% level when applied to the GEM central tracker at a luminosity of 10^{33} $cm^{-2}s^{-1}$. The number of fake tracks, at the same time, is at a level of approximately 0.5% of the tracks found, or 0.02 tracks per event above 5 GeV/c and 0.2 total fake tracks per event.

This pattern recognition algorithm has also been used in conjunction with the helical fitter to determine the momentum resolution, impact parameter resolution, and z vertex resolution over the entire eta range of the GEM tracker, for momenta from 1 GeV/c to 500 GeV/c and separately for electrons and muons. The results can be found in GEM TN-93-470.