ALEPH 93-23
MINIV 93-01
A. Bonnissent et al.
11.02.1993

Searching for the origin of VDET pattern recognition errors using the Monte Carlo truth

Alain Bonissent,John Carr and John Drinkard
CPPM Marseille

January 21, 1993

1 *Introduction*

The current Vdet hits to tracks association algorithm is very efficient, as can be seen from many physics analyses. As we shall show below, some 3% of the hits were associated to the wrong track. If one takes into account the fact that a typical track should use 4 hits (one in each layer and each view), this is far from being negligible.

These errors are often assumed to be due to the nature of the algorithm, which processes one track at a time. It is generally believed that a global treatment or refinement would permit better overall results, by optimising the full assembly of tracks and hits, therefore eliminating the dependency on the order in which the tracks are processed.

Such an algorithm which would work on hadronic events does not exist yet, although several ideas have been proposed or tested. Information on the origin of these errors might help to specify the exact nature of the problem, and suggest solutions to solve it.

On Monte-Carlo events, it is possible to know which track(s) a given hit originated from, and compare this with the reconstruction output.

The level of noise, or extrapolation errors, due to imperfect treatment of the multiple scattering or alignment, can be different on real data. Nevertheless, Monte-Carlo events are convenient to trace the origin of the pattern recognition inefficiencies in a somewhat idealised environment : everything which goes wrong here is unlikely to improve on real data.

2 *Navigating from tracks to Vdet hits*

The relation between reconstructed tracks and Vdet hits, as established by Julia, is stored in the VCPL bank :

$$FRFT < -VCPL- > VDXY or VDZT$$

For the truth information, we first have to navigate from FRFT to FKIN tracks, through the TPC or ITC hits. This is nicely done using the subroutine VMCTRK from the SETQUAL package. Then, navigation to VDET hits follows the following path :

$$FKIN < -VDTD < -VFHL < -VDXY or VDZT$$

This is done by the subroutine VTRUTHL from the SETQUAL package.

With these tools, for each hit in either bank VDZT or VDXY, the list of truth or Julia associated tracks are established and compared.

3 *Results*

In what follows, we consider independently the XY and ZT views. The word hit refers to a row in either VDXY or VDZT bank.

A sample of 2000 hadronic Monte-Carlo events was used, comprising some 40000 tracks and 100000 hits. The events were generated by Galeph version 250, with database 161, and reconstructed with Julia version 260, using database 165. Dead line drivers are simulated, with a time dependent efficiency map which reproduces the situation of real data. Dead strips (pinholes) were not simulated. This will later be referenced to as sample A. Sample B differs from sample A by the fact that the dead line drivers have been removed : this simulates an idealised perfect Vdet. Sample C is a more realistic situation, simulated with the most recent code : the appropriate proportion of line drivers are dead, and the effect of pinholes was simulated by randomly suppressing the appropriate proportion of clusters, in order to mimmick the situation during 1992. Each data set comprises the same number (2000) of events.

Bad hits are those which are erroneously assigned to a track which did not really create them. If all reconstructed tracks and hits are taken into account, the ratio of improperly associated (bad) hits to the total number of hits is 4.0%. For physics considerations, one should concentrate on so called "good tracks". We defined good tracks as :

- $N_{TPC} > 4$
- $d_0 < 2cm$
- $|Z_0| < 5cm$
- $cos(\theta) < 0.95$
- not issued from a reconstructed V_0

Only hits which were assigned to a good track were considered in the table 1, which gives their classification, in percentage of the total number of hits on good tracks :

Data set	A	B	C	
Total wrongly associated hits	3.2%	3.0%	4.4%	1
Spurious	1.0%	1.0%	1.4%	2
Caused by other track(s)	2.2%	2.0%	3.0%	3
Caused by more than 1 track	0.3%	0.3%	0.4%	4
True track was processed before	0.8%	0.8%	1.1%	5
True track was processed after	1.1%	1.0%	1.5%	6
... and was closer	0.8%	0.8%	1.3%	7

Table 1: hits distribution. Proportions are on total number of hits assigned to tracks. Samples are : A, standard 1991 Monte-Carlo; B, same with no dead line drivers; C, 1992 simulation (dead line drivers and pinholes are taken into account).

In row 1, we find the total proportion of association errors. It is of the order of 3-4% for all data sets. This can be split into 1.6% for tracks which have 2 hits , and 9.5% when the track has only 1 hit in the relevant view. Row 2 refers to hits which are either caused by a track which was not reconstructed (low momentum, looper, etc...), or by noise. The rest of bad hits have been caused by an other track (could be good or bad); row 1 is the sum of rows 2 and 3.

The hits accounted for in row 4 were created by several tracks. It is difficult to decide if they are assigned to the correct one. Such hits are likely to create trouble, because the cluster barycenter can be far away from any of the correct tracks. Their proportion is small.

The following rows refer to the problem of one track "stealing" a hit which should belong to some other. Row 5 corresponds to the hits for which this is irrelevant, because the correct track did not pick the hit, although it was processed first. These plus the multi-track hits represent half of the wrong hits.

In row 6, we find the complement, i.e. the hits where the correct track was processed later. A refined algorithm should be able to reduce this class of wrong hits, but this is likely only if the hit was closer to the correct track than to the wrong one. We see in row 7 that this is only of the order of 1% of the total number of hits. This is a rather small number, even if we keep in mind that 4 hits are normally used by a track.

A clue on the origin of the association errors is given in figures 1 and 2. The distribution of the distance between the impact (in the relevant Vdet layer) of the track which caused the hit and the impact of the track to which it was associated is shown, in the relevant Vdet layer and view. In each figure, we have the two plots refering to mistakes in XY and ZT views respectively. In the ZT view for example, we see that for ZT mistakes the distance is rather small, reflecting the distribution of the track extrapolation errors, while for the XY mistakes, we have a broader distribution. In the XY view, the situation

Data set	A	B	C	
Total missing hits	12.7%	10.2%	15.8%	1
Stolen by other track	1.0%	1.2%	1.6%	2
Existent but not used	1.6%	1.8%	2.4%	3
Non existent	10.1%	7.1%	11.8%	4
Geometric loss	2.5%	2.5%	2.5%	5
V_0 not reconstructed	3.8%	3.8%	4.2%	6
Vdet inefficiency	3.8%	0.8%	5.1%	7

Table 2: Missing hits distribution. Proportions are on number of hits which should be present, counted as 1 per view and per layer, for each track whose extrapolation hits this layer. Samples are : A, standard 1991 Monte-Carlo; B, same with no dead line drivers; C, 1992 simulation (dead line drivers and pinholes are taken into account). The total number of missing hits, line 1, is the sum of lines 2,3 and 4; Non existent hits, line 4, is the sum of lines 5, 6 and 7

is reversed, with smaller extrapolation errors and therefore a sharper peak for the XY mistakes. This is the manifestation of the ambiguities which always occur when two independent and orthogonal coordinate measurements are combined

In the context of a new Vdet design, with extended geometry, some attention should be given to the plot in figure 1. It shows that the distribution of the ZT distance is limited to some 5 cm, which is approximately the extension of one wafer, although the XY readout is performed on two wafers in parallel (the strips are connected). This is due to the fact that tracks are grouped into jets. When the electronic circuits will be located at the extremities of the faces, the difference in dimensions of a face in the XY and Z coordinates may require that several ZT strips in the same face should be connected together. If their distance is larger than 3-5cm, the plot shows that this should cause little problems.

We also investigated the tracks which intercept a given layer, and which do not have a hit in this layer, in each view separately, and for "good tracks" only. This is summarised in table 2, for our three datasets. (all number are per track, per layer and per view) :

These numbers confirm that the assignment of hits to tracks is close to optimal : when a track has no hit, it most often means that no hit was available, irrespective of what the other tracks did (only a small fraction of missing hits were stolen by an other track). It also shows the effect of the various hardware failures and radiation damage : the loss is here of the order of 5%, which should be fixed by the future radiation resistant design. Figure 3 a) gives the distribution of the Z coordinate for all Vdet hits assigned to reconstructed tracks, figure 3 a) shows the distribution of the Z coordinate of the impact on a Vdet layer for those tracks which have no hit in the corresponding view (XY or

ZT). In the ZT view, the spikes reflect the geometric acceptance of the wafers, with a gap of 2mm between two wafers. This is responsible for some 2.5% of hits missing. The remaining is due to dead electronic channels, and fluctuations in the energy deposited by the tracks. The latter effect can best be estimated in column B, where we have no dead channels at all. The value of 7.1% is the sum of 2.5% geometric loss, 3.8% which correspond to not reconstructed V_0 or materialisations and 0.8% physical losses. Figure 3 c) displays the ratio of missing over existing hits. Apart from the spikes which correspond to the gap between wafers, the profile is flat at the expected value of some 5%.

4 *Conclusion*

The results of this investigation show that the main source of problems is not the sequential and independent processing of tracks : if this was entirely cured, only 30% of the incorrect associations would be removed, amounting to only 1% of the total number of hits.

From this study, we believe a significant source of misassociations lies in the one-dimensional nature of the detector : the two coordinates are measured and treated by the pattern recognition algorithm independently of each other. Because the tracks appear grouped into jets, the probability is high that two tracks are very close in e.g. the x-y plane. Since the pattern reconstruction algorithm uses the correlation between the hits in the two layers and the outer tracking, this effect is much lower for two hits tracks. With the new Vdet design, with an increased angular coverage of the second layer, and lower radiation damage, the number of one hit tracks should be lower, resulting in an improved efficiency of the hits association.

Finally, the present scheme of processing one track at a time, after ordering by increasing outer tracking error seems to be close to optimal. It is unlikely that changes at this level could lead to significant improvement.

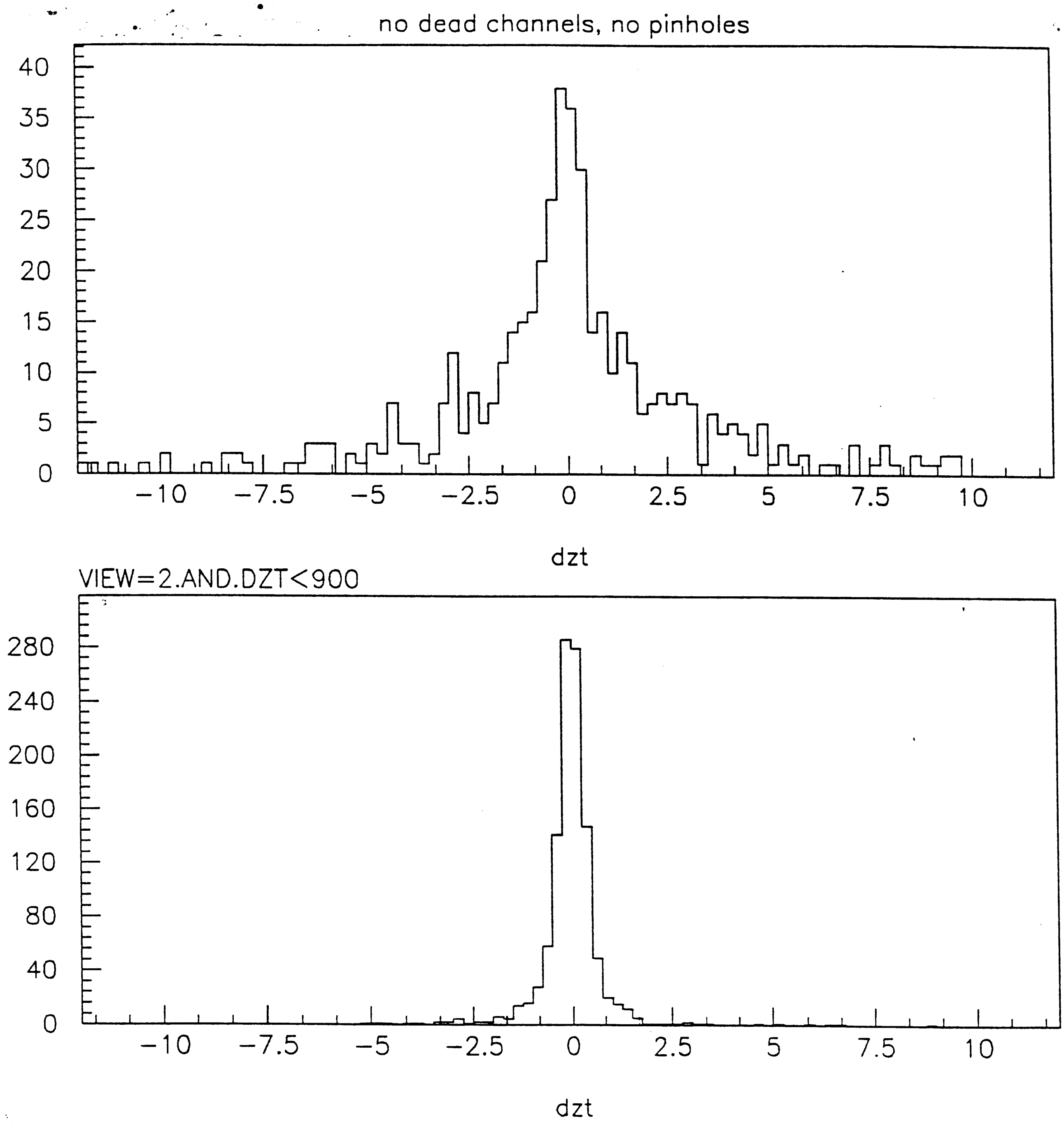

Figure 1: Distance, in the Z direction, between a wrongly associated hit and the impact of the track which caused it. Top : misassociations in the XY view; bottom : misassociations in the Z view.

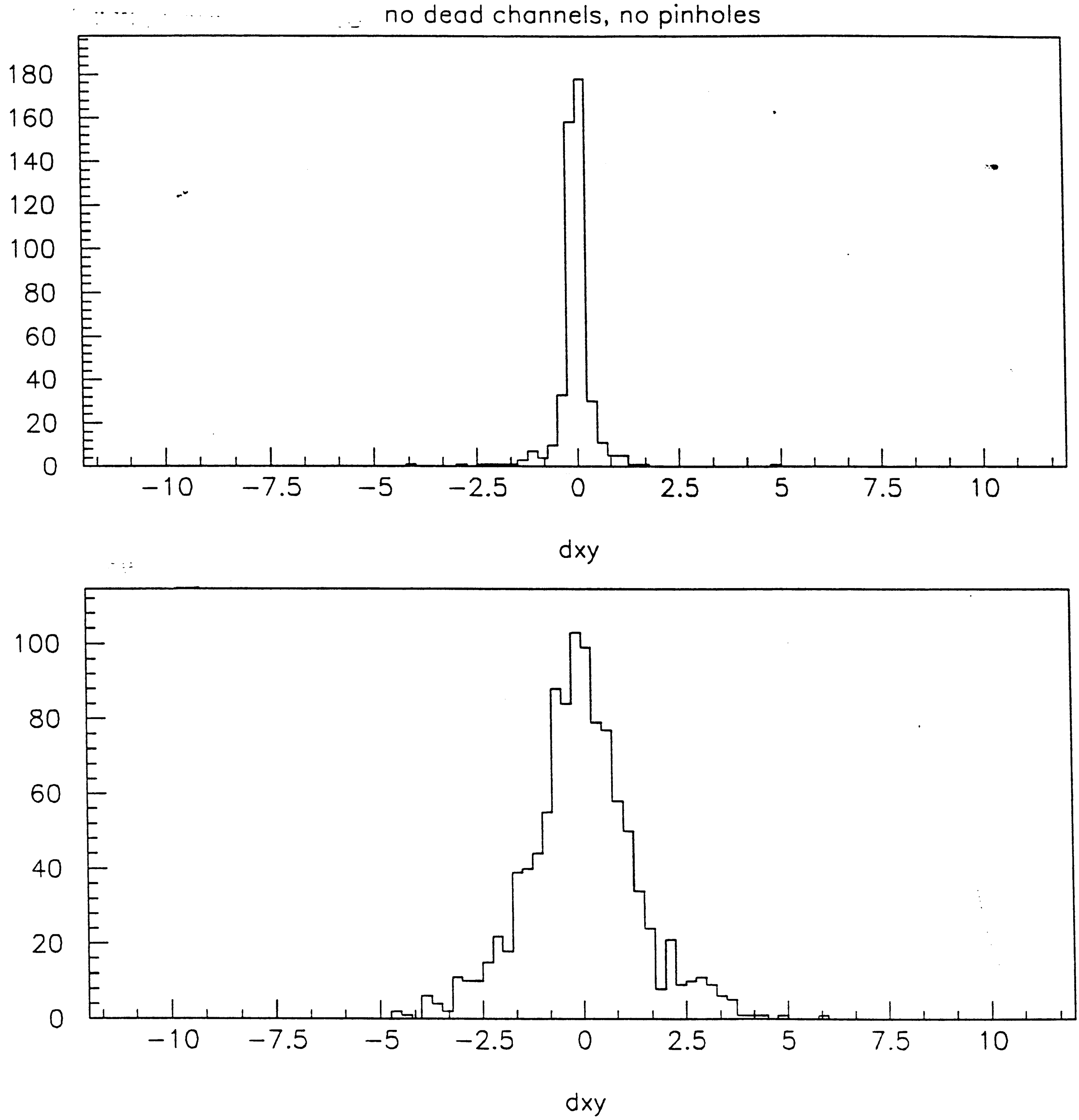

Figure 2: Distance, in the XY plane, between a wrongly associated hit and the impact of the track which caused it. Top : misassociations in the XY view; bottom : misassociations in the Z view.

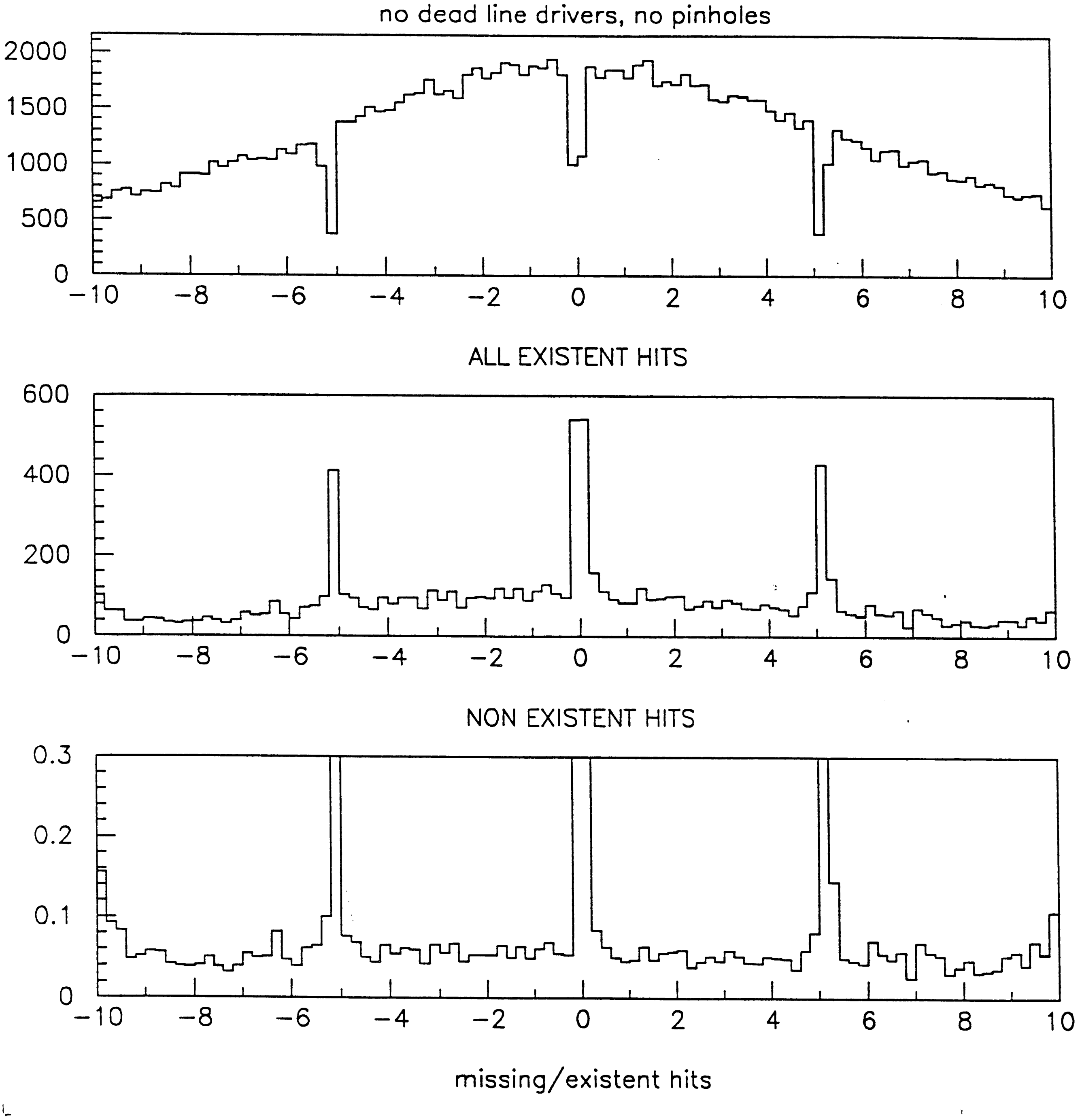

Figure 3: Distribution of the Zcoordinate for sample B (no dead channels). Top : all existing hits, all layers; center : missing hits, the position of the track's impact on the Vdet layer where the hit is missing is plotted; bottom : ratio of top and center plots, bin per bin.