HARP Collaboration

HARP Memo 04-102
12 June 2004
http://cern.ch/dydak/cluster.ps

On TPC cluster reconstruction

F. Dydak, M. Gostkin, Yu. Nefedov, J. Wotschack, A. Zhemchugov

Abstract

For a bias-free momentum measurement of TPC tracks, the correct determination of cluster positions is mandatory. We argue in particular that (i) the reconstruction of the entire longitudinal signal shape in view of longitudinal diffusion, electronic pulse shaping, and track inclination is important both for the polar angle reconstruction and for optimum $r \cdot \phi$ resolution; and that (ii) self-crosstalk of pad signals calls for special measures for the reconstruction of the z coordinate. The problem of 'shadow clusters' is resolved. Algorithms are presented for accepting clusters as 'good' clusters, and for the reconstruction of the $r \cdot \phi$ and z cluster coordinates, including provisions for 'bad' pads and pads next to sector boundaries, respectively.

1 Introduction

The TPC is a powerful instrument which measures 3-dimensional coordinates of clusters along tracks, which lend themselves to an easy reconstruction of the track trajectories without unphysical solutions. The main problem of each TPC is that systematic errors may introduce biases and deteriorate resolutions. The HARP TPC is no exception from this long-standing experience.

The HARP TPC is beset by several hardware problems which could have been avoided if more time than 17 months had been available for design and construction of the instrument.

These hardware problems are (listed in their order of importance):

1. **Decisively important:** static and dynamic distortions resulting from (i) the static inhomogeneity of the magnetic field, and (ii) three static and three dynamic inhomogeneities of the electric field;
2. **Important:** uni-directional, bi-directional and self-crosstalk of pad signals;
3. **Less important:** different amplification factors and non-linearity of preamplifier response;
4. **Somewhat important:** the 100 ns problem;
5. **Marginally important:** the 3.6 μs problem.

A concerted effort has been made by our group, to correct adequately for distortions [1, 2, 3, 4], for crosstalk [5, 6], for different amplification factors of the preamplifiers [7], and for the 100 ns and 3.6 μs problems [8]. The overall conclusion is: all above hardware deficiencies can be understood and corrected by appropriate software.

The only persistent inadequacy of the TPC is that $\mathrm{d}E/\mathrm{d}x$ cannot be determined with sufficient accuracy to permit pion/electron separation. Fortunately, this entails no loss of physics since the barrel RPC's which surround the TPC, show better than design-performance for particle identification.

In this paper, we discuss the algorithms to reconstruct correctly the $r \cdot \phi$ and z coordinates of clusters. While there is in comparison with earlier work [9] not much difference of opinion on the $r \cdot \phi$ coordinate, there is quite some on the z coordinate: (i) we use the **entire longitudinal signal shape** for every pad of a cluster, (ii) we take the **charge-weighted time average** of the cluster's 'leader' pad as estimator of the cluster's z position, and (iii) we pay due tribute to the consequences of **self-crosstalk**.

We present in this paper our algorithms for accepting clusters as 'good' clusters, and for the reconstruction of the $r \cdot \phi$ and z cluster coordinates, including provisions for 'bad' pads and pads next to sector boundaries, respectively.

This paper is concerned solely with the reconstruction of the cluster coordinates *per se*. In a sequel to this paper, we shall discuss the assignment of coordinate errors.

We wish to stress that in the discussion of 'statistical precision' versus 'systematic precision', we have a strong preference for systematic precision. This is because the final coordinate precision of the HARP TPC will not be dominated by statistical effects but by the accuracy with which systematic effects will be corrected.

2 Longitudinal cluster shape

2.1 Basic considerations

We recall briefly the essentials of the longitudinal cluster shape, which were discussed already in an earlier paper [6]. The longitudinal signal shape is a convolution of

1. the **longitudinal diffusion** which causes a Gaussian jitter in the arrival of the primary electrons at the sense wires;
2. the **electronics response** of the preamplifier which is asymmetric in time (the rise is faster than the fall); and
3. the **inclination of the track** with respect to the pad plane, in conjunction with the relatively large radial size of ~ 16 mm of a TPC pad.

We remind that there is an important fourth effect to be taken into account: pad-specific **self-crosstalk** lengthens the longitudinal signal shape. We shall come back to this effect in Subsection 2.2.

The longitudinal diffusion increases with the square root of the driftlength. It is approximated by a Gaussian with an r.m.s. of ~ 4 mm $\cong 80$ ns at 1 m driftlength, valid for the TPC gas of 91% argon and 9% methane at atmospheric pressure [10].

The preamplifier response to a δ-charge input is approximated by the function [11]

$$f(t) = t^2 \exp\left(-t/100\right) ,$$

where t is the time in ns. Figure 1 shows a simulation of the longitudinal signal shape with the above preamplifier response, diffusion from a drift distance of 1 m, and for a track with 45° polar angle. The plot is for the sole purpose of illustration, a precise reproduction of data is neither intended nor warranted; the plot shows the signal **before** sampling, digitization and pedestal subtraction. Figure 1 also demonstrates that the assignment of a z cluster coordinate from the cluster time, employing the drift-time/drift-distance relation, is a non-trivial issue. We concluded that **the best estimator which can be obtained from the longitudinal signal shape, is the charge-weighted time average.**

Other estimators like the leading edge of the pulse, or the trailing edge, were discarded as too biased (neither the leading-edge nor the trailing-edge estimator takes track inclination into account). The average between leading and trailing edges was discarded as less reliable than the charge-weighted time average.

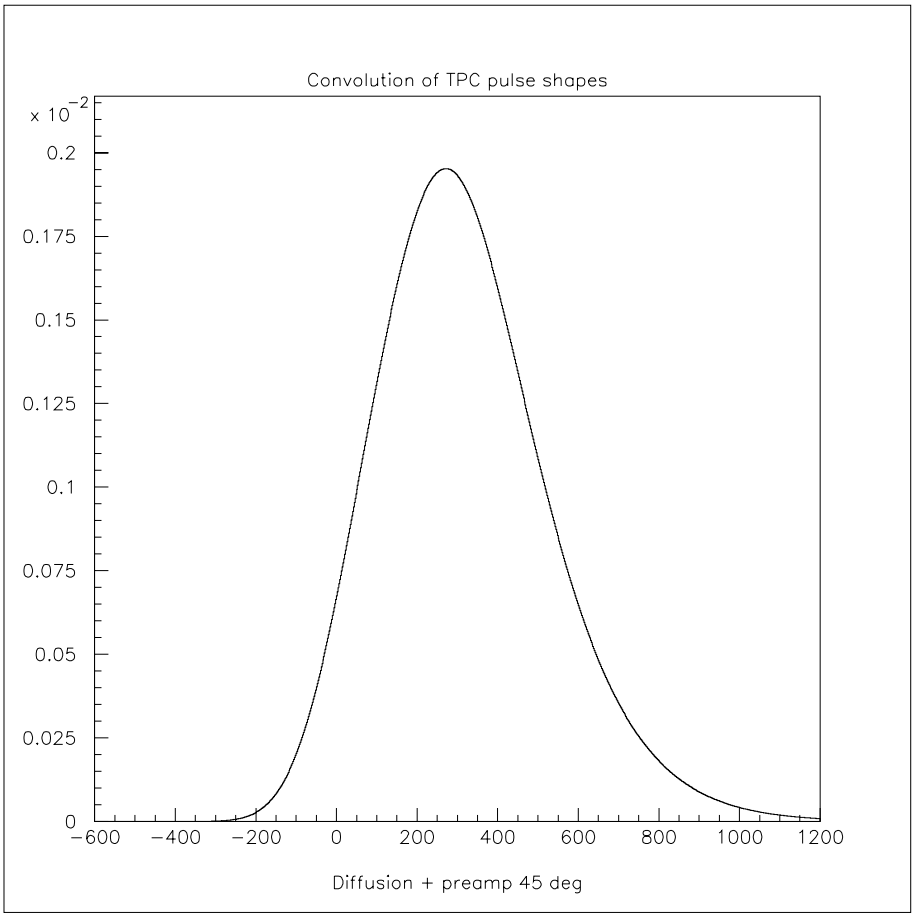

Figure 1: Simulation of the longitudinal signal shape, including preamplifier response and diffusion from a drift distance of 1 m, for a track with 45° polar angle.

The simulation shows that the charge-weighted time average is delayed by ~ 300 ns w.r.t. the real time of the cluster. Accordingly, the charge-weighted time average will have to be reduced by ~ 300 ns to obtain a correct estimate of the cluster time.

The charge-weighted time average needs the prior definition of a charge threshold. For this, 20% of the maximum of the charge was chosen. This choice reflects a compromise between 'not too low' in order to minimize systematic errors from (i) low-charge tails from unavoidable inadequacies of the CDM crosstalk correction, and (ii) inadequacies of the pedestal subtraction; and 'not too high' in order not to lose unnecessarily statistical precision.

Figure 2 shows for different points (r, z) inside the active TPC volume the effect of track inclination on the pulse length, defined as the difference between the leading and the trailing edges at the level of 20% of the maximum charge. For this purpose, only tracks emanating from a target located at $z = 0$ were considered. The strong dependence of the pulse length on track inclination, especially at very small or very large polar angles, is apparent.

The graphical content of Fig. 2 is numerically reproduced in Table 1 which gives for 15 bins in z and six bins in r the estimation of the pulselength at 20% of the maximum pulseheight, for tracks emanating from $z = 0$. This matrix of numbers is available at

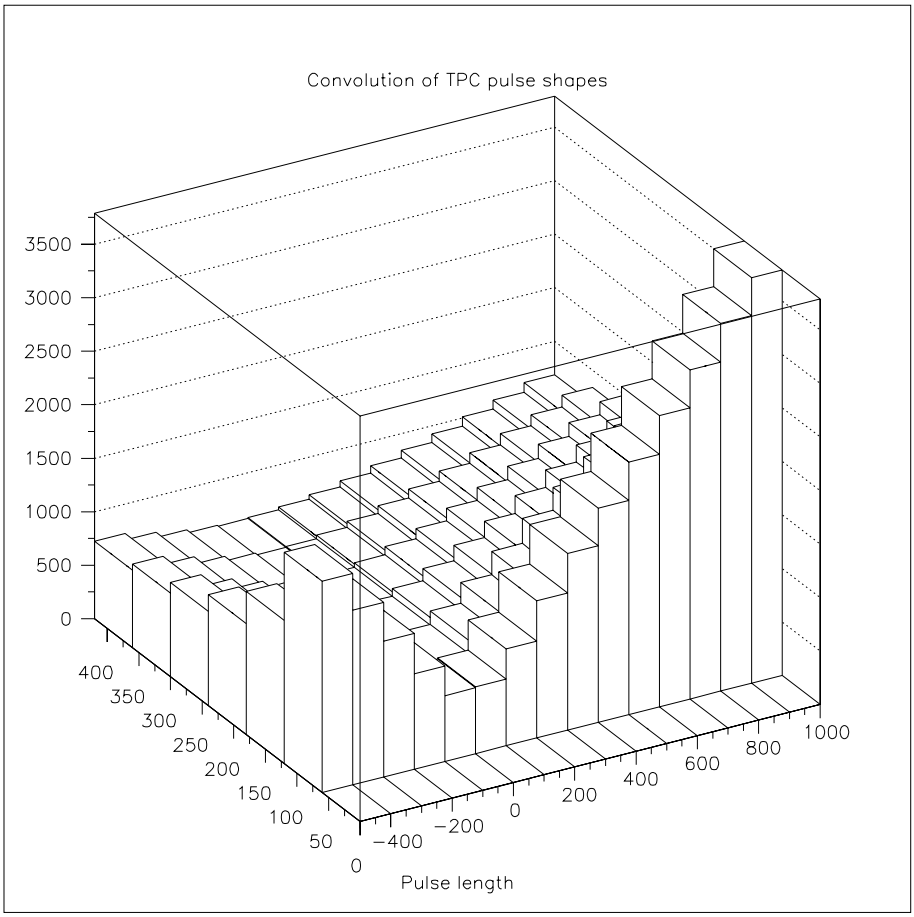

Figure 2: Simulation of the time difference [ns] between the leading and trailing edges of longitudinal pulseshapes at the 20% threshold level, at different points (r, z) within the active TPC volume, for tracks emanating from a target at $z = 0$.

http://cern.ch/dydak/TPCpulsecorrmatrix

Figure 3 shows for the same points in the active TPC volume, and for the same tracks, the correction to be applied to the charge-weighted time average above a threshold of 20% of the maximum pulseheight, in order to obtain a correct estimate of the true time of the cluster. The main point is not so much the precise numerical value which must anyway be tuned from data; rather, it is the approximate constancy across the active TPC volume, and the approximate size of the correction.

The graphical content of Fig. 3 is numerically reproduced in Table 2 which gives for the same 15 bins in z and six bins in r as above an estimation of the time correction to the charge-weighted time average above a threshold of 20% of the maximum pulseheight, for tracks emanating from $z = 0$. Also this matrix of numbers is available at

http://cern.ch/dydak/TPCpulsecorrmatrix

Table 1: Estimation of the pulselength [ns] at 20% of the maximum pulseheight, at given points (r, z) inside the active TPC volume, for tracks emanating from $z = 0$.

	90 mm	150 mm	210 mm	270 mm	330 mm	390 mm
-450 mm	1976	1338	1036	880	782	724
-350 mm	1632	1108	884	768	704	662
-250 mm	1268	888	744	674	636	614
-150 mm	892	700	636	606	592	582
-50 mm	614	586	578	576	572	572
50 mm	622	594	586	584	582	582
150 mm	906	718	656	628	614	606
250 mm	1282	910	770	704	670	648
350 mm	1650	1130	914	802	742	704
450 mm	2002	1366	1068	918	826	772
550 mm	2358	1582	1236	1038	920	844
650 mm	2720	1796	1402	1166	1020	924
750 mm	3074	2012	1562	1298	1126	1010
850 mm	3436	2230	1712	1428	1230	1096
950 mm	3790	2446	1868	1550	1338	1186

2.2 Effects from self-crosstalk of pad signals

Beyond the effects of longitudinal diffusion, electronics response and track inclination, the longitudinal pulse profile is affected by uni-directional, bi-directional and self-crosstalk. While distortions of the longitudinal pulseshape from uni-directional and bi-directional crosstalk are corrected by the CDM algorithm [5, 6], distortions from self-crosstalk are, by construction, not.

Throughout this paper, when we show results from clusters along physics or cosmic-muon tracks, the CDM crosstalk correction has been applied beforehand, i.e. the effects of uni-directional and bi-directional crosstalk have been systematically removed. However, the results from pulser data stem from pulse-injection events which have not been corrected for crosstalk. *A priori*, this difference may introduce a bias in conclusions.

Obviously, uni-directional crosstalk does not matter for the purpose under consideration, but bi-directional crosstalk does. Although possible in principle, we considered it too time-consuming to correct all pulser events for crosstalk. To avoid any bias in conclusions, pads which are known to participate in bi-directional crosstalk have been excluded in subsequent plots of pulser data (though we convinced ourselves that conclusions do not change whether or not these pads are excluded).

For the record, we give here the list of pads which participate in bi-directional crosstalk (the numbering within a sector is between 1 and 662):

Sector 1:
14 15 17 18 67 68 69 70 71 96 97 122 123 242 287 395 423 424 430 431 641 642 648 649

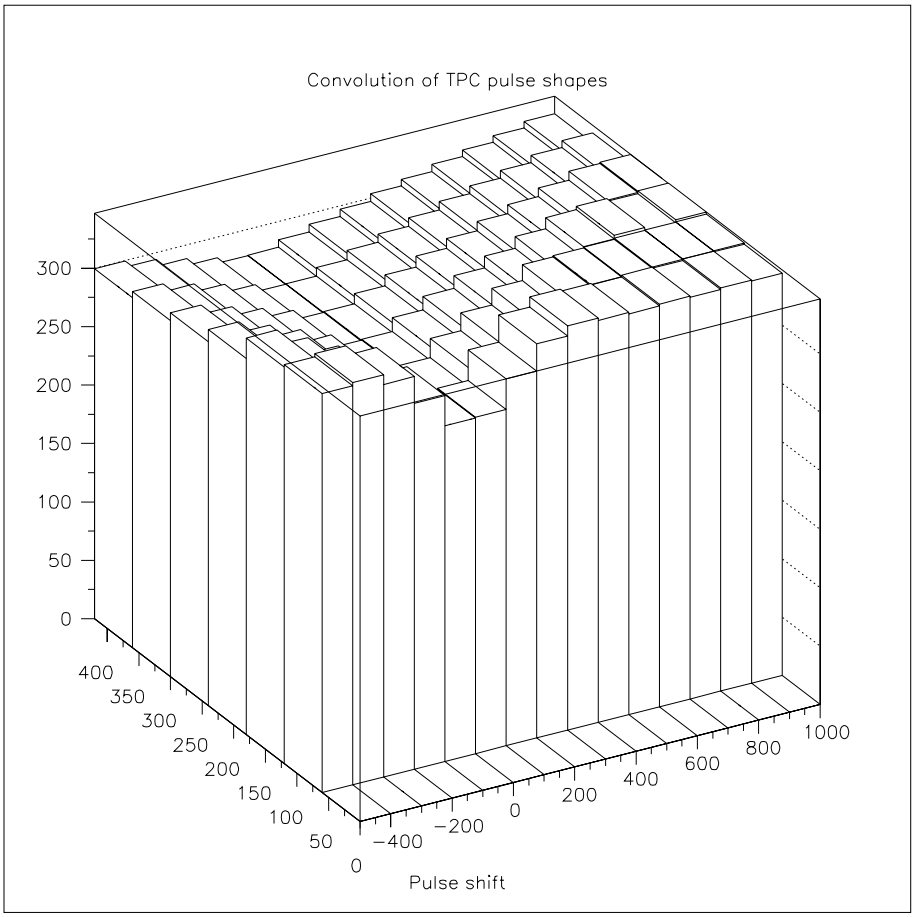

Figure 3: Simulation of the time [ns] to be subtracted from the charge-weighted average time above a 20% threshold, in order to obtain a correct estimate of the cluster's true time.

Sector 2:
14 15 67 68 118 119 120 144 145 149 150 242 287 395 423 424 430 431 529 555 556 641 642

Sector 3:
17 18 42 70 71 96 97 242 286 287 648 649

Sector 4:
14 15 17 18 67 68 70 71 144 145 242 286 287 395 423 424 430 431 648 649

Sector 5:
14 15 17 18 43 67 68 70 71 96 97 99 122 123 144 145 149 150 423 424 555 556 641 642

Sector 6:
14 15 17 18 67 68 70 71 242 287 641 642 648 649

This list of pads which participate in bi-directional crosstalk, is available at

http://cern.ch/dydak/bidirectionalpads.txt

Figure 4 shows from pulser data the FWHM of the pulses for all pads, numbered from 1 to

Table 2: Estimation of the time correction [ns] to the charge-weighted time average above a threshold of 20% of the maximum pulseheight, at given points (r, z) inside the active TPC volume, for tracks emanating from $z = 0$.

	90 mm	150 mm	210 mm	270 mm	330 mm	390 mm
-450 mm	342	340	322	312	305	300
-350 mm	344	327	313	304	298	294
-250 mm	336	313	302	295	291	288
-150 mm	313	298	291	287	284	282
-50 mm	287	282	280	278	278	277
50 mm	288	283	280	279	278	278
150 mm	314	299	292	288	285	284
250 mm	337	314	304	297	293	291
350 mm	346	328	315	306	301	297
450 mm	344	342	325	315	308	304
550 mm	343	343	335	323	315	309
650 mm	345	344	344	331	322	315
750 mm	344	345	347	338	328	321
850 mm	345	345	345	346	335	327
950 mm	344	346	347	347	341	332

3972 (to create this plot, existing files of histograms with a 100 ns binning of the longitudinal profile were exploited). One observes three distinct bands in the FWHM which henceforth we refer to as 'Type1' (small FWHM), 'Type2' (medium FWHM) and 'Type3' (large FWHM).

Type1 pads have 'weak' self-crosstalk, Type2 pads have 'medium ' self-crosstalk, while Type3 pads have 'strong' self-crosstalk.

With a view to comparing pulser data with cosmic-muon data, we show in Fig. 5 the same as in Fig. 4 but for clusters along cosmic-muon tracks. Again three distinct bands show up, albeit at larger FWHM values, as expected.

From Fig. 4 it is easy to determine cuts in FWHM which classify pads according to the Type1, Type2 and Type3 self-crosstalk categories. These have the **average** longitudinal pulse shapes (with 100 ns binning) shown in Fig. 6. While the three plots with a linear scale (left) demonstrate the difference in the **width** of the signals, the three plots with a logarithmic scale (right) emphasize the behaviour of the pulse tails. One conclusion to be drawn is that in some 'afterpulse' pads, there is activity after the main pulse.

This effect of pads with self-crosstalk is the cause of what we came to call 'shadow clusters': non-physical secondary clusters which follow typically 1–2 μs after a first cluster.

Shadow clusters are a nuisance for pattern recognition and therefore should be suppressed. After investigation of the problem, we concluded that afterpulse pads should not be determined from pulser data alone (in order to avoid classification on the basis of very small pulseheights), but mainly from physics data. We shall come back to this problem in Subsection 2.4.

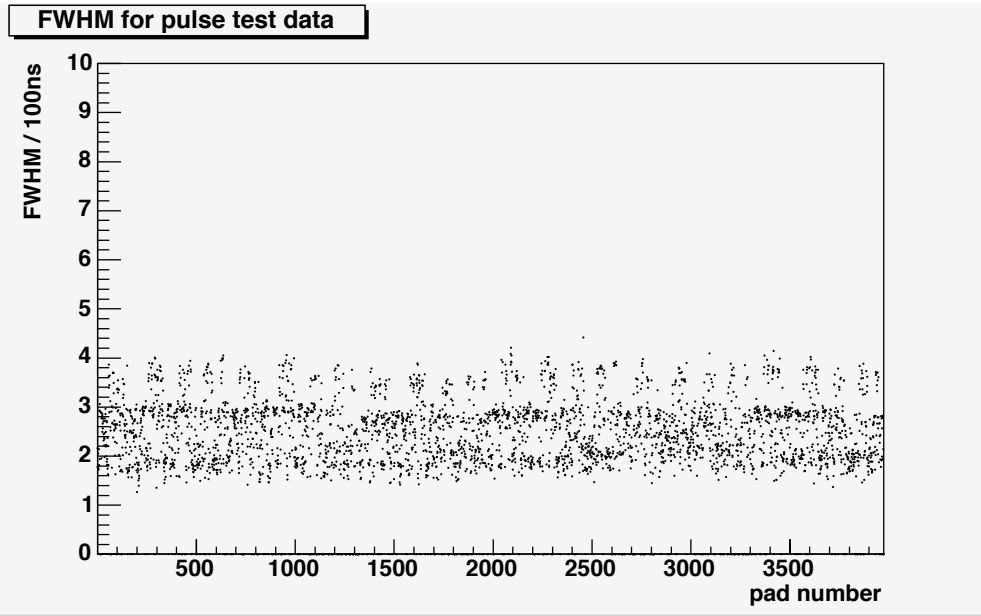

Figure 4: Longitudinal FWHM of pulses from pulser data, as a function of pad number from 1 to 3972.

In order to corroborate the self-crosstalk conjecture, we study the correlation of Type1, Type2 and Type 3 pulseshapes with various parameters. If self-crosstalk were an intrinsic feature of one or the other of the four preamplifiers located together in one physical chip, one would observe a strong correlation between Type1, Type2 and Type2 pulseshape on the one hand, and the preamplifier input number which ranges from one to four, on the other hand. Figure 7 demonstrates that this is not the case: the same preamplifier input number shows up in all three types of pulseshapes [12]. Therefore, self-crosstalk is **not** an internal feature of one or the other preamplifier located in the same chip.

Is self-crosstalk a feature correlated with the preamplifier chip as a whole? No strong correlation is visible in Fig. 8: the same preamplifier-chip number shows up in all three types of pulseshapes (the limitation to the first 16 chips in each sector is for plotting purposes only; we checked that the remaining 17 to 186 chips behave the same way).

The conclusion is that self-crosstalk is, like unidirectional and bidirectional crosstalk, an effect **solely caused by the layout of the printed paths inside the 5-layer motherboard.** Contrary to what was believed before (see, for example, Ref. [13]), there is nothing like 'chip' crosstalk; there is only 'motherboard' crosstalk.

This conclusion is confirmed by the physical location of pads of Type1, Type2 and Type3 pulseshape, shown in Figs. 9, 10 and 11, respectively. These plots demonstrate that the only strong correlation of Type1, Type2 and Type3 pulseshape is with the physical location of the pad on the motherboard. Especially Fig. 11 is telling because it reflects the repetitive pattern of analogous printed paths with increasing radius and pad row number, respectively.

How reproducible is self-crosstalk between TPC sectors? We show in Fig. 12 the r.m.s. dispersion of FWHM between sectors, as a function of the pad number within a sector. If each pad exhibited the same level of self-crosstalk, the dispersion would take its minimal

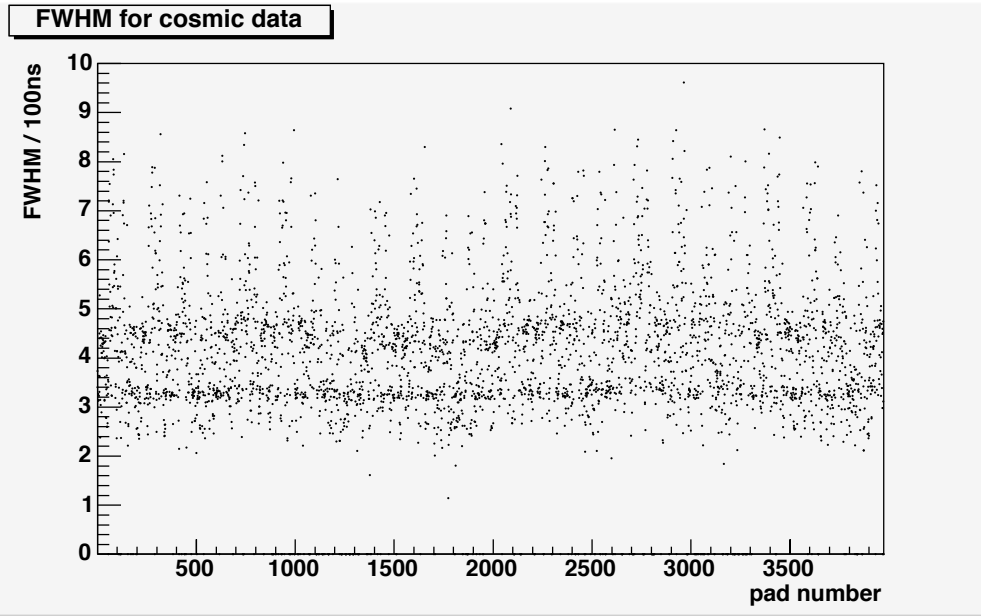

Figure 5: Longitudinal FWHM of pulses in clusters of cosmic muons, as a function of pad number from 1 to 3972.

value of ~ 0.1. This is generally not the case; hence **self-crosstalk is not uniform across the TPC sectors.**

This conclusion confirms the experience from uni-directional and bi-directional crosstalk which was shown not to be uniform across TPC sectors [6]. We conjecture that the mother-board is non-uniform in the distances between sub-layers at the micron level, which on the one hand can plausibly be expected, and on the other hand is sufficient to cause sizeable differences in coupling capacitances which are the origin of crosstalk.

The overall conclusion is that like for uni-directional and bi-directional crosstalk, **the mapping of the effects of self-crosstalk must be done individually for all 3872 pads.**

Figure 13 gives the distributions of FWHM for Type1, Type2 and Type3 pulseshapes. There is satisfactory similarity between pulser data (which serve to define the three categories) and cosmic-muon data (where the categories are utilized).

Self-crosstalk requires that the z coordinate of a cluster which is determined from the charge-weighted time average of its 'leader' pad (see Subsection 2.3), must receive specific corrections according to the Type1, Type2 and Type3 category of its pulseshape.

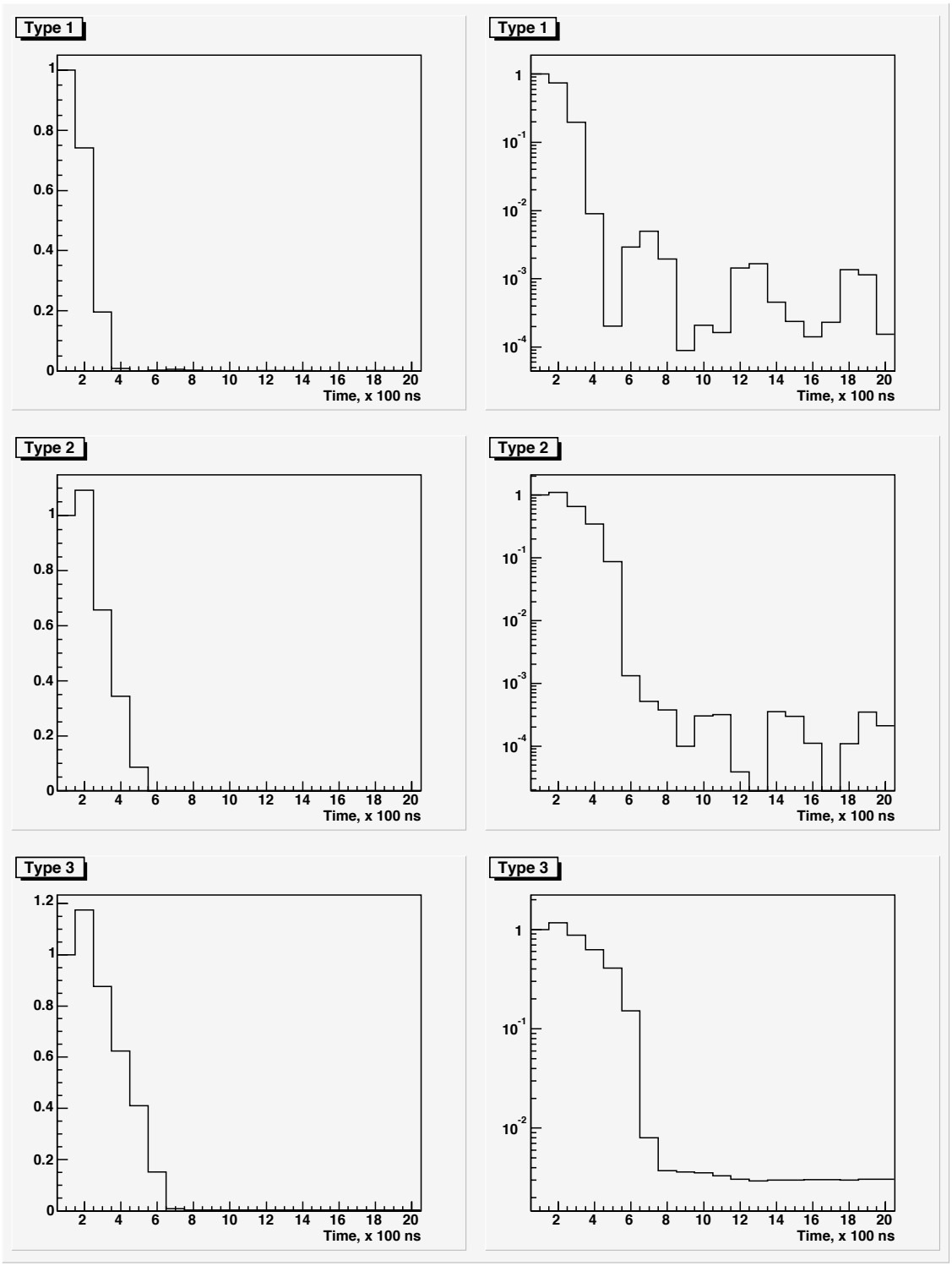

Figure 6: Average longitudinal pulseshapes of pulser data, for the Type1 (top), Type2 (middle) and Type3 (bottom) self-crosstalk categories, on a linear (left) and on a logarithmic scale (right).

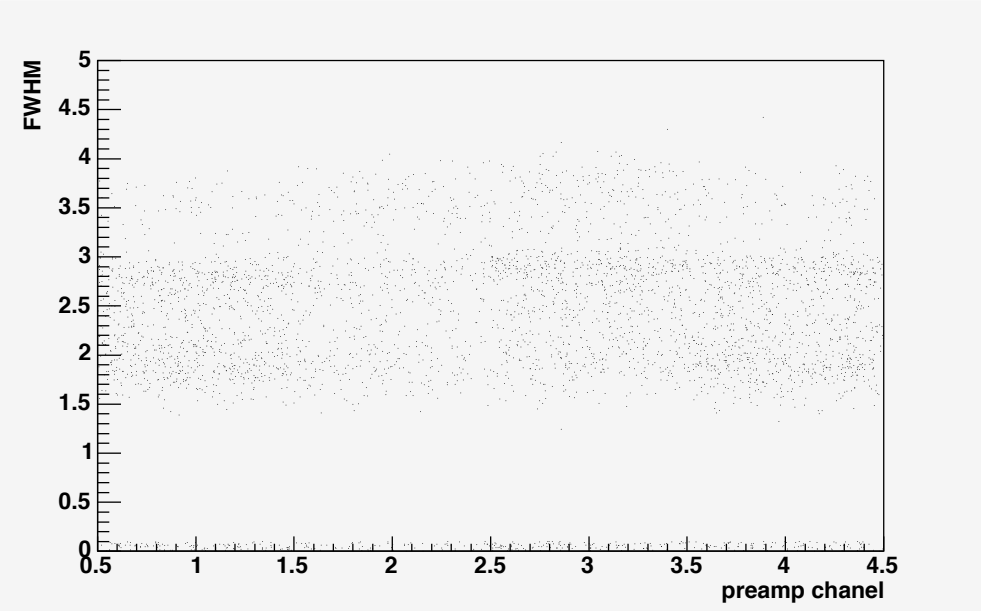

Figure 7: Longitudinal FWHM of pulseshapes from pulser data versus preamplifier input number 1 to 4.

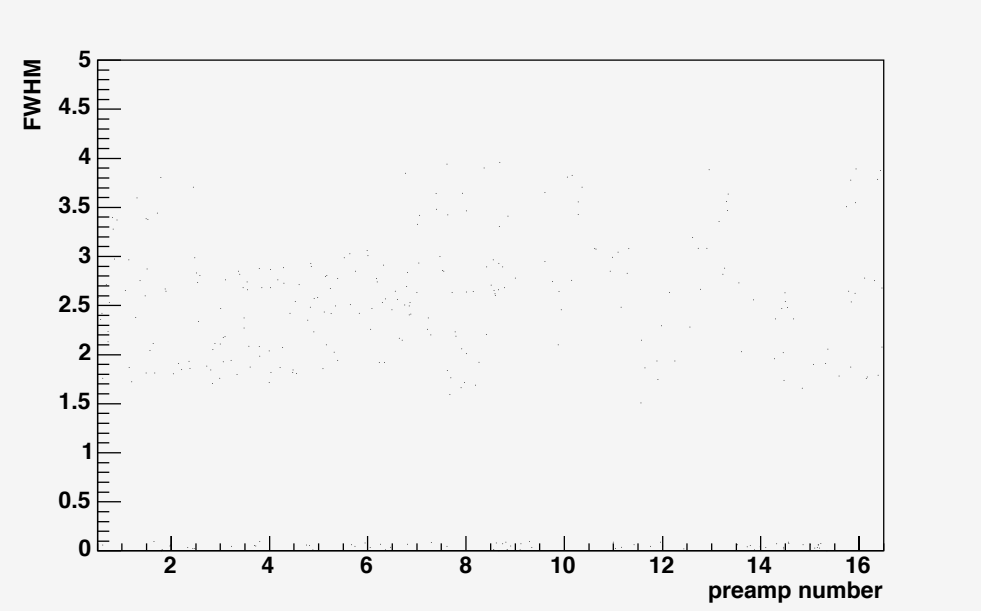

Figure 8: Longitudinal FWHM of pulseshapes from pulser data versus preamplifier-chip number 1 to 16.

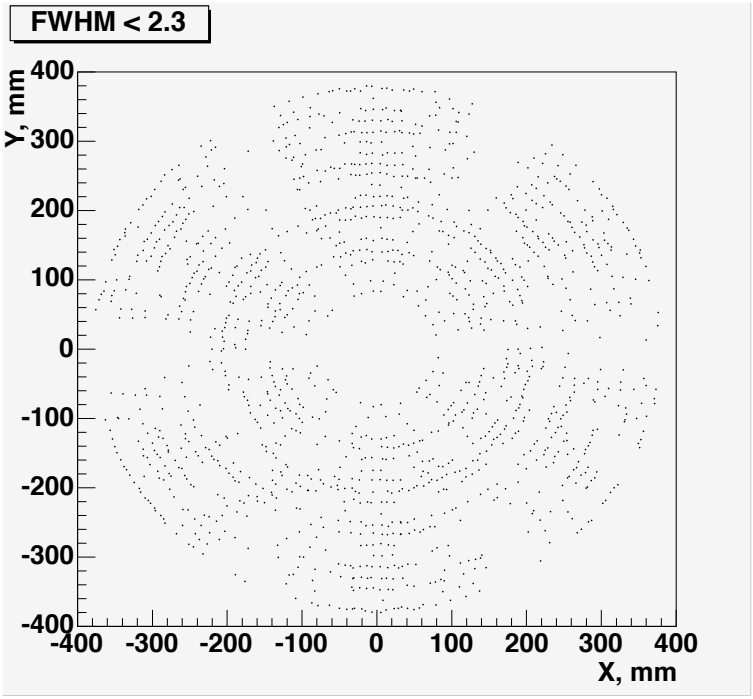

Figure 9: Position of pads with Type1 pulseshape.

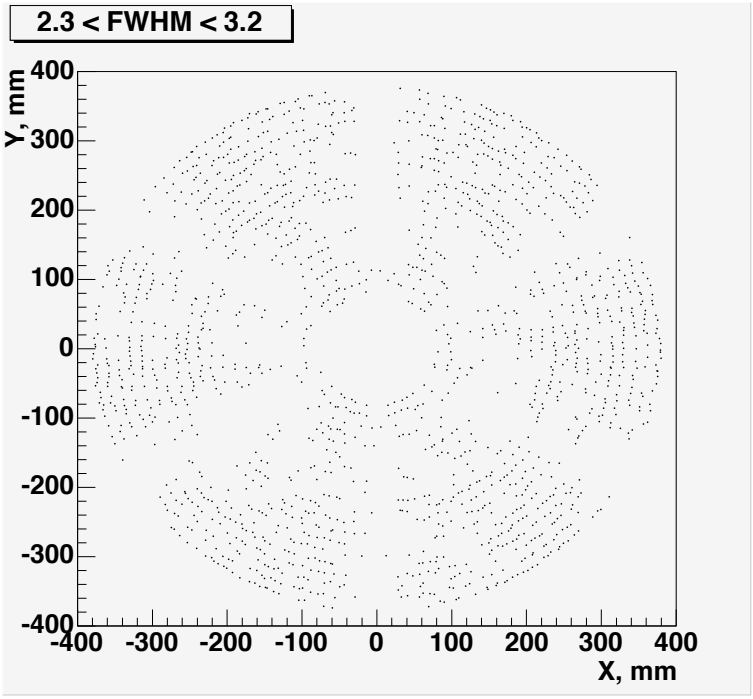

Figure 10: Position of pads with Type2 pulseshape.

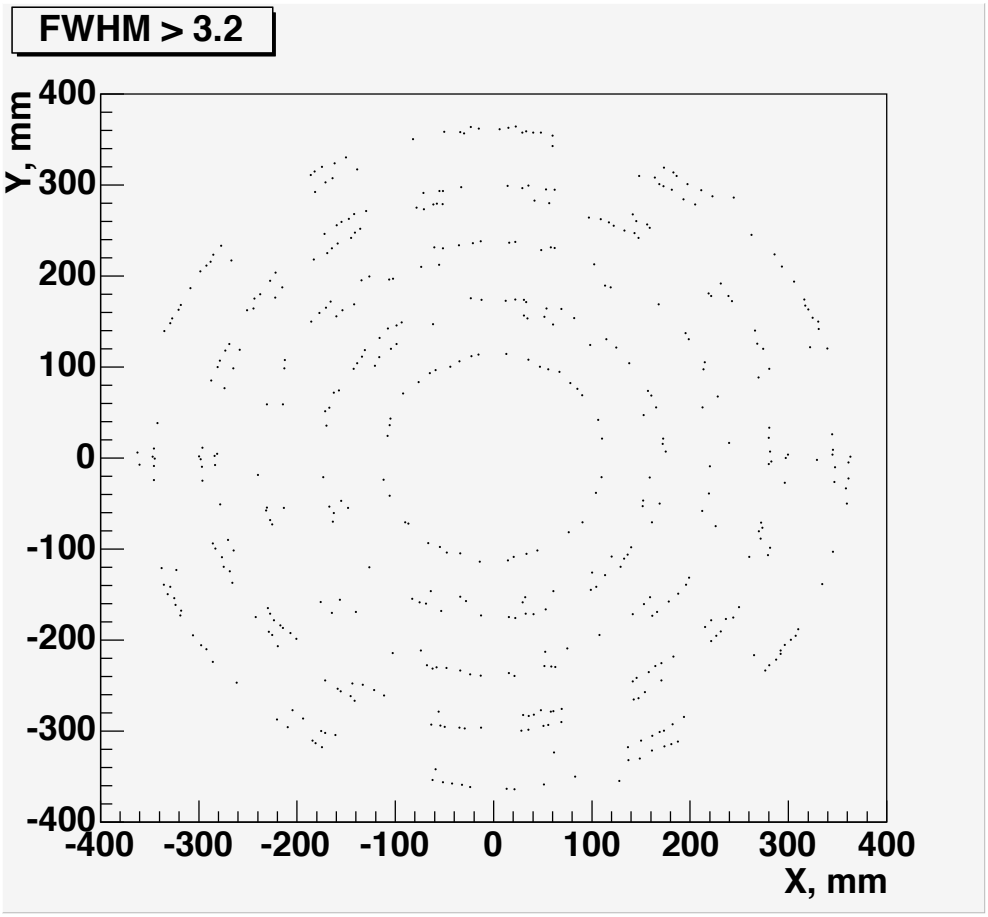

Figure 11: Position of pads with Type3 pulseshape.

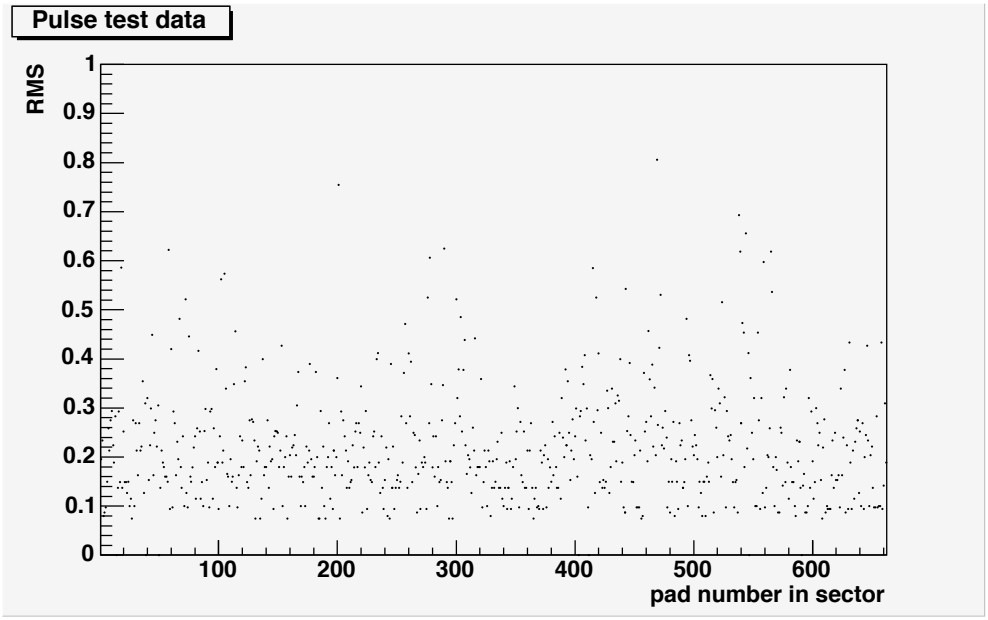

Figure 12: r.m.s. dispersion of the FWHM of the longitudinal pulseshape of pulser data across the six TPC sectors.

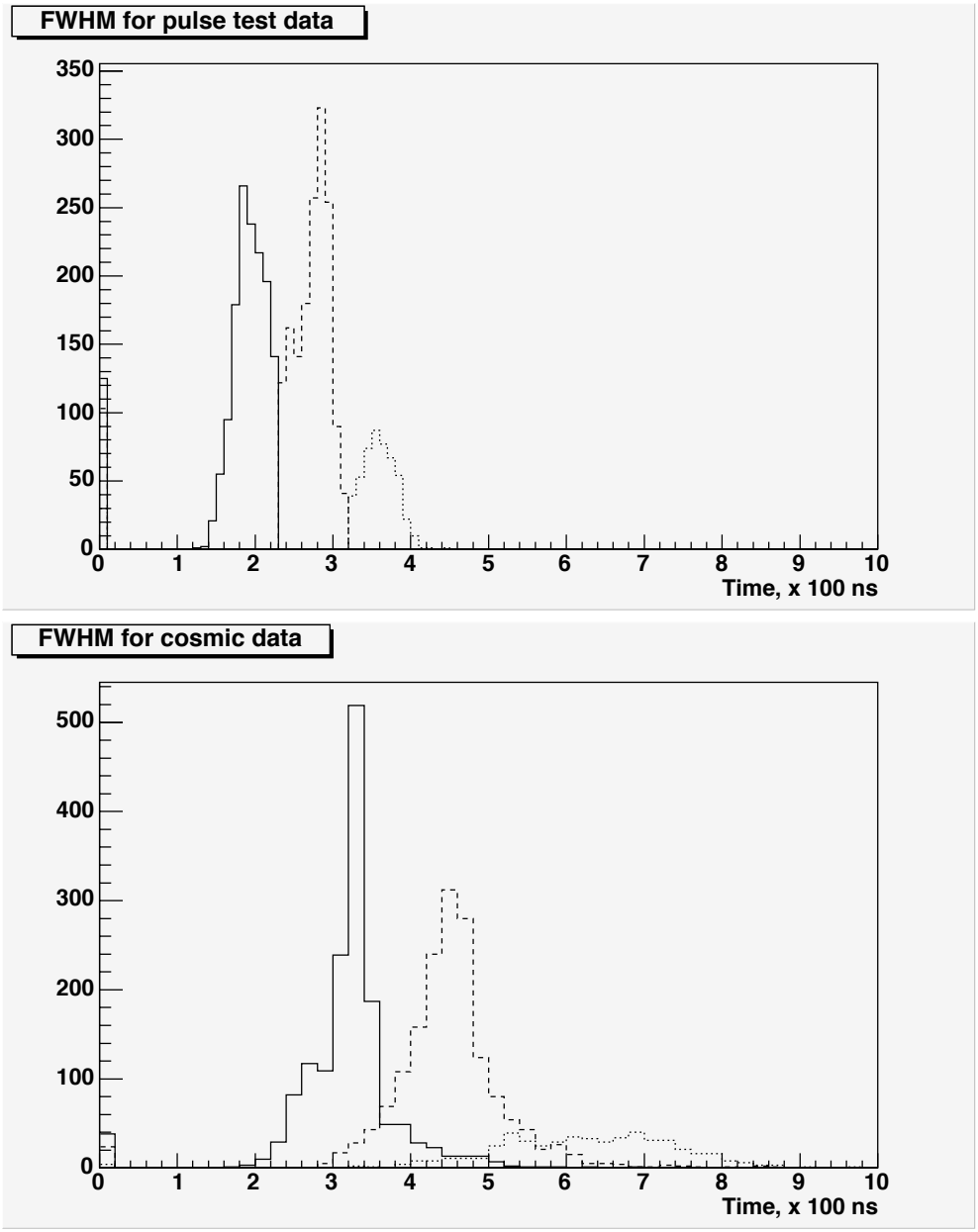

Figure 13: Distribution of FWHM of the longitudinal pulseshape of pulser data (top) and cosmic-muon data (bottom), for the Type1 (full line), Type2 (broken line) and Type3 (dotted line) self-crosstalk categories.

2.3 Longitudinal coordinate assignment

Since the three pads which maximally contribute to a cluster (see Section 3) will in general belong to different pulseshape categories, the z position is determined from the 'leader' pad of a cluster only: that is the pad with the largest time-integrated charge. This choice has two motivations: (i) the largest charge will be least affected by inadequacies of crosstalk correction and pedestal subtraction, and (ii) the need of combining the z information from pads with different Type1, Type2 and Type3 pulseshape categories is avoided.

The z position of a cluster is the charge-weighted time average of its leader pad; the average is taken over all 100 ns samples from the time of the first sample with pulseheight above 20% of the maximum pulseheight, until the time of the last sample with pulseheight above the 20% threshold which is followed by two consecutive empty samples; the maximum accepted number of samples is 40.

The above concept is not optimal for clusters along beam-muon tracks; however, the main interest in such clusters is to recognize beam-muon tracks, not to reconstruct their trajectory with ultimate systematic accuracy.

Figure 14 shows the experimental counterpart of Fig. 2: the pulse length of the leader pad of clusters along physical tracks emanating from a thin target at $z = 0$, as a function of the cluster's (r, z) position. The comparison of the two figures is reassuring (note that the simulation had no provisions for sampling, digitization, pedestal subtraction, and acceptance): the same basic dependence of the pulse length on the (r, z) position that was simulated, and similar pulse lengths, show up in the data.

Self-crosstalk calls for corrections of the charge-weighted time averages of clusters. At this point of our work, these corrections are not yet numerically finalized because they depend on the pattern recognition algorithm which selects clusters along a track candidate. In turn, the pattern recognition depends on the prior application of the final distortion corrections. Therefore, for the time being, we restrict ourselves to point out the **approximate** size of the correction. We show in Fig. 15 the z residuals of physics tracks in +8.9 GeV/c data as a function of the FWHM of the clusters' leader pads as determined in the pulser data. The correction for the 100 ns problem [8] has been applied to the data. Figure 15 demonstrates that the pad-dependent correction caused by self-crosstalk is in the range of several millimetres, and thus constitutes an important correction of the z coordinate.

After the self-crosstalk correction, all pads are at equal footing with respect to the still missing global correction of the z coordinate: the additive constant to be determined from data, which we expect to be ~ 300 ns $\simeq 15$ mm (see Table 2). This (approximate) constant will be determined from the fit of physics tracks emanating from a thin target the geometrical position of which is precisely known. However, for this determination, the prior application of corrections for static and dynamic TPC distortions is mandatory, for their shift of the cluster positions in the radial direction.

In order to avoid problems (i) with laser signals, and (ii) with partially reconstructed pulseshapes, the cluster time is requested to be larger than 5 μs and smaller than 30 μs.

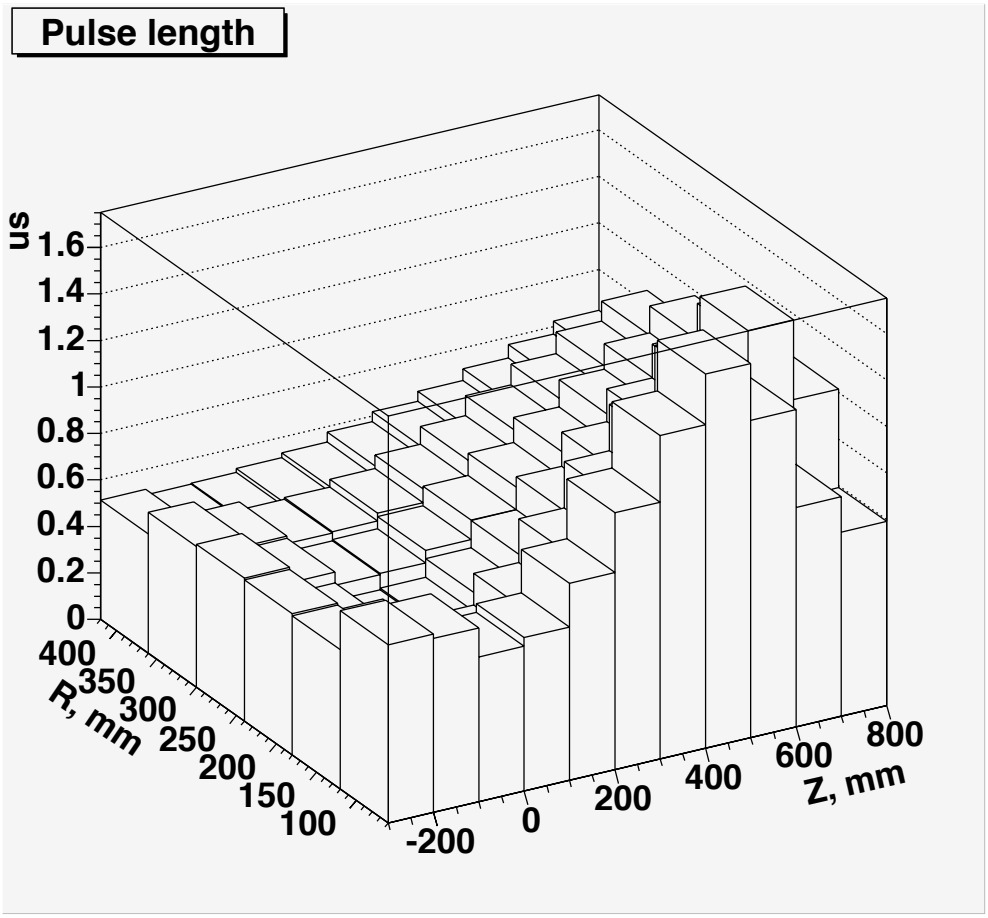

Figure 14: Pulse length [μs] of the leader pad of clusters along physics tracks emanating from a thin Be target at $z = 0$, as a function of the cluster's (r, z) position.

2.4 Handling of shadow clusters

The origin of shadow clusters is self-crosstalk in certain afterpulse pads. 36 such pads can be easily identified in pulser data. Since the afterpulse phenomenon is independent of the physical configuration of events, the bulk of afterpulse pads can effectively be determined from physics events: their signature is a higher than normal occurrence of clusters with a distance less than 1 μs to the first cluster.

This procedure resulted in further 285 pads. The total list of 321 afterpulse pads reads as follows:

Sector 1:
58 68 118 119 151 201 237 257 259 260 262 274 275 276 279 281 286 289 290 294 418 421 422 423 425 429 431 432 433 437 439 440 444 529 540 546 550 552 555 557 558 559 576 582 586 612 624 625 627 629 630 631 632 636 637 643 645 649 655

Sector 2:
14 60 119 147 254 257 260 261 266 269 279 281 286 288 291 417 421 422 423 425 438 439 440 455 540 544 552 555 557 558 559 625 627 636 643 644 648

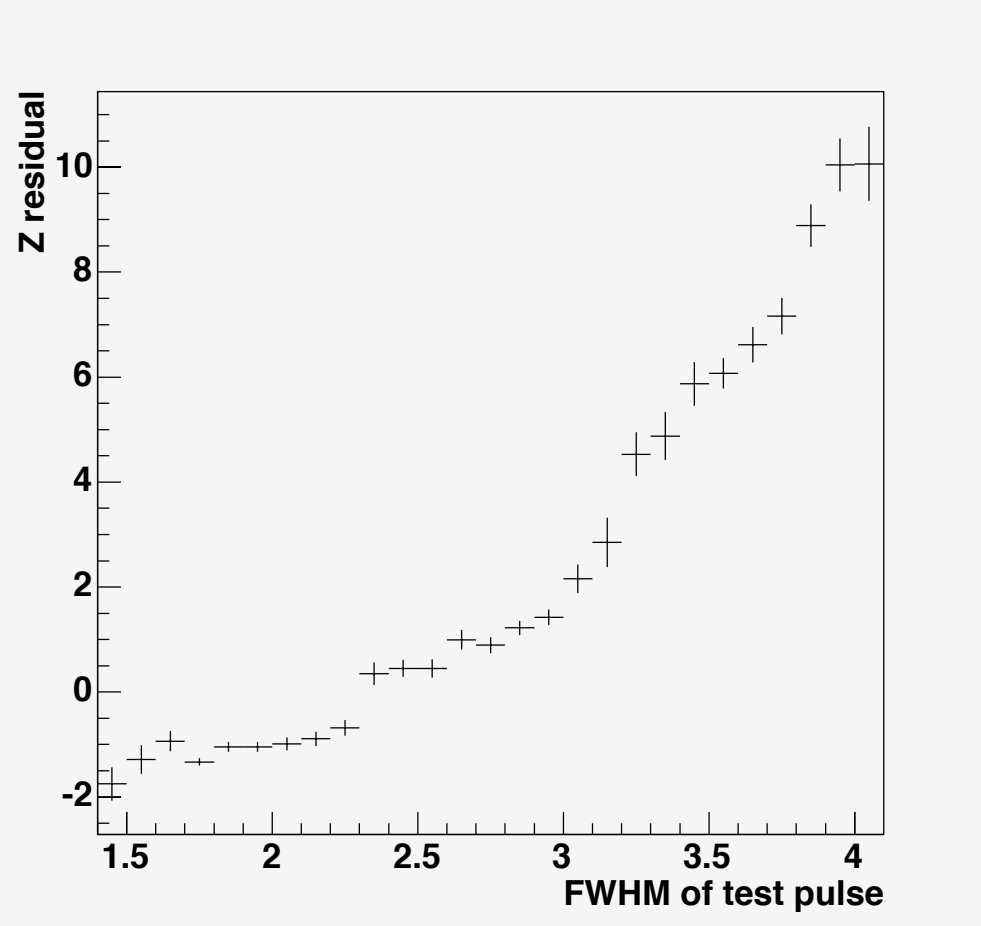

Figure 15: z residuals [mm] from the fit of physics tracks versus the FWHM of pads as determined from pulser data.

Sector 3:
14 67 68 93 100 118 119 120 255 259 260 271 274 275 276 281 287 288 289 293 322 331 420 425 432 437 439 440 444 455 458 517 529 550 551 552 553 554 556 558 560 572 611 612 624 625 628 629 630 631 632 634 636 637 642 644 645

Sector 4:
14 68 86 102 117 118 119 120 153 154 155 242 257 274 275 276 278 279 286 287 289 293 322 416 421 422 425 439 440 443 455 529 540 544 551 552 553 554 555 556 557 558 565 569 576 580 612 624 625 626 627 629 634 636 637 641 642 643 644 645 649

Sector 5:
98 117 118 119 154 257 260 262 274 275 276 286 287 289 291 310 330 417 420 421 423 425 429 430 431 433 444 453 467 542 551 552 555 557 559 570 576 590 624 625 627 634 636 643 644 649 655

Sector 6:
64 68 71 95 96 97 100 113 118 119 120 257 274 275 276 279 281 286 288 293 330 332 333 417 420 421 422 423 431 432 433 437 439 444 455 468 529 540 542 544 550 552 553 555 556 560 576 612 624 625 627 629 630 631 632 636 637 641 643 644

The list of afterpulse pads is available at

http://cern.ch/dydak/afterpulsepads.txt

In the reconstruction of clusters with a afterpulse pad as leader pad, all clusters following the first cluster with the same leader pad at a distance of less than 1 μs are discarded.

3 Transverse cluster coordinate

A cluster can minimally consist of one pad, and can maximally consist of three adjacent pads.

3.1 Transverse coordinate assignment

The r coordinate of a cluster is the mean radius of the cluster's pad ring.

The $r \cdot \phi$ coordinate of a cluster of pads within one pad row is determined as the charge-weighted average of maximally three adjacent pads:

$$r \cdot \phi = \frac{\sum_1^N Q_i \times (r \cdot \phi)_i}{\sum_i^N Q_i} \ ,$$

where $1 \leq i \leq N$, with N denoting the number of pads of which the cluster consists (minimally 1, maximally 3). Q_i is the time-integrated charge and $(r \cdot \phi)_i$ the pad centre position of the i.th pad.

In an earlier paper [6], we discussed the TPC's 'effective pad response function', and concluded that for any radial track, only **three** pads carry significant pulseheight (we intentionally ignore charges in more distant pads since these charges cannot be large and are likely to cause more trouble with systematic error than they do any good).

Of course, the above argument does not apply for non-radial tracks. The most extreme example are the large-radius sections of a helical track. However, the recipe works well also in such cases: rather than combining many pads within one pad row into one cluster, we generate several clusters, each with not more than three pads.

The time-integration in pads adjacent to the leader pad, which tend to have small charges and therefore 'holes' in their sampling trains, takes place over the **same** pulselength as in the leader pad: the $r \cdot \phi$ position resolution depends critically on the **complete** collection also of small charges in pads adjacent to pads with large charges.

4 'Good' versus 'bad' clusters'

Plots shown in this Section refer to +8.9 GeV/c data, where 'bad' pads were eliminated according to the criteria discussed in Ref. [8].

In line with our *leitmotiv* that reduction of systematic error is more important than gain in statistical precision, we apply relatively strong cuts to retain only 'good' clusters for pattern recognition. The selection criteria depend on the number of pads per cluster.

For 1-pad clusters, our variable of choice is the time-integrated charge divided by the number of 100 ns samplings (henceforth called 'average charge/sampling'), after crosstalk correction and pad equalization. Fig. 16 shows its distribution separately for the six sectors.

In Fig. 16, all 1-pad clusters are shown, **without regard of association with a track**. There is a sizeable contribution of non-physical clusters from noise and imperfections of crosstalk correction, so different sectors may behave differently because of intrinsic differences in crosstalk, pad signal amplification and pedestal subtraction, which will lead to different fractions of 1-pad, 2-pad and 3-pad clusters, despite of later pad equalization. Especially sector 2 looks different in comparison with the other five sectors.

We retain 1-pad clusters as 'good' clusters if their average charge/sampling is above 25 units.

For 2-pad and 3-pad clusters, we show in Fig. 17 the scatterplot of the average charge/sampling of one pad versus the average charge/sampling of its neighbor pad (for a 3-pad cluster, two entries are made per cluster). All clusters are shown, after crosstalk correction and pad equalization. The non-physical spike which sticks out at low average charge/sampling, is eliminated by the following requirement:

- in 2-pad clusters: at least one pad must have an average charge/sampling above 25 units;
- in 3-pad clusters: if one or both of the outer pads have an average charge/sampling less than 25 units, they are removed and the cluster is checked for qualification as 2-pad and 1-pad cluster, respectively.

We show in Fig. 18 the frequency of 'good' 1-pad, 2-pad and 3-pad clusters for the six TPC sectors. Sectors 1, 3, 4 and 6 resemble each other closely while in sectors 2 and 5 clusters are missing relative to the other sectors. We recall that sectors 2 and 5 stick out as the sectors with the largest numbers of 'bad' pads [8].

Furthermore, we show in Fig. 19 the left-right asymmetry of 'good' 2-pad clusters, defined as $(Q_{\mathrm{R}} - Q_{\mathrm{L}})/(Q_{\mathrm{R}} + Q_{\mathrm{L}})$, where Q refers to the time-integrated charge of the left (L) or right (R) pad of a 2-pad cluster. The data are in general satisfactory, except sectors 2 and 5 again which are to be kept in mind for investigations of possible biases in physics analyses.

Analogously, we show in Fig. 20 the left-right asymmetry of 'good' 3-pad clusters (for a three-pad cluster, half of the middle charge is added to the left charge, and the other half to the right charge). The plots look satisfactory.

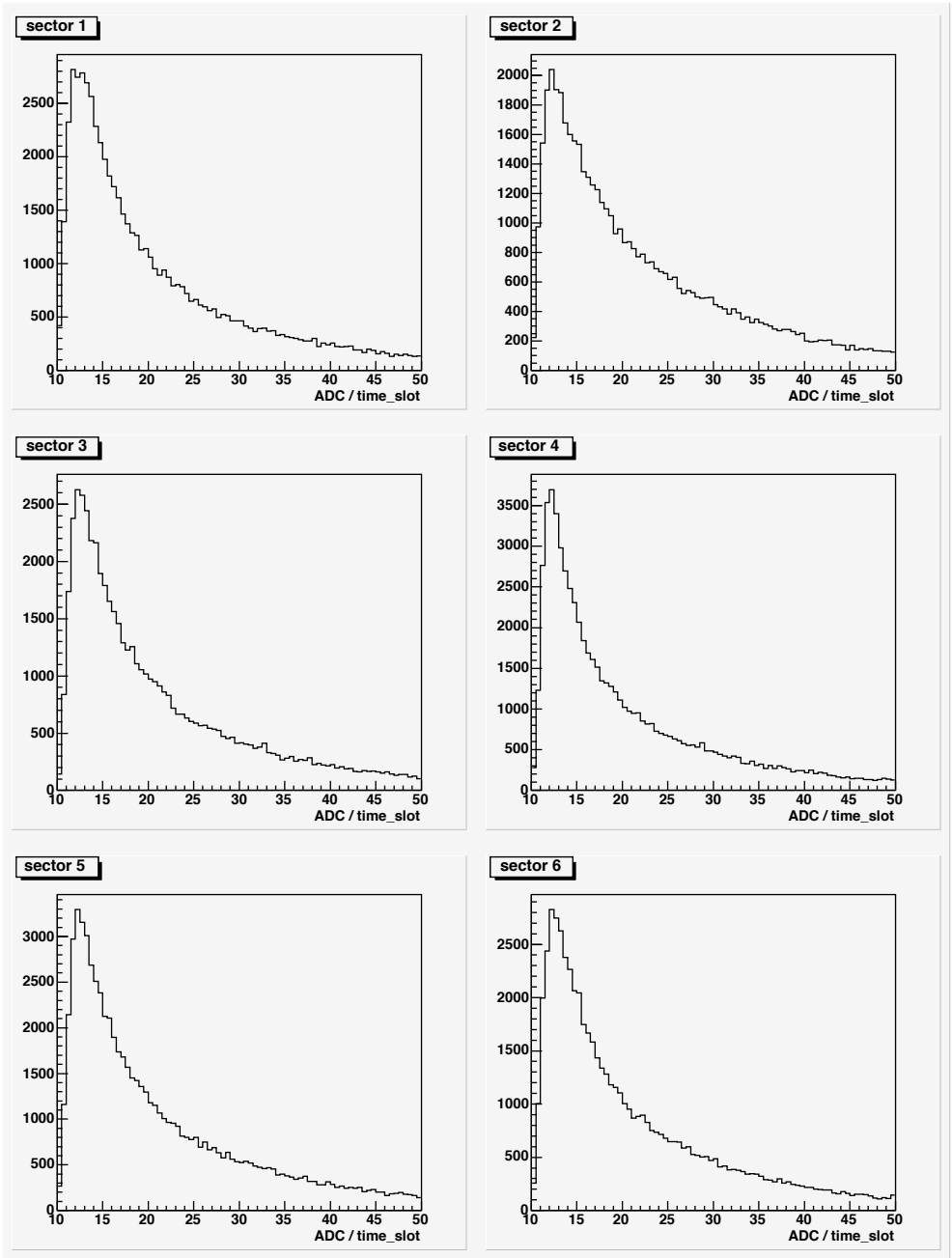

Figure 16: Average charge/sampling of 1-pad clusters, after crosstalk correction and pad equalization yet before the 'good' cluster selection, separately for the six TPC sectors.

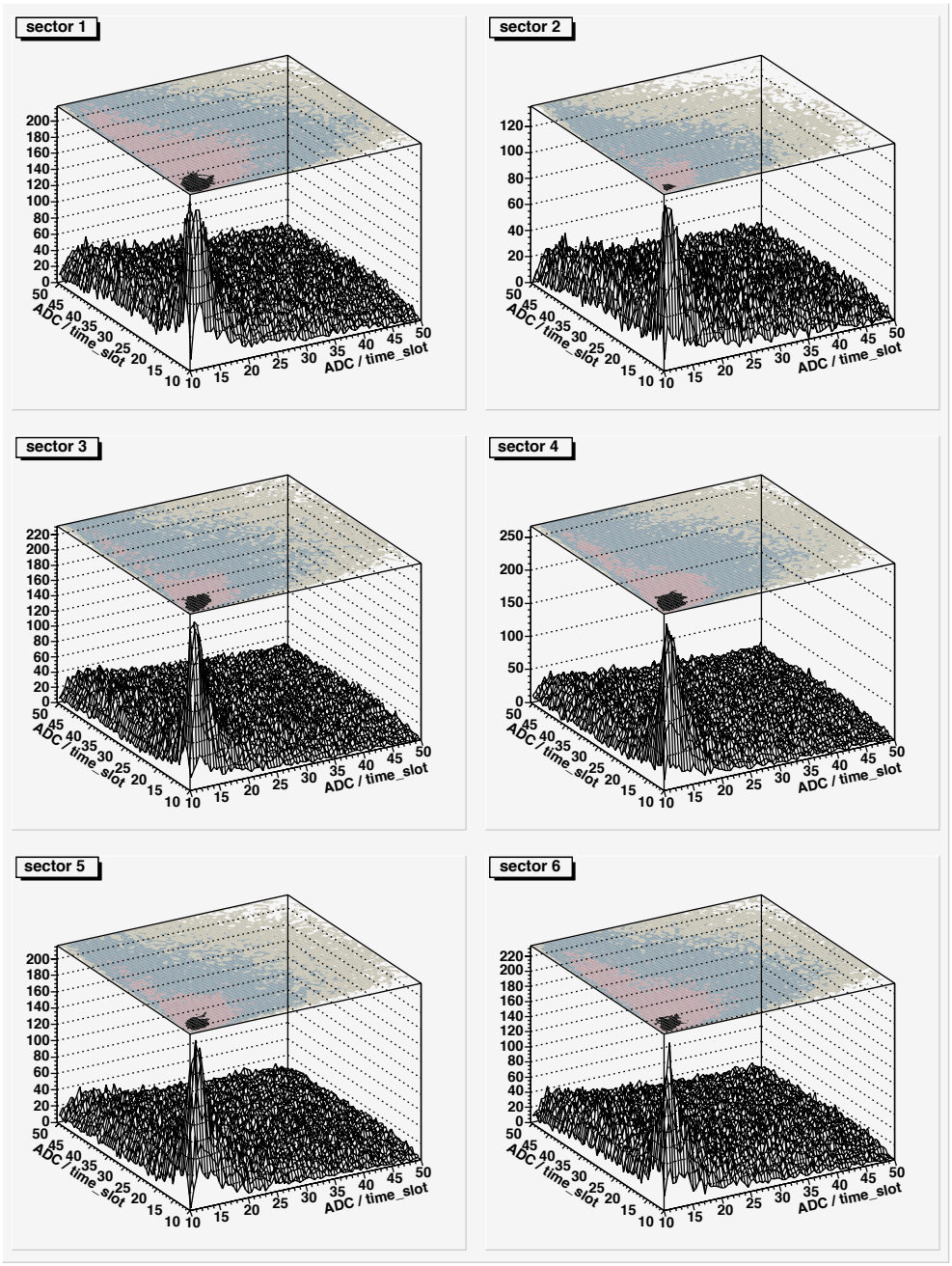

Figure 17: Scatterplot of the average charge/sampling of one pad versus the charge of the neighour pad, for 2-pad clusters and 3-pad clusters, after crosstalk correction and pad equalization yet before the 'good' cluster selection, separately for the six TPC sectors.

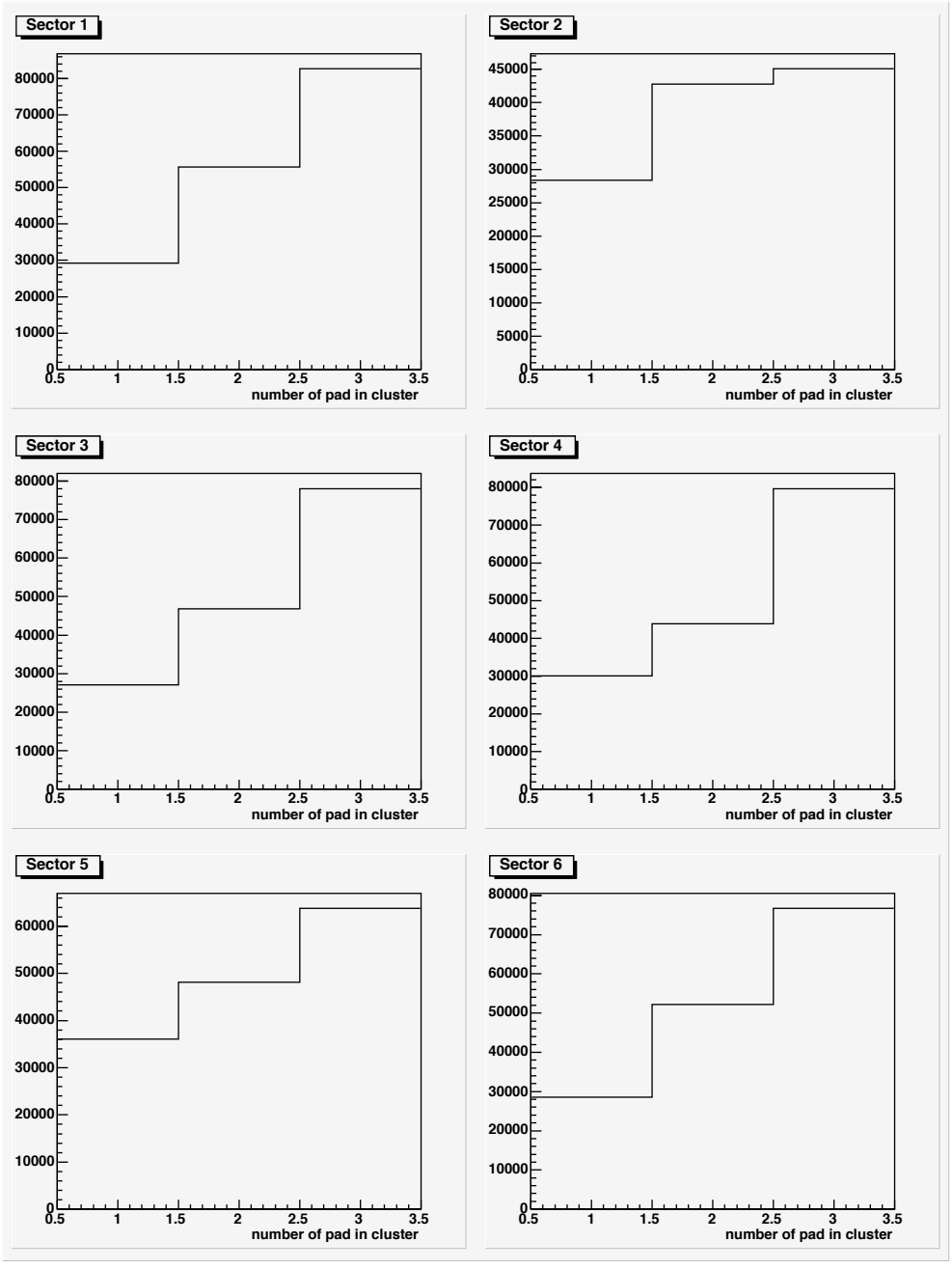

Figure 18: Frequency of 'good' 1-pad, 2-pad and 3-pad clusters in physics data, separately for the six TPC sectors.

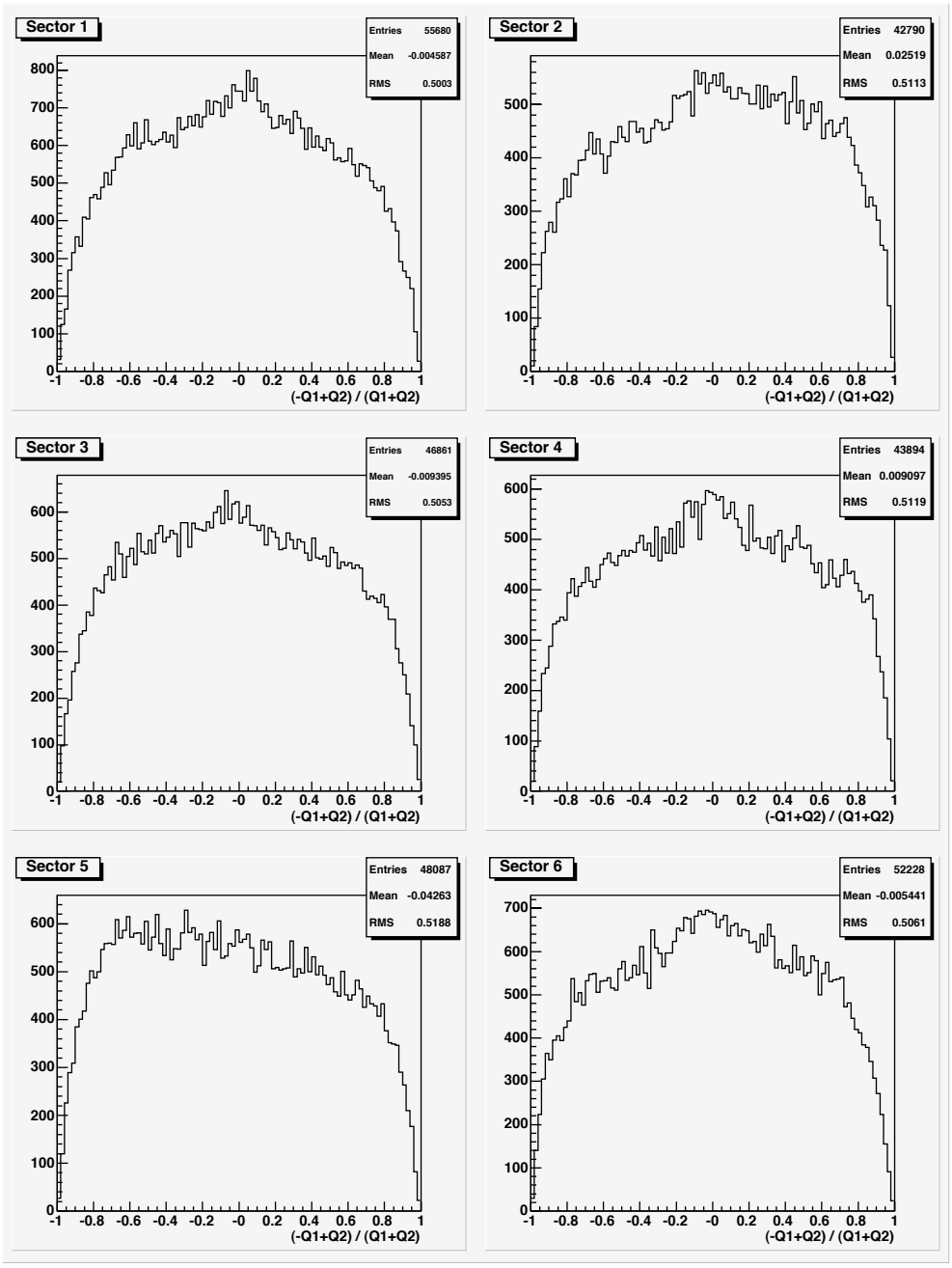

Figure 19: Left-right asymmetry of 'good' 2-pad clusters, separately for the six TPC sectors.

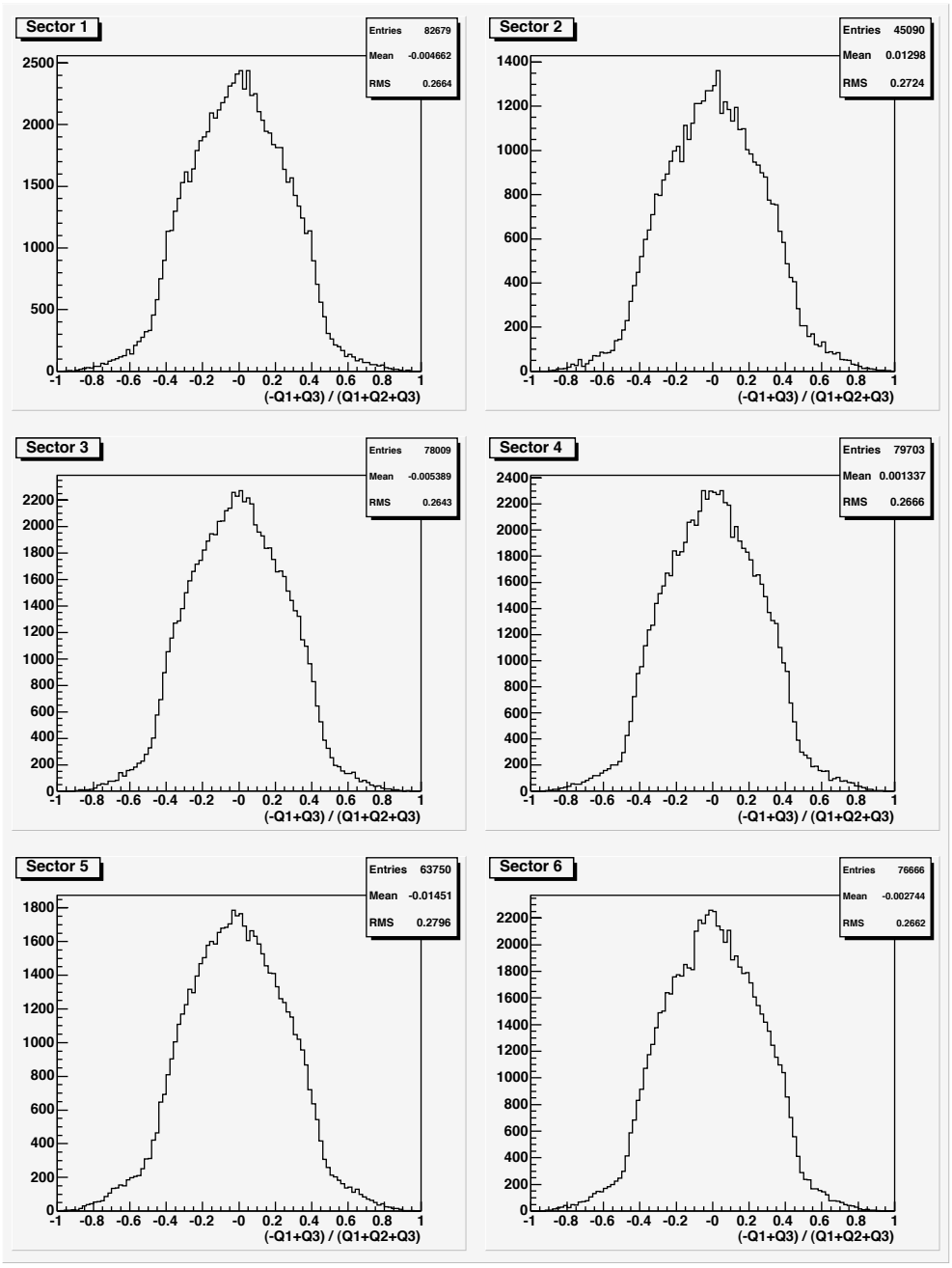

Figure 20: Left-right asymmetry of 'good' 3-pad clusters, separately for the six TPC sectors.

5 Handling of bad pads and of pads next to sector boundaries

Here, we present our handling of bad pads for the determination of the $r \cdot \phi$ coordinate of a cluster (for the water data, the list of bad pads is given in Ref. [8]).

Pads next to sector boundaries are handled as if the (physically not existing) adjacent pad were a bad pad.

Bad pads are taken into account only if they are adjacent to a live pad in the same pad row (further bad pads adjacent to a bad pad are ignored).

While we retain clusters with bad pads in order to aid pattern recognition, we shall assign a large error to the $r \cdot \phi$ position of such clusters, with a view to avoiding a bias in track fitting as much as possible.

In this Section, we mean by the 'charge' of a pad again its time-integrated charge.

Consider a 2-pad cluster, with the pad right or left of the live pad 'bad'. With a view to making pads at sector boundaries as useful as possible, we adopted the following algorithm: we assign to the bad pad the difference between the most probable total charge of 'good' 2-pad clusters (see the distribution of the total charges in Fig. 21) and the charge observed in the live pad, and calculate accordingly the $r \cdot \phi$ position; if the difference is zero or negative, we ignore the bad pad and consider the cluster as 1-pad cluster.

Consider a 3-pad cluster, with both pads right and left of the live pad 'bad'. In this case, we take the centre of the the live pad as $r \cdot \phi$ cluster position.

Consider a 3-pad cluster, with one pad next to the live pad with the larger charge. In this case, the likelihood is large that charge is missing which contributes significantly to the $r \cdot \phi$ position. We assign to the bad pad the difference between the most probable total charge of 'good' 3-pad clusters (see the distribution of the total charges in Fig. 21) and the sum of the charges observed in the two live pads, and calculate accordingly the $r \cdot \phi$ position; if the difference is zero or neative, we ignore the bad pad and consider the cluster as 2-pad cluster.

Consider a 3-pad cluster, with one bad pad next to the live pad with the smaller charge. If the smaller charge is less than 30% of the larger charge, the likelihood is large that little or no charge is missing in the bad pad. Therefore, we ignore the bad pad and consider the cluster as 2-pad cluster. If the smaller charge is equal or larger than 30% of the larger charge, we assign to the bad pad the difference between the most probable total charge of 'good' 3-pad clusters and the sum of the charges observed in the two live pads, and calculate accordingly the $r \cdot \phi$ position; if the difference is zero or neative, we ignore the bad pad and consider the cluster as 2-pad cluster.

By construction, 3-pad clusters with a bad pad in the middle do not exist.

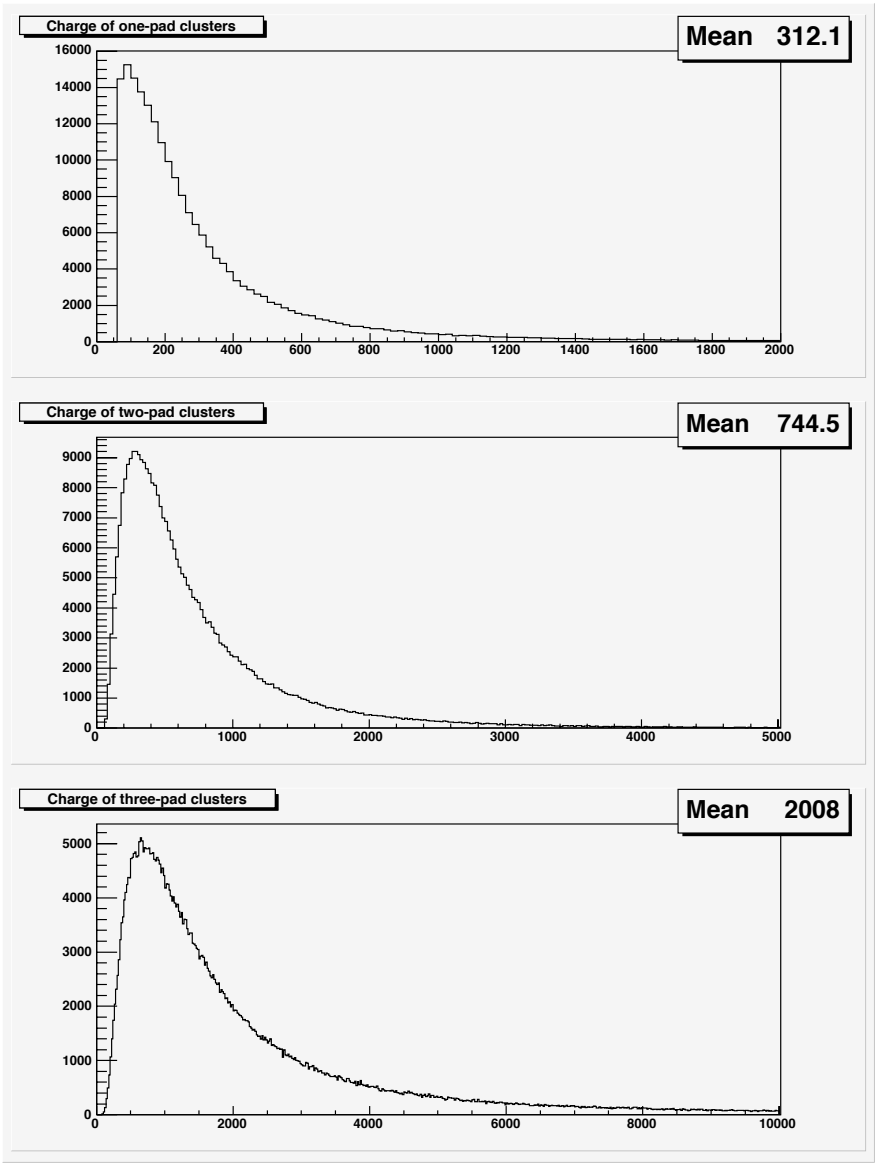

Figure 21: Total charge of 'good' 1-pad (top), 2-pad (middle) and 3-pad (bottom) clusters.

References

[1] F. Dydak, On distortions of TPC coordinates: inhomogeneities of electric and magnetic field, HARP Memo 03-001 (6 June 2003), *http://cern.ch/dydak/TPCdistortions.ps*

[2] F. Dydak, A. Krasnoperov, Yu. Nefedov, TPC track distortions: correction maps for magnetic and static electric inhomogeneities, HARP Memo 03-002 (30 June 2003), *http://cern.ch/dydak/TPCdistortions2.ps*

[3] TPC track distortions III: *fiat lux*, HARP Memo (in preparation).

[4] TPC track distortions IV: *post tenebras lux*, HARP Memo (in preparation).

[5] A. De Min *et al.*, TPC cross-talk correction: CERN–Dubna–Milano algorithm and results, HARP Memo 03-003 (6 July 2003), *http://cern.ch/dydak/crosstalk3.ps*

[6] F. Dydak *et al.*, Performance of TPC crosstalk correction, HARP Memo 04-101 (7 May 2004), *http://cern.ch/dydak/crosstalk4.ps*

[7] A. De Min and F. Dydak, Analysis of HARP TPC krypton data, HARP Memo 04-103 (23 April 2004), *http://cern.ch/dydak/kryptonCalib.ps*

[8] F. Dydak *et al.*, Water data: bad pads, 3.6 μs and 100 ns problems, HARP Memo 04-104 (27 May 2004), *http://cern.ch/dydak/badpads.ps*.

[9] S. Borghi, M.G. Catanesi and R. Veenhof, Clustering Algorithm, HARP Internal Note (March 2003).

[10] The ALEPH Handbook, Report ALEPH 89-77 (1989), editor W. Blum.

[11] J.C. Legrand, private communication.

[12] We thank G. Prior for help with the relation between pad numbers and preamplifier input numbers.

[13] M. Chizhov *et al.*, Implementation of the TPC crosstalk correction, HARP Internal Note (11 September 2002), *http://cern.ch/dydak/crosstalk.ps*.