ALEPH 92-52
SOFTWR 92-06
J. Carr et al.
1.4.1992

Simulation of VDET

J. Carr, P. Coyle, J. Drinkard, D. Rousseau

Centre de Physique des Particules de Marseille

1 Introduction

The Monte Carlo (MC) program for simulating the Aleph Minivertex Detector (VDET) has been under semi-continuous development since 1986 [1]. Last year the program was debugged and implemented for simulation of VDET using the 1990 or 1991 detector configurations. In this note we describe the organization of the software and algorithms used in the generation and digitization of the signals, and, for 1991 conditions, describe the code developed this year to simulate inefficiencies. All results described in this note refer to the 1991 geometry and simulation, as 1991 encompasses the period of useful VDET data.

2 Organization Within Galeph

The VDET simulation is integrated in the standard fashion within Galeph, as illustrated by the calling tree in figure 1a. Initialization includes reading the data-base (VRDDAF,VRDGAL) and building the geometry (AGVDET). Hit generation begins in the routine VDHIT, where energy released from charged tracks into the silicon (as calculated by Geant) is summed at each tracking increment. VDHIT also saves the entry and exit points of a track's intersection with a VDET face. When the tracking is finished, the routine VDASIG copies these numbers to the named bank VDHT, which is written to the POT. Digitization of the signals is driven by the routine VDDIGI, which is described below.

The organization VDDIGI is sketched in figure 1b, the objective being to transform the deposited charge into a cluster (i.e. a set of strip addresses with strip signals given in terms of ADC counts), which approximates that which

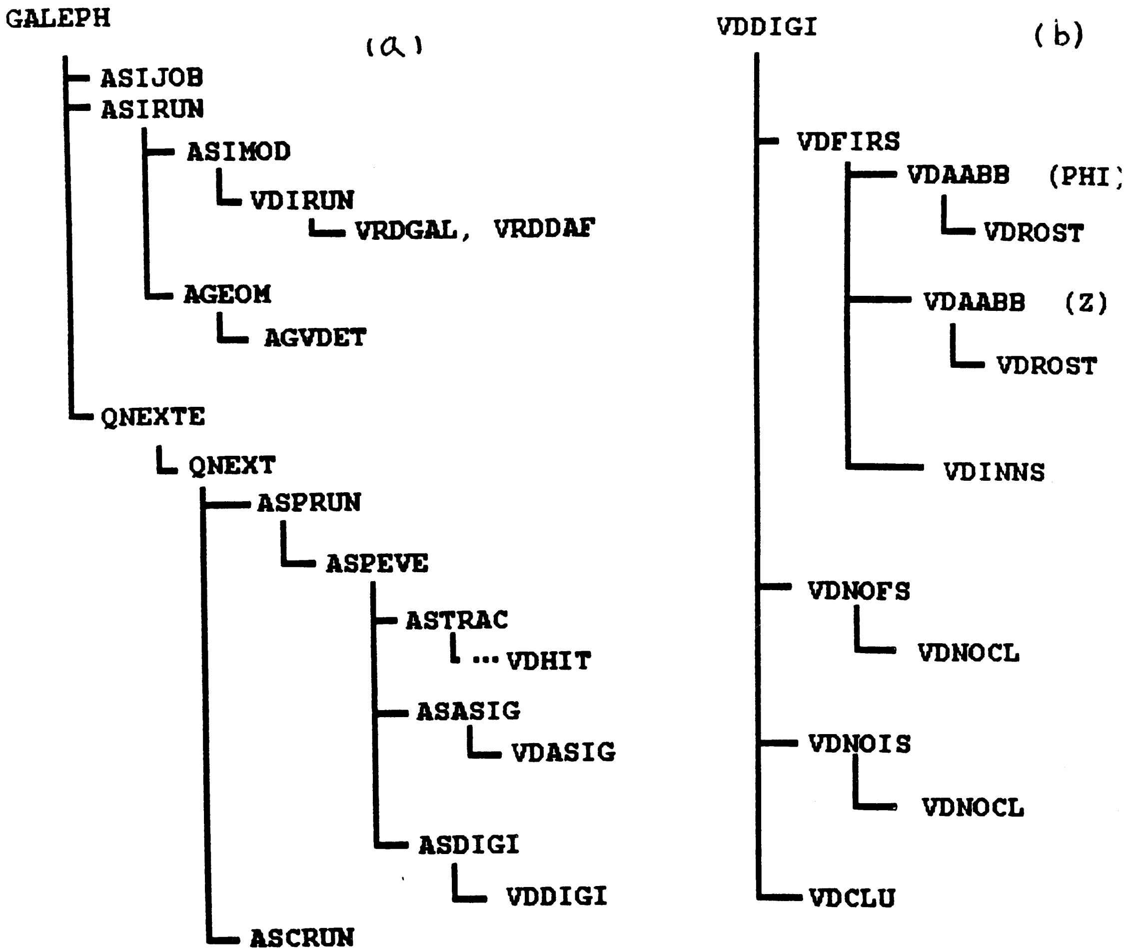

Figure 1: Calling trees for (a) VDET simulation routines within GALEPH, and (b) the digitzation control routine VDDIGI.

is seen in real data. First, several work banks are booked to keep track of the history of each fired strip, then the routine VDFIRS is called to do the actual digitization for each hit. The output of VDFIRS is an array of fired strips for each dimension, ϕ (U) and Z (W), for each VDET hit. The routine VDINSS inserts this output to the track-to-digit history work banks. After digitizing real hits in VDFIRS, VDDIGI calls the routine VDNOFS, which performs an electronics noise simulation (described in the next section). Optionally, the user may require spurious hits, i.e., hits not associated to tracks. The routine VDNOIS serves this purpose, though it is not used by default in Galeph, since it has been shown that such spurious hits do not occur in significant amounts in the actual detector.

Clustering of the fired strips takes place in VDCLU, which stores the results in work banks and returns control to VDDIGI. At this point, all track to digit history information, as well as the clustering results, are stored in work banks. The last function of VDDIGI is to create from these the raw pulse-height banks, VPLH; the hit address bank VHLS; and the final track-to-digit history bank, VDTD.

3 Simulation of a VDET Hit

The task of digitizing a VDET hit and simulating the readout is performed in various subroutines: VDAABB, which simulates the charge diffusion and transport; VDROST, which projects the charge seen by each strip onto the readout strips; VDNOCL, which adds correlated noise to strips including and surrounding each fired strip; and by VDCLU, which groups fired strips into clusters using a simple algorithm. Non-zero pedestals and common mode noise, which are conditions present (and corrected for) in real data, are not simulated in the MC.

The following subsections describe in more detail each of these routines.

3.1 VDAABB

The routine VDAABB subdivides the energy from VDHT into 40 separate "clouds", whose initial positions lie along the trajectory of the track, which is taken as the line joining the entry and exit points through the wafer. The charge distribution for each segment diffuses as it drifts toward the strips, with a profile which is Gaussian and a width which is proportonal to $\sqrt{t}$, where t is calculated by integrating the relation between drift velocity and electric field. The latter is calculated exactly assuming a simple electrostatic model. The effect of the ALEPH magnetic field on the drift of holes is modeled by introducing a Lorentz angle, θ_L to modify the direction of the drift[2]. Strips in the path of $\pm 3\sigma$ window for a given cloud receive a normalized fraction of that cloud's charge. The process is illustrated in Figure 2a. Note that this procedure involves all VDET strips, i.e. not just read-out strips.

The Lorentz angle results in an average shift of about 7μ in the final r-ϕ position; since this shift is effectively removed in the real data by the software alignment, we perform a "special" alignment (an alignment only in r-ϕ) on the MC which removes the shift. This special alignment file is included in the data base for versions ≥ 152.

3.2 VDROST

This routine simulates the effect of having just a subset of the strips read out electronically (recall that in ϕ, every fourth strip is a read out, while in z, every second strip is read out). VDROST loops over all fired strips, building an array of readout strips. When a readout strip is encountered, it is given a signal which consists of a fraction of its original signal (as assigned in VDAABB), plus contributions from seven neighbouring strips on each side. The coupling constants which determine the relative contribution of each strip are calculated in reference 1. Figure 2b shows the cluster of figure 2a, projected onto readout strips using this technique.

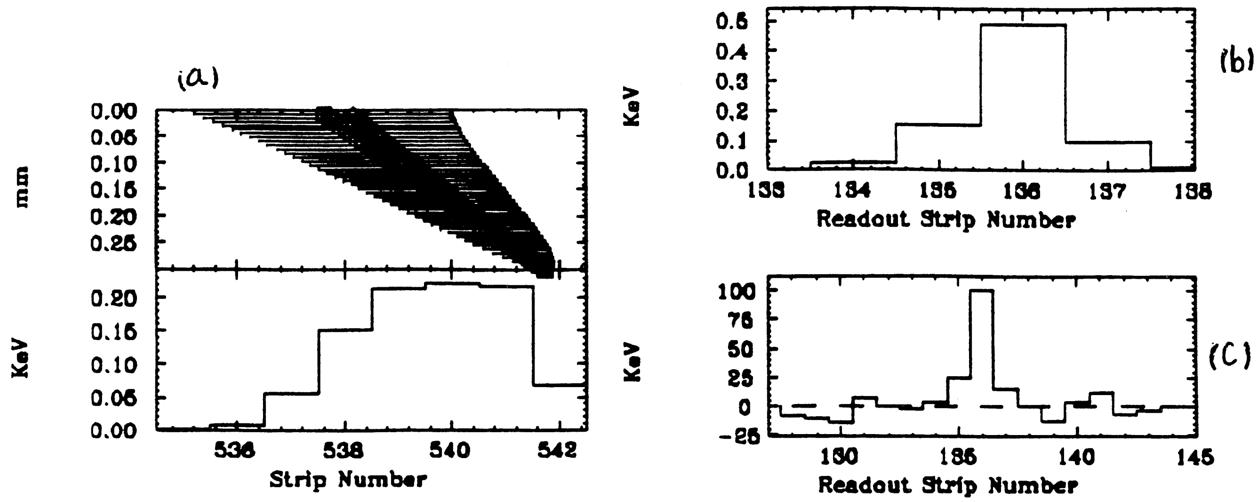

Figure 2: Evolution of a cluster generated by a single muon, as seen from the phi strips. (a): trajectory of muon is shown by the boxes; horizontal brackets represent the active area seen by strips due to diffusion at the 3σ level; the diamonds represent the center of the charge centroid after Lorentz drift. The histogram is the sum of collected charge on each strip. (b): The same cluster projected onto readout strips, and (c), after the addition of noise.

3.3 VDNOFS

For simulation purposes, the noise in the VDET consists of two contributions; (a) a parallel current source which represents all the noise contributions from the detector, the bias network and all noise from the electronics not directly related to the amplification mechanism (b) a series voltage source at the input of the electronics which includes noise due to the amplification.

VDNOFS is the control routine for adding these noise contributions to the clusters. For each fired strip it calls the VDNOCL routine, in which the fired strip and each of the four neighbouring strips either side of the fired strip, are assigned a series and parallel noise charge randomly choosen from gaussian distributions. The gaussians have mean zero and sigmas determined from studies of the noise observed in the test beam data[3]. These noise charges are then added to the charge of the fired strip. Correlation between the series noise of the fired strip and the series noise of the closest four neighbouring strips is also taken into account using a correlated noise formula[4]. The coupling constants used for the correlated noise come from analytic calculations (see reference 1). Figure 2c shows the cluster of figure 2b after the addition of noise.

3.4 VDCLU

The routine VDCLU takes the strip array and looks for contiguous groups of strips above some threshold. The threshold in GALEPH for 1991 simulation is st to 23 ADC counts; commensurate with the typical thresholds used in the Sciroccos. There is no summed pulse height cut, either in the MC or in the Sciroccos (though the facility exists in the latter). Once the clusters are determined, the signals are converted to charge and scaled by a gain factor so that the signals have units of ADC counts.

4 Performance

In this section, agreement between the MC and data is checked for various quantites, explained below:

- Position resolution. This is calculated by constraining a track to hit two layers of the detector and measuring its residual with respect to a third (i.e. there can be up to four layers when one considers overlaps), and is compared for the data and MC[5] for the W dimension in figure 3. The true resolution of the MC, determined using truth banks VDHT and VDTD, is also included for reference in this plot, and shows that the three-layer method reproduces reasonably well the actual resolution in the MC. The plots show that the measured resolution in W for the MC is systematically lower by $20 - 40\%$, depending on track incident angle. In U (not plotted), one notices a similar trend. The reason for this discrepancy is not fully understood; two likely explanations are (1) the absence of misalignments and geometrical distortions in the MC, and (2) disagreements in average signal-to-noise.

- Pulse height. The summed pulse height (i.e. the sum of strip signals for strips used in the cluster calculation) is compared for the data and MC in figure 4, along with the corresponding plots for cluster dimension. Slight disagreements are apparent in each set of plots, presumably because the MC still needs some fine tuning. These discrepancies are not expected to affect physics significantly.

- Signal-to-noise. Plotted for data (fill 779) and MC in Figure 5 is the single-strip RMS noise estimator,

$$\frac{PH_U - PH_W}{\sqrt{N_U + N_W}},$$

  where PH is the pulse-height for the U or W dimensions and N is the corresponding strip number. This estimator, though slightly biased here because of pulse height cuts applied in the offline clustering, still gives a good idea of the RMS strip noise[6]. The agreement is satisfactory, showing sigmas of 10.3 and 10.4 ADC counts for the data and MC, respectively. Translated into effective signal to noise, we obtain an approximate value

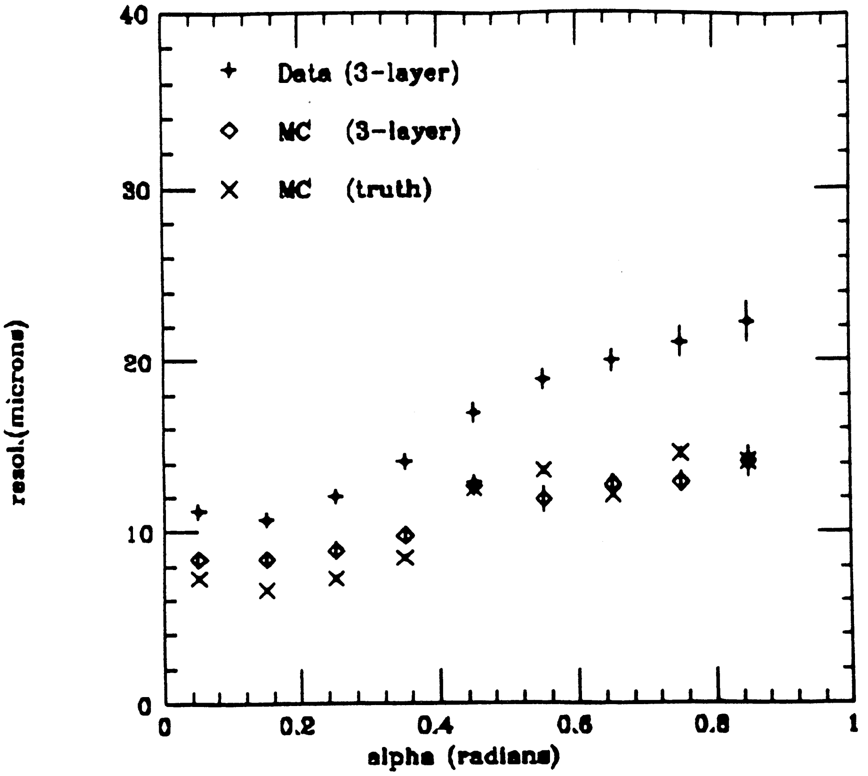

Figure 3: A comparison in W position resolution between data and MC, as a function of track incident angle, alpha.

of 14, though, using an un-biased calculation, one can show that the actual value is closer to 19[7].

It should be noted that while the RMS noise in the MC and data agree within to 10% averaged over a large fraction of the data, we have noticed discrepancies of up to 30% for certain time periods (e.g. the time period corresponding to fills 565-566).

- Linearity in charge division. Since VDET is a silicon detector with capacitively coupled channels, the position resolution should be a strong function of "floating" strip pitch (25μ in U, 50μ in W), as opposed to the readout strip pitch (100μ in both dimensions). As a check on this, we plot in figure 6 the quantity

$$\eta = \frac{PH_N}{PH_N + PH_{N+1}},$$

where PH_N and PH_{N+1} are the two highest pulse heights in the cluster, subscripted by their strip numbers. η gives a reasonable estimate cluster position modulo readout strip, assuming most of the charge is confined to two strips. For normally incident tracks, one would thus expect enhancements in η for W near 0.0,0.5 and 1.0, and for U near 0.0,0.25,0.5, 0.75 and 1.0. The plots in figure 6 show this expected behaviour for the data, though it is evident that diffusion has the effect of washing out the expected peaks in U. Furthermore, there exist fairly significant discrepancies between the data and MC in U layer 2 and W, layer 1. This disagreement, which has negligible physics consequences, has been traced to a bug in the MC[8].

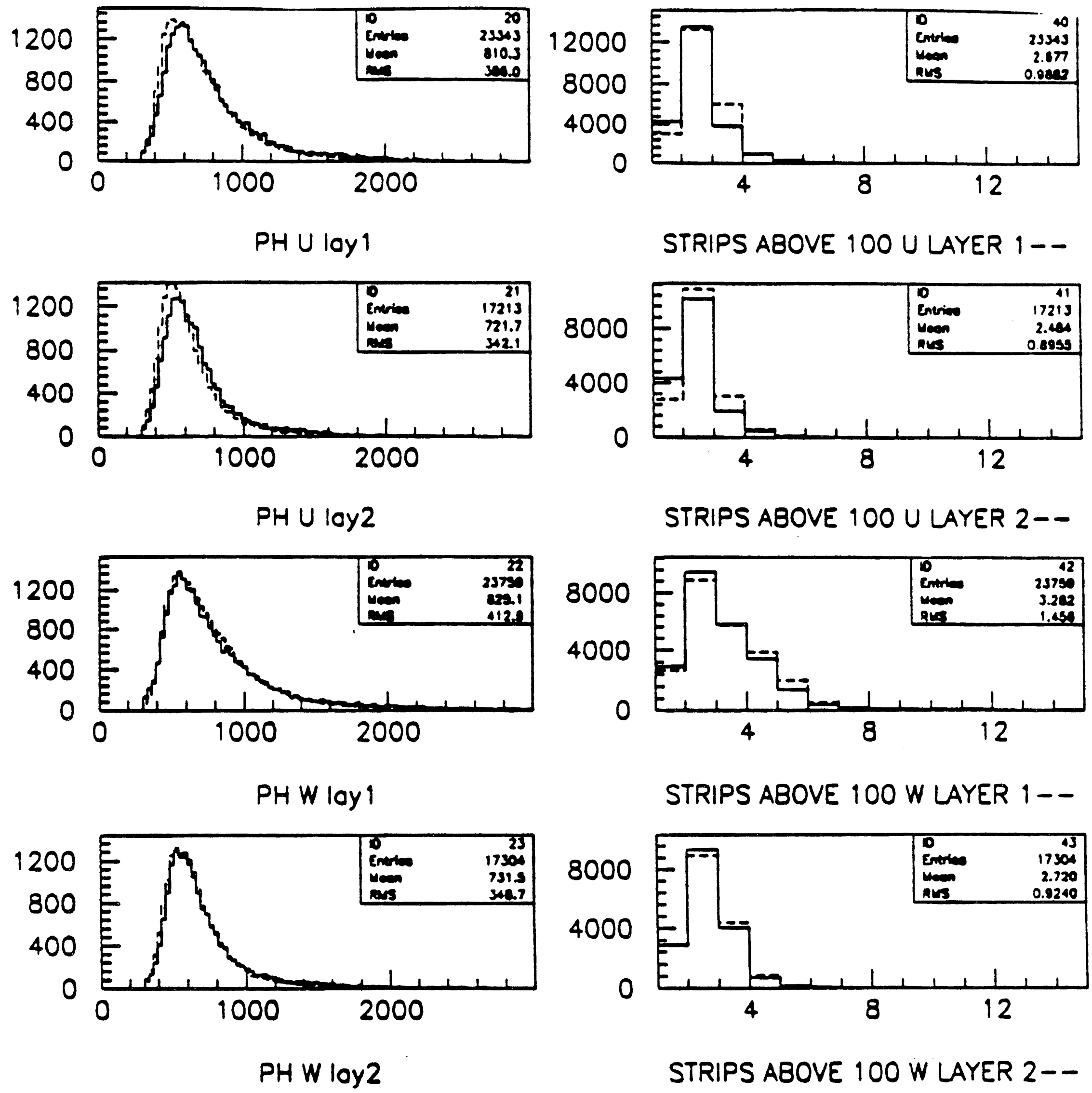

Figure 4: First column shows a comparison in summed pulse heights for data (solid) and MC (dashed) for each dimension in each layer. The second column shows the corresponding cluster dimension for strips with signals exceeding the minimum pulse height of 25 ADC counts (the figure is 100 counts in "offline" units).

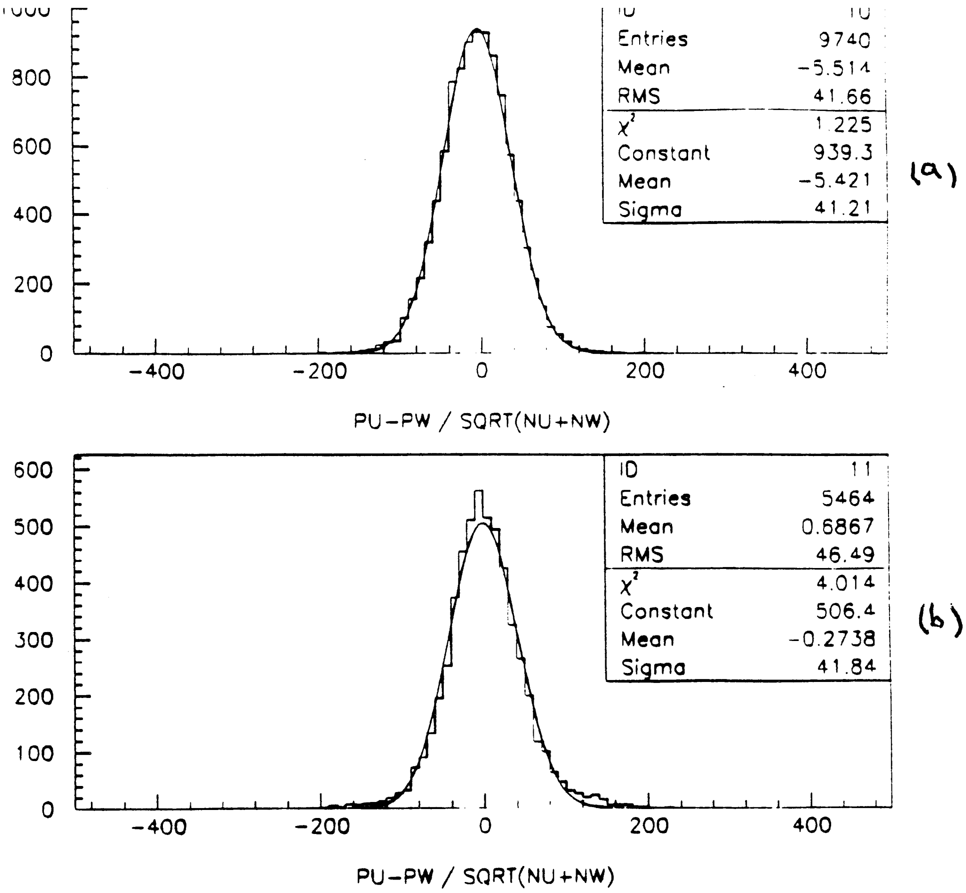

Figure 5: Distribution in single-strip RMS noise for the MC (a) and for the data (b). The units are in ADC counts × 4.

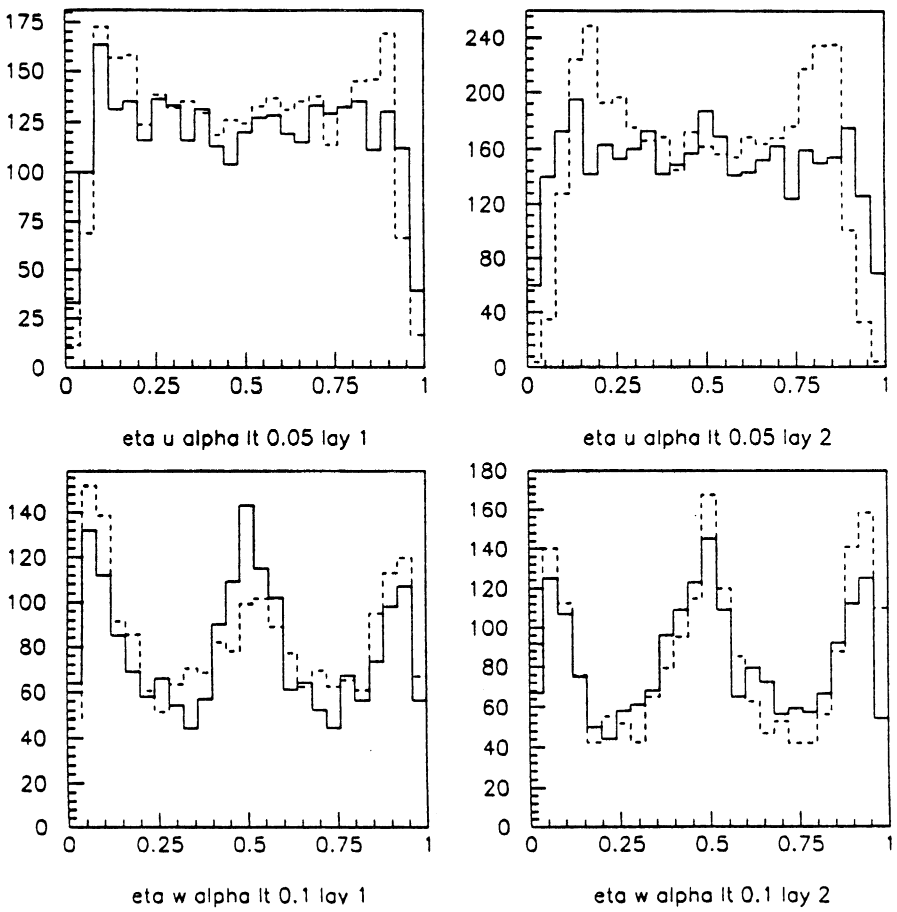

Figure 6: Distribution in η for the data (solid histogram) and MC (dashed).

5 Simulation of Inefficiencies in 1991

5.1 Philosophy

For runs in which the VDET was included in the partition, the major source of inefficiency was assumed to be due to line-drivers (hereafter referred to as "lines") which, for one reason or another, were dead for all or part of the run. Other sources of inefficiency, such as dead or noisy channels, or cutoffs from poorly adjusted thresholds, were considered to be less serious in comparison[9].

The present MC simulates only the effect of dead lines. All other sources of inefficiency are assumed to be negligible, or naturally modeled in the normal reconstruction chain. Compared to a precise calculation of efficiency, this approach has the advantage of not being sensitive to systematics in tracking, and it avoids the complication of the breaking down the calculated efficiency into its various sources. The latter task would be necessary to avoid double-counting the contributions inherently present in the reconstruction (i.e. from clustering). Finally, time dependence, already a serious problem in the simple flagging of dead lines, would further complicate a more sophisticated approach of efficiency determination.

5.2 Dead Line Simulation

Simulation of dead lines is performed at the JULIA level; i.e., nothing is done in Galeph. JULIA first reconstructs hits in the normal fashion, with the creation of the position banks VDXY and VDZT. After hit reconstruction, the routine VDMCEF (for JULIA versions ≥ 257) is called, which loops over hits, flagging those (bit 30=1, quality word in VDXY/ZT) which occur in the region of the detector corresponding to a dead line for a given time period. These flagged hits are then ignored in the pattern recognition and track refit. Every 100 events, VDMCEF makes a full rotation through all time periods.

The efficiency maps are constructed in the following way: All 1991 data (to run 12817[10]) corresponding to SCANBOOK-selected VDET runs (with XVDEOK filtering of VDET hits) are analyzed in a run-by-run search for dead lines. A line is considered dead if it has a measured efficiency, $\epsilon_{\rm ld}$, of below 30% for a given run, where $\epsilon_{\rm ld}$ is calculated as follows:

$$\epsilon_{\rm ld} = N_c/N_e,$$

where N_e is the number of "good" FRFT tracks extrapolated to within 4 σ of their errors to the line, and N_c is the number of VDXY/ZT hits actually found on that line, within some distance to the extrapolated hit. In the U dimension, this "proximity window" is ±1.2mm; in the W dimension it is ±5.5mm. Moreover, the requirement $N_e \geq 10$ is imposed to give the calculation at least marginal statistical significance. The track quality cuts, for reference, are:

- momentum of at least 1 Gev/c.
- $|d_0| < 0.2$ cm
- $|z_0| < 10$ cm
- at least 4 ITC coordinates
- at least 6 TPC coordinates
- a track-fit χ^2 per d.o.f. no greater than 4.

A plot which correlates dead lines in time is shown in Figure 7a, where N_e is the number of extrapolated tracks, incremented by run, up to run 12817. Using this plot, we divide the 1991 "VDET" running period into ten blocks, assigning a different efficiency map (where the map contains 1's and 0's), for each time block. The MC model of Figure 7a is shown in Figure 7b. Averaged efficiencies for data and MC are shown in figures 8a and 8b, respectively.

The efficiency maps are stored in the data-base (versions ≥ 160) in the bank VDEM. Also in this bank are numbers for each map describing the percentage of the useful VDET data that each map corresponds to, and two numbers giving an approximate idea of the run-interval each map was calculated from. This last piece of information should be treated cautiously, however, since a time block does not necessarily imply consecutive runs. The total length of the bank is currently $1 + (288 + 3) \times N$ words, where the first word gives the number of time blocks, N (which as mentioned above, is for now set to ten).

Comparing more closely the base-line efficiencies in figure 8, one notices a systematic difference between data and MC at the level of 4%. If one computes the efficiency after removing the dead lines, one obtains for the MC approximate values of 98% for Z lines and 94% for ϕ lines; the corresponding values for real data are 94% and 90%, respectively. For reference, if one uses truth information to calculate the "true" efficiency in MC, one obtains a figure of about 97% for both ϕ and Z. The origin of the disagreement in the MC between "truth" and the "blind" efficiency calculation we have described is most likely a result of tracking and association errors. The disagreement between data and MC is not fully understood, though some of the discrepancy may be explained by the absence of dead individual channels in the MC (i.e. see reference 9).

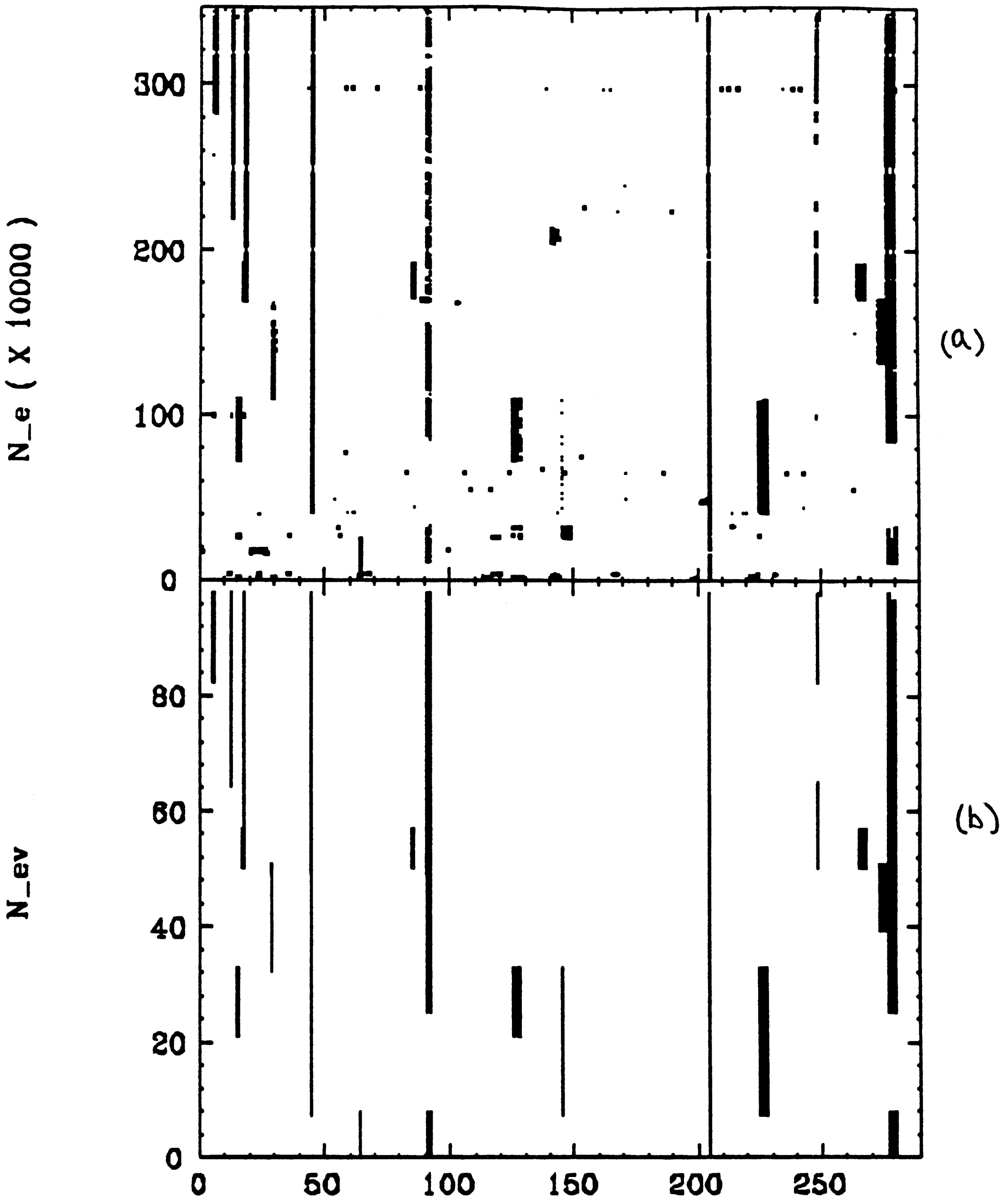

Figure 7: Correlation of dead line drivers with time for data (a) and the MC model (b). Time is represented in units of 10^5 track-extrapolated hits.

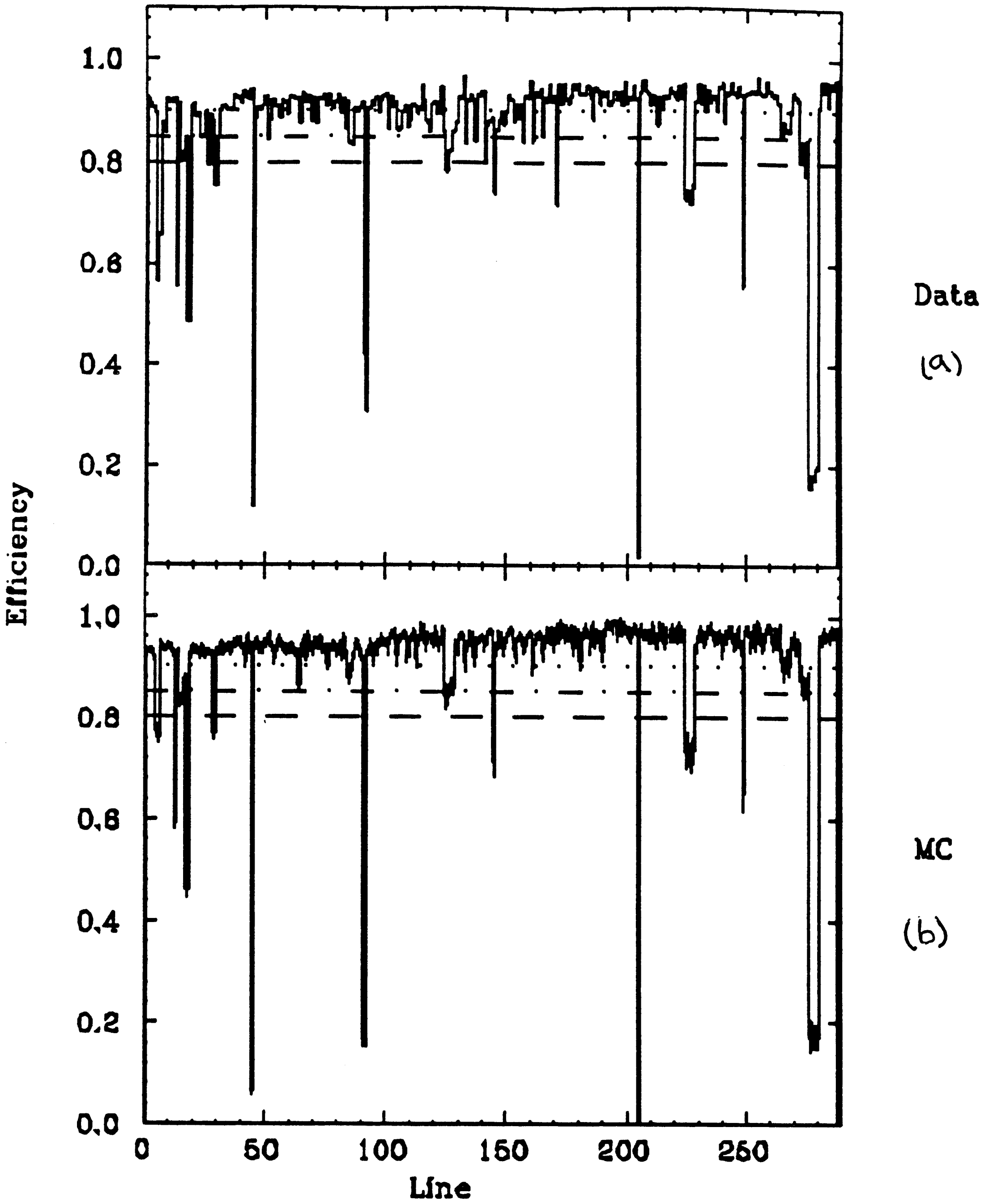

Figure 8: Averaged efficiency including dead lines for data (a) and for the MC (b). For reference, the solid dashed level represents 80%, the dot-dashed line 85% and the dotted line 90% efficiency.

6 Conclusion

In this note, we have described the algorithms and status of the 1991 VDET Monte Carlo. In a comparison with the 1991 data we can make the following conclusions:

- Reasonable agreement with the data is found in the MC signal-to-noise simulation.
- The position resolution is, for quasi-normal tracks, about 25% better in the MC ($\approx 9\mu$) than for the data ($\approx 12\mu$).
- In the absence of dead lines, the efficiency calculated using a crude method averages about 4% higher in the MC. The obvious way to "force" agreement would be to flag dead an additional percentage of the clusters in the MC; indeed, the structure to do this is already present in VDMCEF. This approach, if taken, would necessitate a more sophisticated approach to efficiency determination than the method we have selected.

To fine tune the 1991 MC into better agreement with the data would first require more detailed and systematic studies concerning each of the above points, followed by adjustments of the relevant MC data-base parameters. No major additions to the software are anticipated. For improvements concerning efficiency directly, current Galeph could then be reproccessd. For adjustments to signal-to-noise or gain, however, Galeph digitization for VDET would first need to be repeated.

We are currently awaiting feedback from physics analysis groups to help decide the extent of future improvements to the MC.

REFERENCES

1. Paolo Walter Cattaneo, Thesis, MPI 1987(unpublished).
2. More details of this treament of diffusion and Lorentz drift may be found in: E.Belau et.al., NIM 214A (1983) 253.
3. G. Lutz MPI-PAE/Exp. El. 173 (Feb 1987).
4. "Noise and Signal Processing Through the Readout Electronics of ALEPH Minivertex Detector" Paolo Walter Cattaneo (unpublished)
5. The three-layer residual measurements were supplied by George Redlinger.
6. Signal-to-Noise in VDET, Aleph note in preparation, J.Boudreau and F. Weber.
7. ibid
8. This bug has been fixed for Galeph versions ≥ 251. We say that physics impact is negligible because the degradation to the MC position resolution arising from it amounted to approximately 1μ.
9. As it turns out, the number of dead channels at the end of 1991 was 6.7%. Of these, various dead lines and one dead module accounted for 70%, with dead, saturated (i.e. from pinholes) and noisy channels accounting for the remaining 30%. These numbers were presented by Alan Litke in the talk "VDET REPAIRS", at the ALEPH Plenary Meeting at CERN, 13-14 February, 1992.
10. The entire data set was not available when these efficiency maps were created. A set of updated efficiency maps, correponding to the entire run (i.e. all reprocessed runs through 13468) will be released in version 165 of the database.