E-787 TN-374
November 2, 1999

Experiment 787
Technical Note No. 374

An algorithm to fit pulses in target CCD data

Milind Diwan, Jim Frank, Vivek Jain

Abstract

This technote describes an algorithm for fitting Kaon fiber pulses in the target CCD data. The code, which was originally written to run on PASS3 ntuples, has been adapted to run as part of KOFIA. Both high gain and low gain CCD pulses are fit, and the results are stored in the PASS3 ntuples. A routine, which runs on these ntuples, has been written to choose between high gain and low gain results, depending on the Kaon energy in a given fiber, E_K(fiber) , and $(T_\pi - T_K)$.

The performance of the pulse fitter is studied using a "Monte Carlo". We have also run the fitter on 1996 Skim5 data. Some preliminary results are presented.

1 Introduction

The main background in the PNN2 region is believed to be due to $K \rightarrow \pi^+ \pi^0$, where the π^+ scatters in the target, loses energy and falls in the kinematic box [1]. One signal for this background can be the presence of a second pulse in a Kaon fiber. It is, therefore, imperative to fit pulses in Kaon fibers and veto events whenever a second pulse, above a given threshold, is found.

We have chosen to use the calibrated pulse shapes (for each fiber) to fit the data rather than the analytical approach used by Joe and Toshio. Results are stored in the PASS3 ntuples and are subsequently used in the PNN2 analysis. We then choose between high gain and low gain results, depending on the Kaon energy in the given fiber, E_K(fiber) , and $(T_\pi - T_K)$.

The performance of this fitter has been studied using a "Monte Carlo". We take pulses in Kaon fibers which have a very good single pulse fits (probability $\geq$ 80%), and add a second pulse, with amplitude varying between 0 and 5 MeV, at different values of $(T_\pi - T_K)$. This is done in bins of E_K(fiber) .

2 Fitting algorithm

We first describe the fitting algorithm. The main routines, shpfit2-hg and shpfit2-lg, are called from setup-pass2. Each pulse is fitted with an average pulse shape determined separately for each fiber. Pulse shapes and the error (in each time bin) associated with the shapes are stored in CFM, separately for high gain and low gain.

The pulse shapes were determined as follows. Firstly, all other target CCD calibrations were done (using KM21 monitors). From a set of ntuples, pulses for both high gain and low gain were chosen to be included in the calibration if (a) E_K(fiber) was greater than 4 MeV, (b) pulse had not overflowed, and (c) the time was within 3 ns of 0. The events also had to have passed a target reconstruction, using SWATHCCD. A spline routine was invoked to line up these pulses in time, with a bin width of 0.5 ns. The pulses were also normalized to unit area. Histograms were saved of these normalized and time-aligned pulses. Then a separate routine read in the histograms, and stepping through each fiber, and each time bin, the average and sigma of the pulse height were obtained. The shapes were massaged to "fix" data bins where the spline fitting occasionally gave results inconsistent with adjacent bins. It also improved the errors in the fast rising part of the shapes. We use the same shapes for unsaturated as well as saturated CCD pulses.

The fitting is done using MINUIT. We first perform a single pulse fit, and if the probability of the single pulse fit is low, $\leq 25\%$, then we do a double pulse fit. The shape for the two pulse fit is determined by adding the pulse shape to itself, displaced by a time which corresponds to $(T_\pi - T_K)$. In the current version of the algorithm, the time of the second pulse is constrained to be within ± 10 ns of $(T_\pi - T_K)$.

The single pulse fit uses two parameters, Kaon amplitude and time, whereas the double pulse fit uses four parameters, Kaon and Pion amplitudes and times. Since the unpacked CCD data (in calibrated mode) already includes several levels of pedestal subtraction, we do not include a pedestal in the fit. The errors, on the pulse shape, used in the fit are scaled up from the errors stored in CFM. The scaling factor is arbitrary and has been chosen to give a flat probability distribution (or as flat as possible). At present, the scaling factors for the single pulse fits are $\sqrt{(1000 * \mathrm{Amp_K})}$ and $\sqrt{(600 * \mathrm{Amp_K})}$, for high gain and low gain, respectively. $\mathrm{Amp_K}$ is proportional to $\mathrm{Ek_{tgc}}$, with the average value of the constant being about 170 for high gain and 42 for low gain data. For these results, we calibrated the energy of the second pulse by equating the ADC energy to the sum of the Pion and Kaon amplitudes as found by the fitter.

In Fig 1, we show the chi-square (per d.o.f.) and probability distributions for a single pulse fit, for $K_{\pi 2}$ peak events, in Kaon fibers which are far away from both the decay vertex and the Pion swath. In these fibers, we expect very few second pulses. The peaking of the chi-square distribution around 1.0 and the flatness of the probability distribution, for all three samples, indicates that the errors are reasonable.

After the PASS3 executable is run, we perform the next stage of analysis on the ntuples. We choose between high gain and low gain fit information, depending upon availability and on E_K(fiber) and $(T_\pi - T_K)$. For instance, if high gain pulses are unsaturated (corresponds to roughly E_K(fiber) < 15 MeV), we use high gain fit results but if E_K(fiber) > 40 MeV, we use low gain results. If E_K(fiber) is between 15 and 25 MeV, we choose depending on $(T_\pi - T_K)$; for $(T_\pi - T_K)$ less than 12 ns, we use low gain results, otherwise we use high gain results (for E_K(fiber) between 25 and 40 MeV, the threshold is 20 ns). At this stage, we only consider those double fits where double pulse fit probability is greater than 0.1%, the fitted second pulse amplitude is positive, and in the case of low gain data, where many fibers fibers are multiplexed into one CCD channel, we require that only fiber per CCD channel is hit. Averaged over events in the KPI2 and PNN2 kinematic regions, roughly 35% of the fit results are based on low gain information. The result of this routine is fed into the TGCUT package.

3 Fitter performance using the 1996 Skim5 data

The fitter was run on 1996 Skim5 events as part of the pass3 executable. Events were selected by making standard KINCUT, TDCUT1 and tight PSCUT cuts, PNN2 box cut, reversing Pveto cuts in the BV and RS, and making Pveto cuts in the beam elements like TG, VC, CO, CM, EC-Up and EC-Dn Inner ring. These cuts enhance cases where one would expect to see the Pion (from

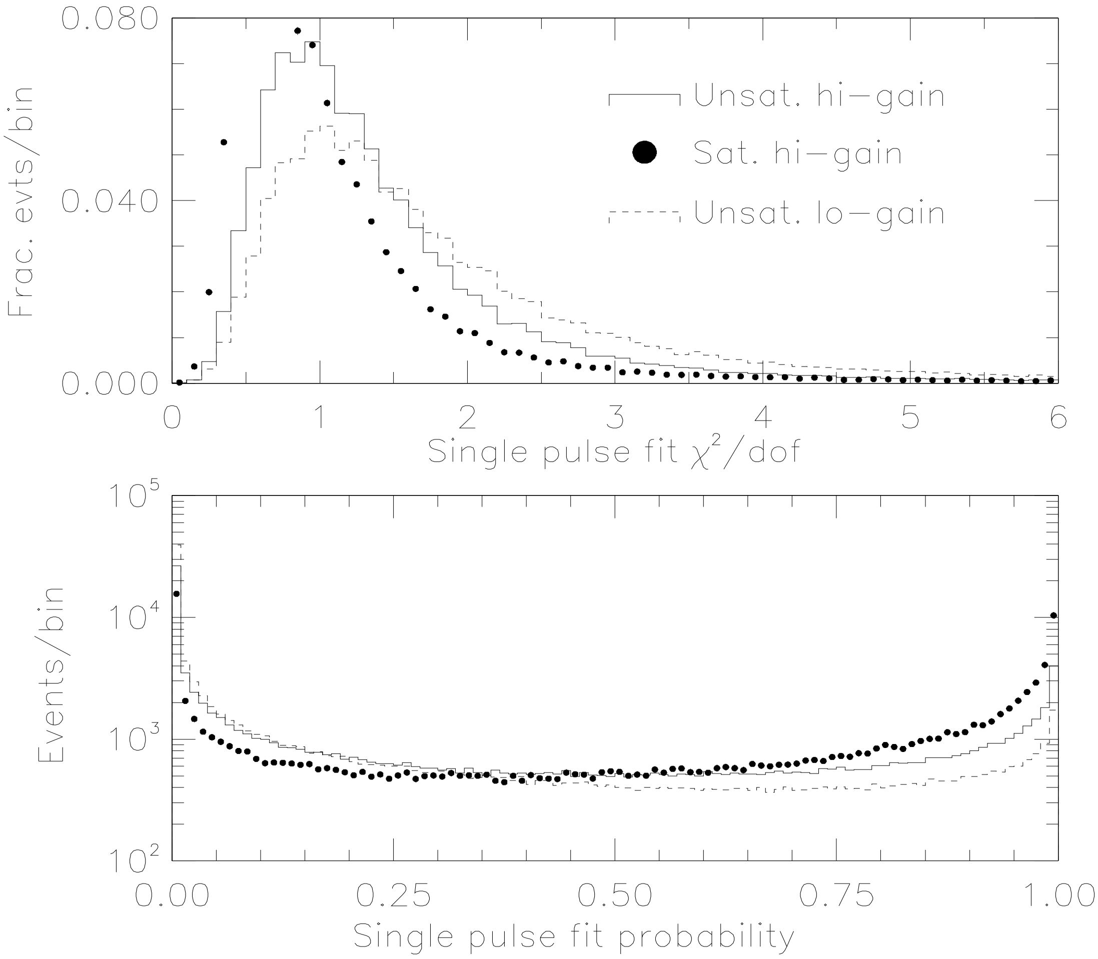

Figure 1: Chi-square and Probability distributions for a single pulse fit

$K_{\pi 2}$ peak events) scatter in the target. We then make TG cuts to remove piscats and further clean up the events. Most of these TG cuts are the standard ones (or newer versions of the standard cuts).

After all these cuts, the rejection of the CCD fitter cut, where we remove events whenever there is a second pulse in any of the Kaon fibers, with energy greater than 1 MeV, is 9.1 ± 2.1, whereas the acceptance of this cut is 55%. It is clear that this cut drastically reduces the background. We do take a 45% acceptance loss, but if we were to optimize R $\times$ A, then the effect of this cut is very beneficial. In fact, with the setup cuts mentioned above, this cutpoint (R=9.1, A=55%) has the best performance.

In Fig 2, we show the energy of the second pulse found in a Kaon fiber for events in the KPI2 and PNN2 kinematic regions. The PNN2 distribution corresponds to the stopping fiber(filled circles), whereas the the KPI2 distributions are for both the stopping fiber (dashed line) and fibers far from the stopping position (solid line). No target cuts, other than the B4EKZ cut, were made to produce this plot. It is clear from here that the fitter is finding higher energy second pulses in the PNN2 sample compared to the KPI2 sample, even in the stopping fiber. This is very promising.

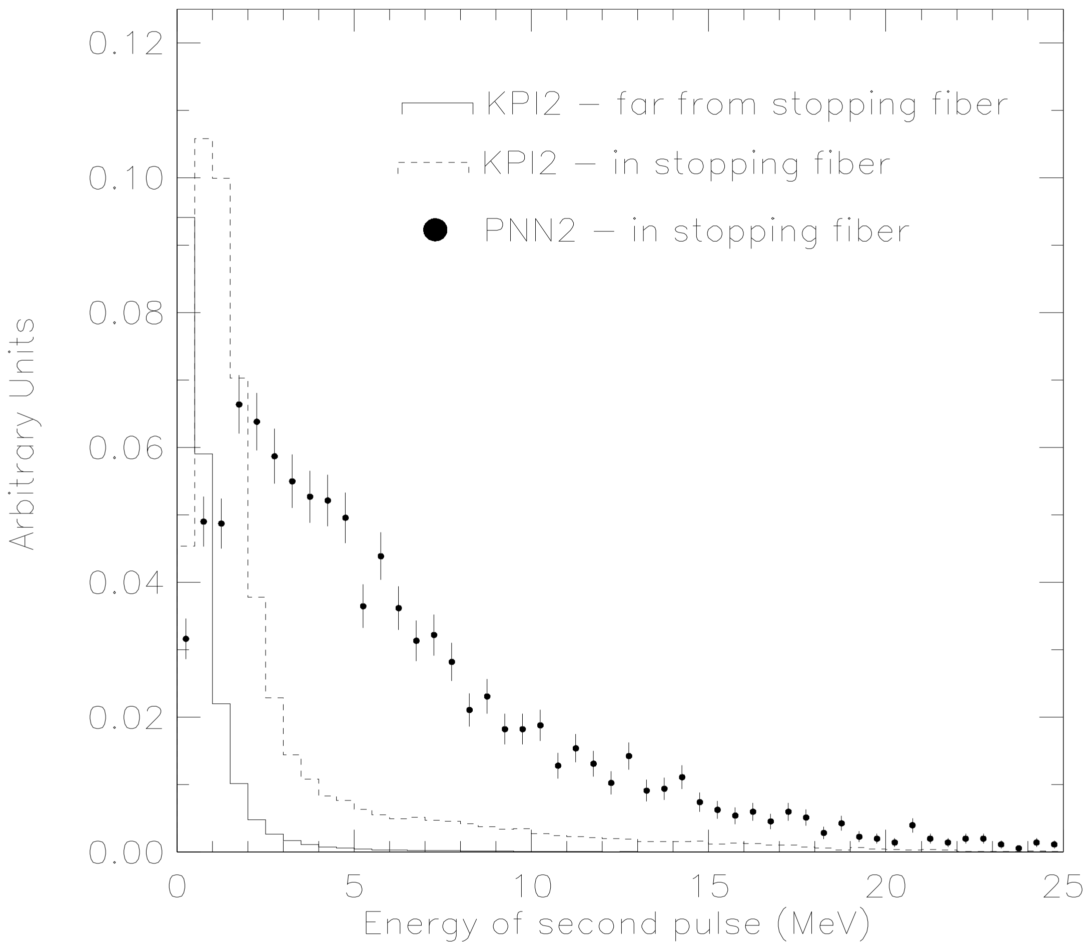

Figure 2: Energy of second pulse in KPI2 and PNN2 event samples. For display purposes, the distributions have been scaled to have the same range along the Y axis.

To understand the systematics involved in the fitting algorithm, we undertook a study which is described next.

4 Fitter Performance

The fitter performance is studied by adding a second pulse of varying amplitude (between 0 and 5 MeV) at different values of $(T_\pi - T_K)$. For this study, the Kaon pulse is required to have a very good single pulse fit probability ($\geq 80\%$). Each chosen Kaon pulse is used 250 times, with 10 bins in second pulse amplitude and 25 bins in $(T_\pi - T_K)$ (where each time bin corresponds to 2ns). The study is done in five bins of E_K(fiber) ; 0-5 MeV, 5-15 MeV, 15-25 MeV, 25-40 MeV and $>$ 40 MeV.

This is done separately for high gain and low gain pulses. In this study, we used about 1000 high gain pulses and 550 low gain pulses.

All Kaon pulses were chosen randomly, the main criteria being that the single pulse fit have a high probability. If there were reflections on the data pulse then there were two possibilities. If the pulse shape accounted for the reflection and the fit was good, then this pulse was accepted in the study. If, however, the fit was bad, for whatever reason, then it was rejected. No explicit effort was made to discard data pulses with reflections.

With this study, we can understand the efficiency of finding the second pulse as a function of E_π, E_K(fiber) and $(T_\pi - T_K)$. In addition, issues like the ratio of amplitude of the fitted second pulse to the input second pulse, difference between the output $(T_\pi - T_K)$ and the input value of $(T_\pi - T_K)$, *etc.* can also be studied. Since the efficiency of finding the second pulse is low at small values of $(T_\pi - T_K)$, one can also make an additional cut on $(T_\pi - T_K)$ and thus improve the peak/tail ratio. This study avoids the pitfalls involved in tuning cuts on small sample of events left after making other cuts.

In Figs 3,4,5(a-e), we show the efficiency of finding a second pulse with energy between 1.0-2.0 MeV, 2.0-3.5 MeV and 3.5-5.0 MeV, respectively, as a function of $(T_\pi - T_K)$ (the errors on the efficiency are smaller than the symbol size and are of the order of 2-5%). The five figures (a-e) are for the five bins in E_K(fiber) . The filled circles are for the high gain data and the open circles are for low gain data. In some of the efficiency plots, the behavior of low gain results seems a bit puzzling (*e.g.*, in Fig 3 for E_K(fiber) bins 0-5 and 5-15 MeV and at times for large values of $(T_\pi - T_K)$). This is an artifact of the procedure of adding a second pulse to a low gain pulse, and not a reflection on the fitter performance.

In Figs 6(a)-(e), we show the ratio of the amplitude of the fitted to the input second pulse as a function of $(T_\pi - T_K)$ for high gain data, and in Figs 7(a)-(e), we show the same distributions for low gain data. It is clear from Figs 6 and 7, that the ratio is very well behaved (in high gain data) for E_K(fiber) less than 15 MeV, and is somewhat worse for E_K(fiber) between 15 and 25 MeV, whereas for low gain data the behavior is reversed, in that the ratio is very well behaved for E_K(fiber) greater than 15 MeV. This is consistent with expectations.

In Figs 8, we show distributions for the difference between the fitted $(T_\pi - T_K)$ and input value of $(T_\pi - T_K)$. One can see that this difference is very tightly distributed around 0. For most values of $(T_\pi - T_K)$ and E_K(fiber) (both high gain and low gain pulses), the fraction of fits where this variable is within ± 5 ns is greater than 98%.

To summarize the above discussion, the fitter performance is optimal if we use high gain data when E_K(fiber) is less than 15 MeV, and low gain data when E_K(fiber) is greater than 15 to 25 MeV. This conclusion is based on studying the efficiency of the fitter, ratio of the amplitude of the fitted and the input value of the second pulse and difference of the fitted and input values of $(T_\pi - T_K)$.

We have also investigated the stability of the fitting procedure to the input time of the second pulse. At present, the starting value in the fit for the time of the second pulse is set to $(T_\pi - T_K)$. We changed the starting value to ($(T_\pi - T_K)$ +4) and found that the fit results were unchanged. We also constrained the output $(T_\pi - T_K)$ to be within ± 2 ns of the input $(T_\pi - T_K)$ (as opposed to the standard 10 ns value), and found no change in the results.

5 Lessons learnt

Since we had run the fitter on PASS3 data before we performed the study described above, we have not fully utilized all the benefits of the fitter. We plan to rerun PASS3 to incorporate some of these conclusions. For instance, while doing the double pulse fit, we plan to add in quadrature, $\sqrt(1000 * \text{Amp}_\pi)$ or $\sqrt(600 * \text{Amp}_\pi)$, for high and low gain data respectively, to the single pulse fit errors. This should increase the second pulse finding efficiency by about 10%.

Another lesson learnt from this study is that since the efficiency of the fitter is low for $(T_\pi - T_K)$ $\leq$ 4ns (for all E_K(fiber) bins except in the 0-5 MeV bin), one can make this cut to gain rejection. Finally, Figs 3, 4 and 5 show that the efficiency of finding a second pulse with E $>$ 1 MeV is well understood, and one can confidently make this cut.

6 Conclusions

We have adapted a fitting algorithm for target CCD pulses to run as part of the pass3 executable, and have studied its performance using both data and a "Monte Carlo". This has proven to be a very valuable tool for the PNN2 analysis.

References

[1] This is conventional wisdom. It has also been investigated using a GEANT based simulation, described in Technote 375.

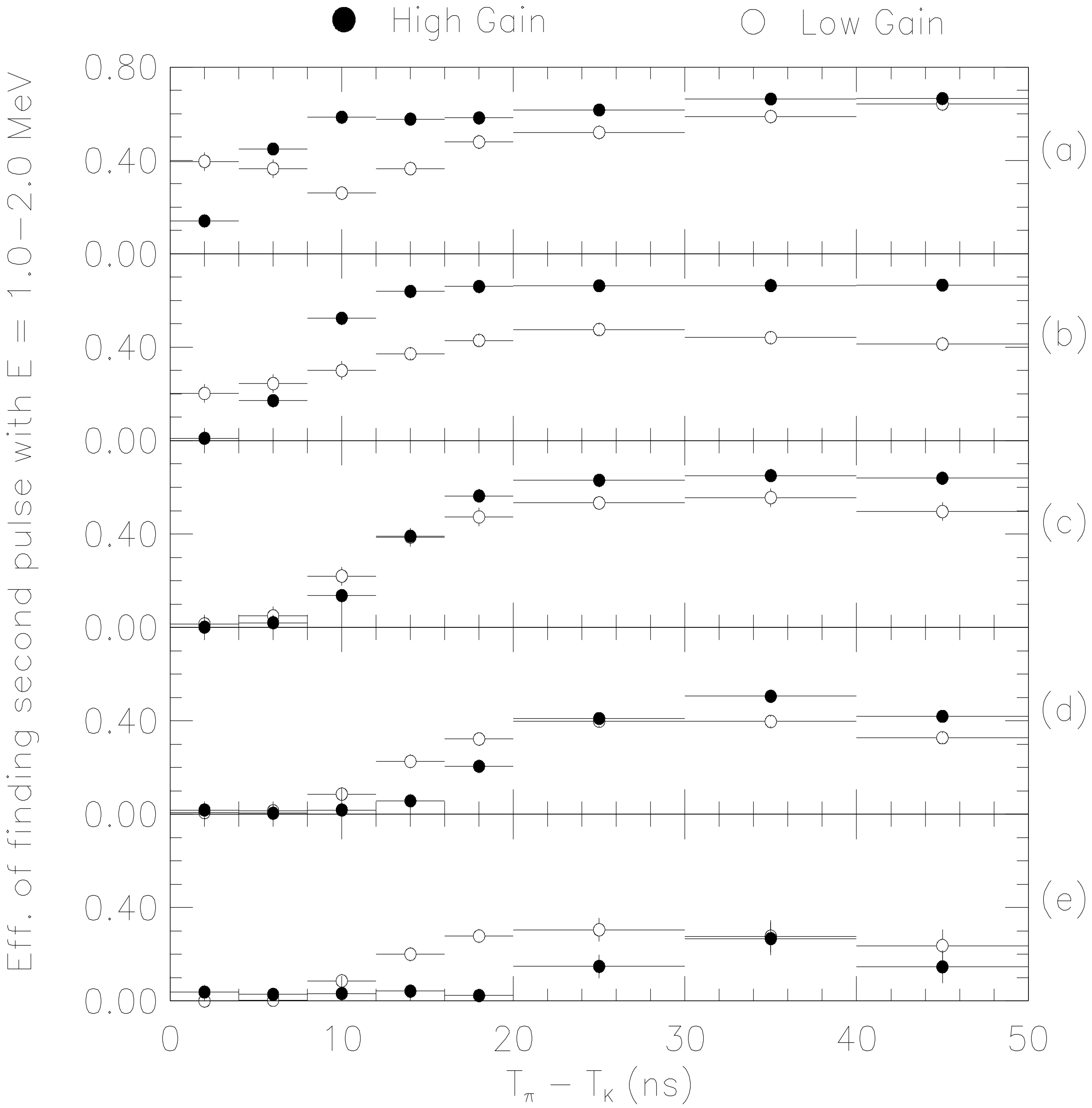

Figure 3: Efficiency of finding a second pulse with energy between 1.0 and 2.0 MeV, as a function of $(T_\pi - T_K)$ in bins of E_K(fiber) , (a) 0-5, (b) 5-15, (c) 15-25, (d) 25-40 and (e) $>$ 40 MeV for high gain (filled circles) and low gain (open circles) data.

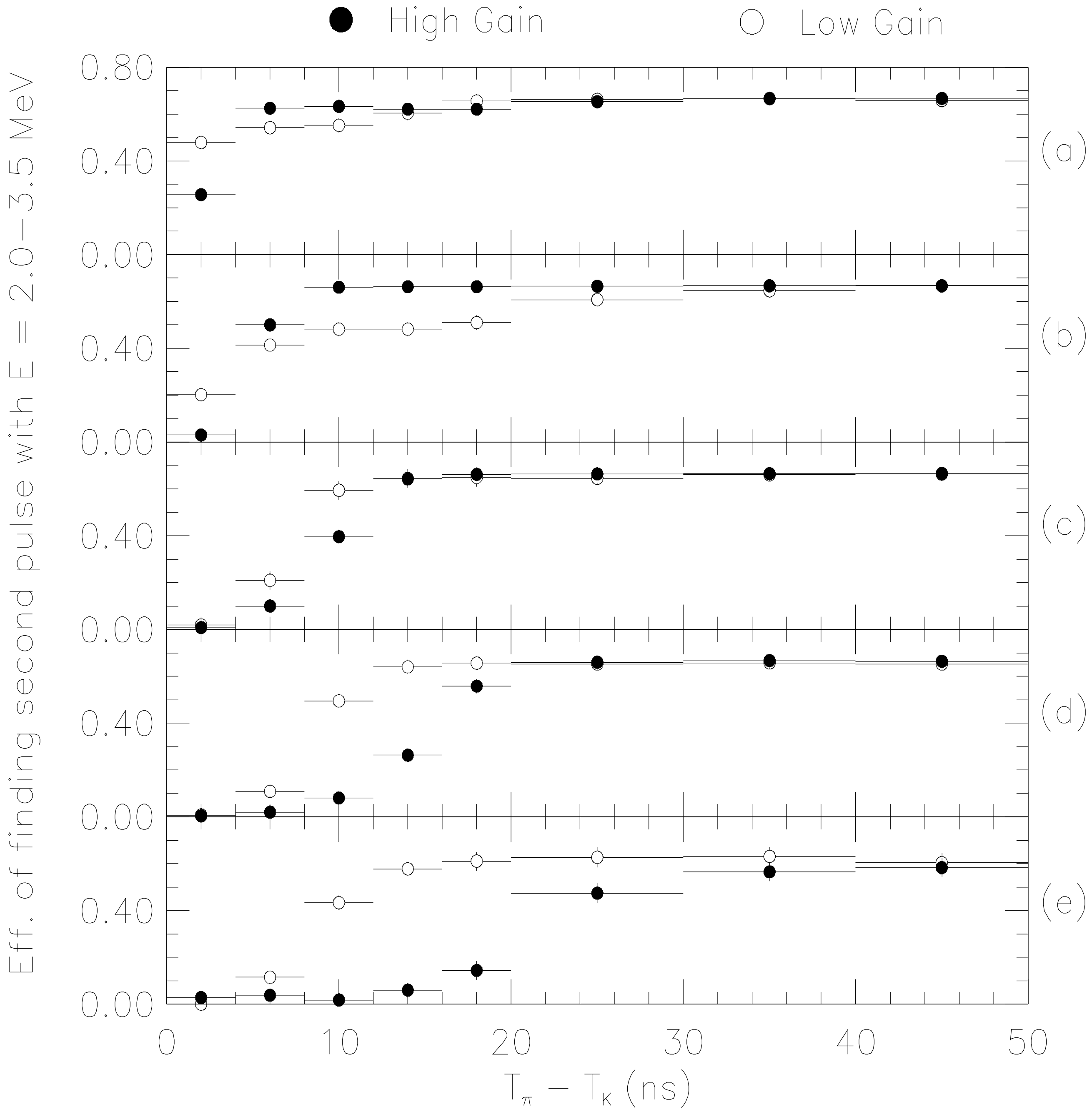

Figure 4: Efficiency of finding a second pulse with energy between 2.0 and 3.5 MeV, as a function of $(T_\pi - T_K)$ in bins of E_K(fiber) , (a) 0-5, (b) 5-15, (c) 15-25, (d) 25-40 and (e) > 40 MeV for high gain (filled circles) and low gain (open circles) data.

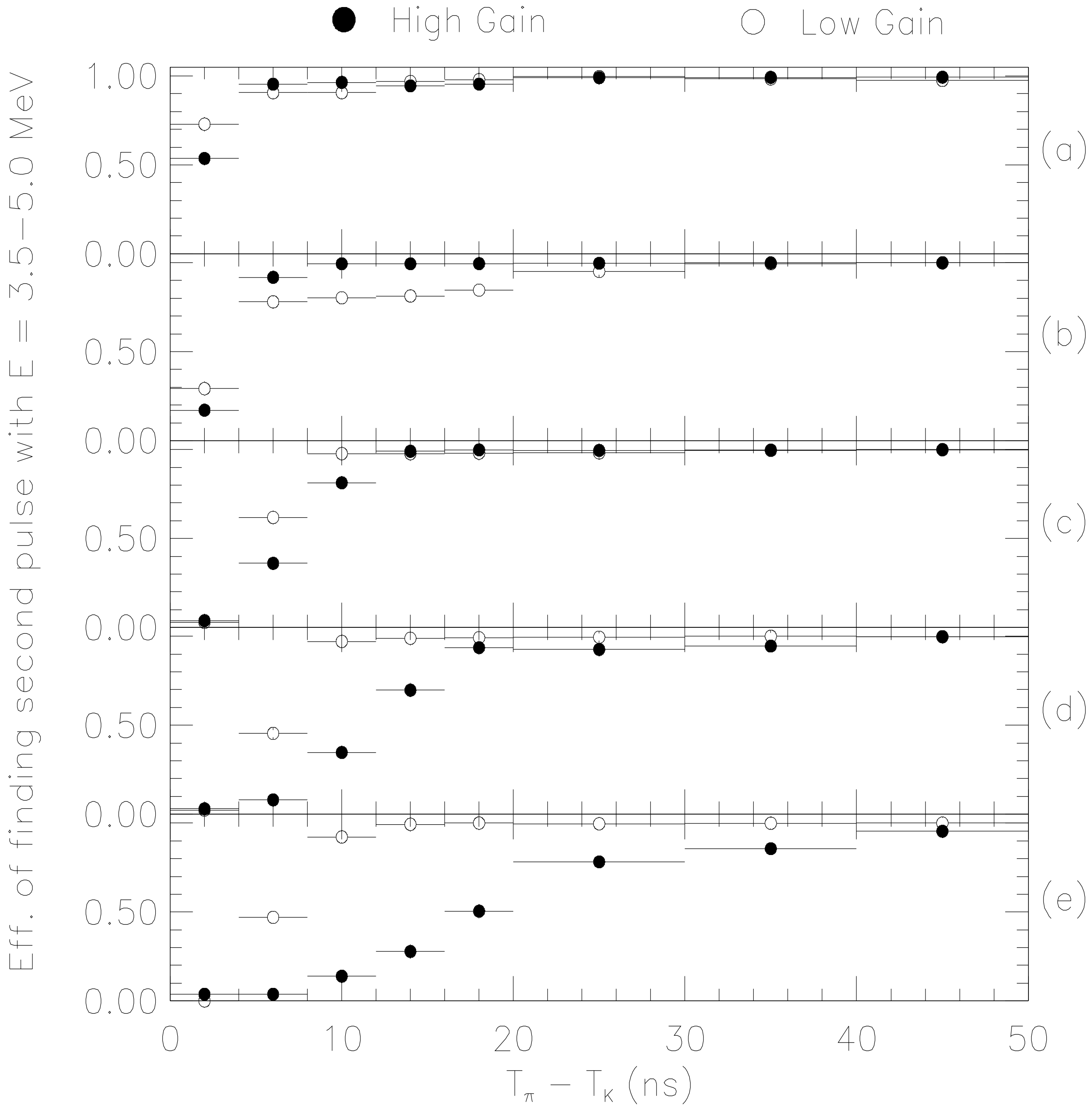

Figure 5: Efficiency of finding a second pulse with energy between 3.5 and 5.0 MeV, as a function of $(T_\pi - T_K)$ in bins of E_K(fiber) , (a) 0-5, (b) 5-15, (c) 15-25, (d) 25-40 and (e) $>$ 40 MeV for high gain (filled circles) and low gain (open circles) data.

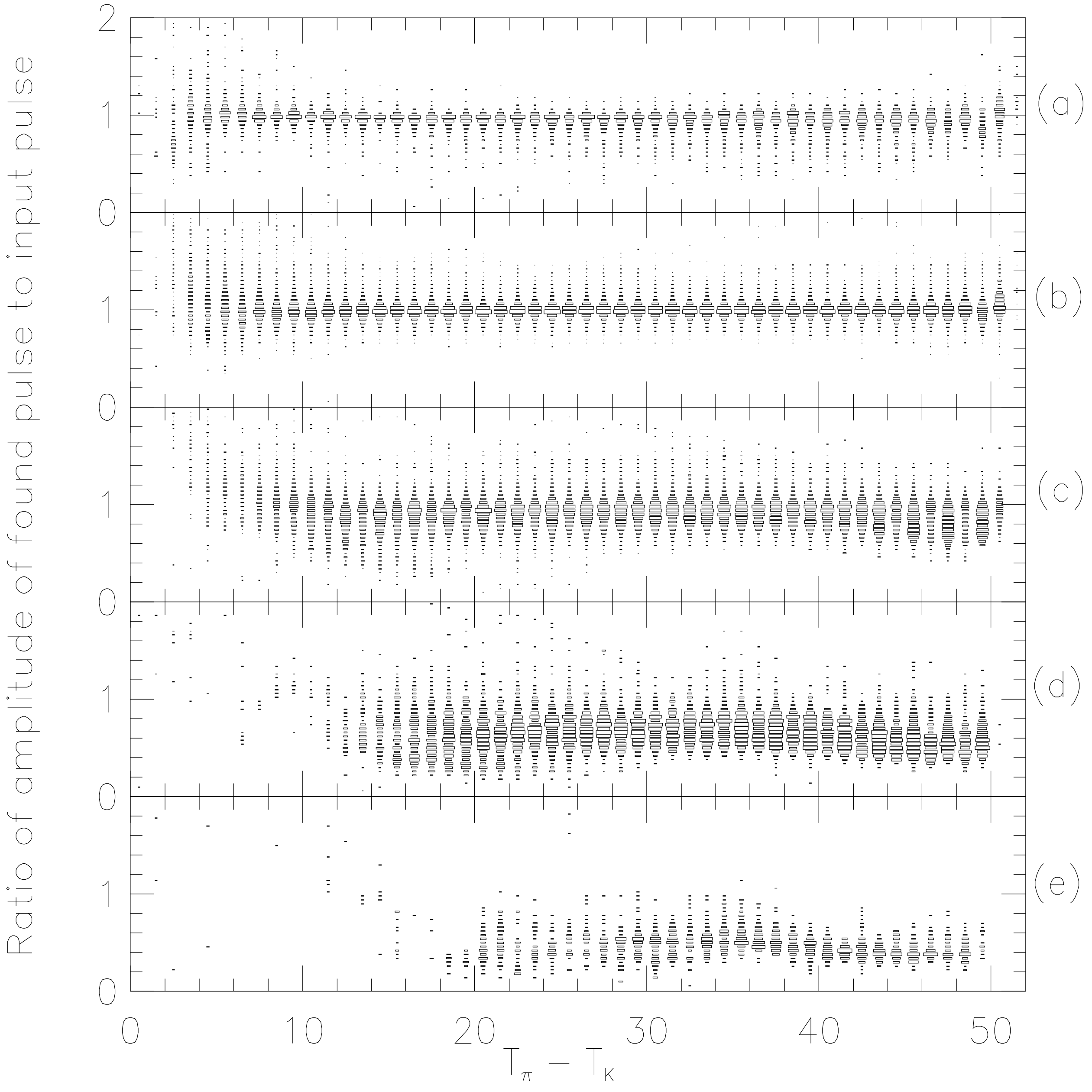

Figure 6: Ratio of amplitudes as a function of $(T_\pi - T_K)$ for high gain data, in bins of E_K(fiber) , (a) 0-5, (b) 5-15, (c) 15-25, (d) 25-40 and (e) > 40 MeV

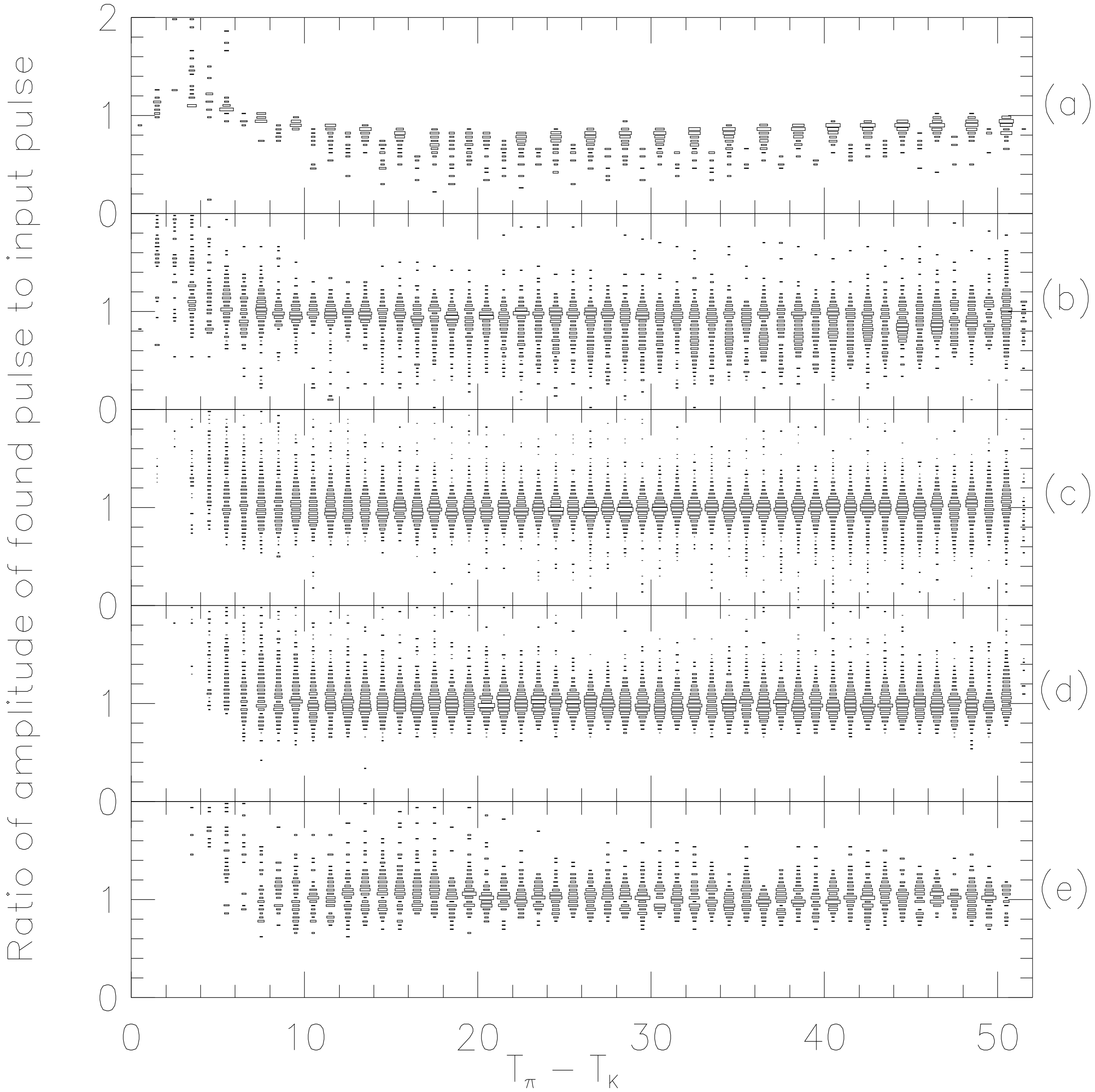

Figure 7: Ratio of amplitudes as a function of $(T_\pi - T_K)$ for low gain data, in bins of E_K(fiber) , (a) 0-5, (b) 5-15, (c) 15-25, (d) 25-40 and (e) $>$ 40 MeV

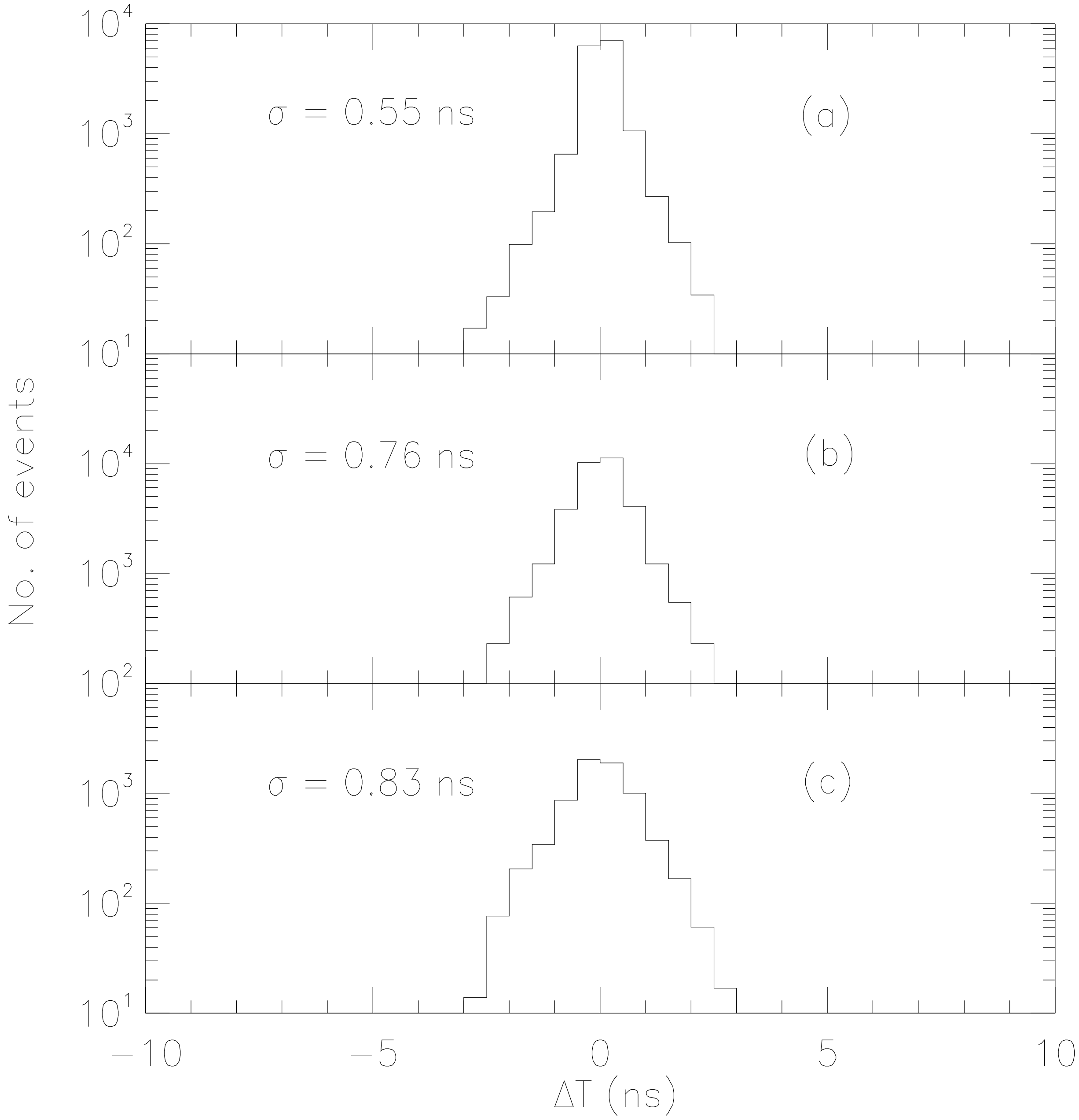

Figure 8: Difference between output and input values of $(T_\pi - T_K)$ for (a) High gain and E_K(fiber) = 0-5 MeV, (b) Low gain and E_K(fiber) = 15-25 MeV, and (c) Low gain and E_K(fiber) > 40 MeV