
GAP-2022-0??

New Baseline Algorithm for UUB Traces



Tobias Schulz, Markus Roth,
David Schmidt, and Darko Veberič

IAP, Karlsruhe Institute of Technology  Germany

October 2022

Abstract

To acquire an unbiased signal from a time trace, it is necessary to correctly identify and subtract a baseline. An improved algorithm for estimating the baselines of SD time traces has been previously developed for UB electronics [1]. However, these electronic boards will be exchanged with the upgraded unified boards (UUB) in the course of the ongoing AugerPrime upgrade, thus making it necessary to revisit and adjust the algorithm to the new electronics. The UUB electronics and its baseline namely have some substantially different behaviour with respect to the baseline which justifies revisiting the algorithm and documenting the modified procedure in this note.

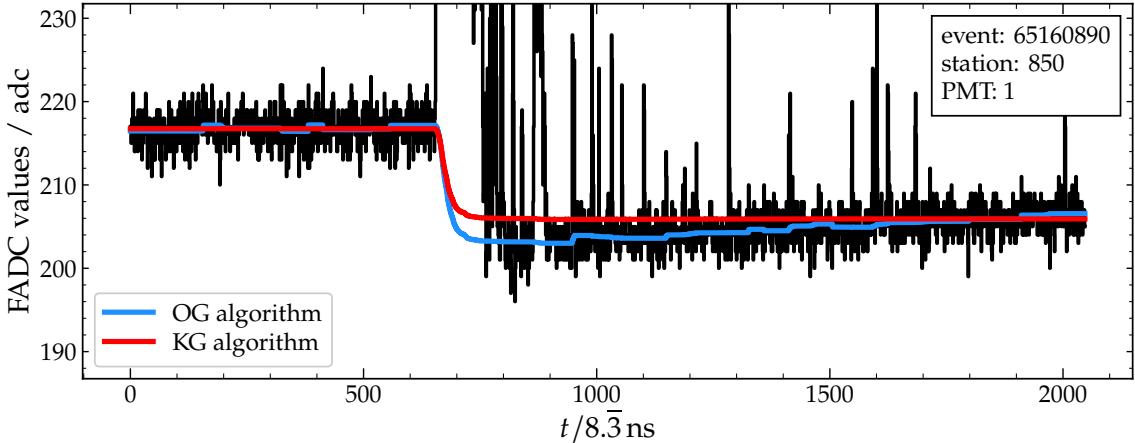


Figure 1: The KG algorithm for UBs computes a constant front and end piece of the time trace and makes a charge-linear interpolation between both of them. The OG algorithm tries to identify flat pieces in the trace and interpolates the baseline between these pieces, resulting in a varying baseline. For this example of a UUB trace both algorithms produce a sub optimal estimate of the baseline, justifying an improvement.

1 Introduction

The KG baseline algorithm was developed based on a physically motivated model of baseline behaviour. Without any signal present in a trace, the baseline of such a trace can be assumed to be constant. Nevertheless, after a significant signal, the baseline of photomultipliers drops, resulting in an undershoot that recovers exponentially with a characteristic decay time τ . For the UB the recovery of the undershoot takes about $300\ \mu\text{s}$, which is followed by an overshoot of the baseline, although with a smaller amplitude as for the undershoot. This overshoot then slowly decreases over a time span of up to 1 to $1.5\ \text{ms}$ [2]. The total trace length of UB traces is $19.2\ \mu\text{s}$ and thus, comparing this time to the aforementioned time scales, the baseline after the undershoot can be safely assumed constant. On the contrary, in the UUB, the recovery of the undershoot is clearly visible due to the shorter decay times, as can be seen in Fig. 1. The nominal decay times in the high gain (HG) are around $100\ \mu\text{s}$ and around $270\ \mu\text{s}$ in the low gain (LG). However, this excludes the AC coupling at the PMT base and thus diverging results are to be expected [3]. The KG baseline algorithm for UBs is not able to take the recovery of the undershoot into account. Namely, a constant front and end baseline estimate based on the truncated mean of the first and last 100 bins of the trace is calculated. Depending on the absolute difference ΔB between these estimates, a baseline is charge-linearly interpolated between the front and end estimates. On the other hand, the OG baseline algorithm identifies various flat pieces and interpolates between them, resulting in a varying baseline estimate that roughly follows the recovery seen in the UUB traces. However, it has been shown that the baseline estimates of the OG baseline algorithm result in false estimates due to the flat piece determination and should not be used [1]. Therefore, a modification of the current KG baseline algorithm is needed in order to take the shorter decay times of the UUB into account.

2 Determination of the decay time

The decay times of multiple events between October 2021 and June 2022 are extracted from the UUB traces.. The amplitude of the undershoot is proportional to the size of the signal that precedes them. A small signal will not produce a significant undershoot and thus the decay time can not be determined reliably. A cut on the charge of the preceding signal is applied to select events with a significant amplitude of the undershoot, such that the decay time can be fitted well.

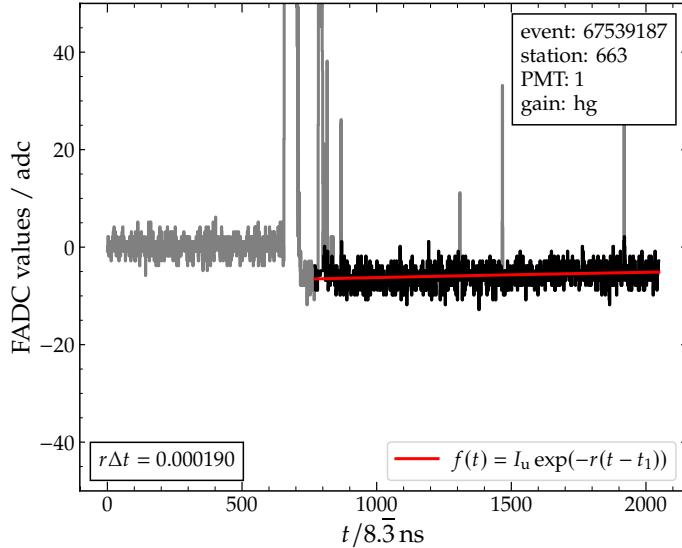


Figure 2: Example final fit of the decay rate r and undershoot amplitude I_u of a trace. After an initial fit, all signal contributions in the trace, that do not belong to the baseline, are removed (in grey) and a final fit of the decay function is performed.

A detailed description on how this cut is chosen is given in Appendix A. This leaves us with a set of 134313 HG traces. Since the data set is too small to include a sufficient number of LG traces with a significantly large undershoot, the following analysis is performed purely on HG traces. An exponential decay of the form of

$$f(t) = I_u \exp(-r(t - t_1)) \quad (1)$$

is used for the fit, where r is the decay rate (which is the inverse of the decay time) and I_u denotes the amplitude of the undershoot immediately after the preceding signal, i.e. at time t_1 . First, a robust baseline estimate of the front baseline is performed. A detailed description of this procedure is given in Ref. [1]. This estimate is then subtracted from the trace for the subsequent fit of the trace. As a rough approximation for t_1 we use the first bin after the trace maximum where the trace becomes negative and half a microsecond is added to avoid any remnants of the shower signal. The amplitude of the undershoot I_u and the decay rate r are determined in an initial fit. After the first fit has been performed, the trace can be cleaned from unwanted signal contributions to improve the quality of a second fit. Using the results from the initial fit, the decay term is subtracted from the trace. Similarly to the estimation of the robust baseline at the beginning of the trace, the mode of the trace is determined and all signal, exceeding a threshold of 2σ relative to the mode is removed. The fit of I_u and r is then repeated with the cleaned trace. The resulting fit can be seen in Fig. 2. A further selection of the data and identification of “bad” PMTs can be done by using both fit parameters I_u and r , and is discussed in Appendix B.

The distribution of all fitted decay rates r is shown in Fig. 3-left. The maximum of this distribution is around $r\Delta t = 1.91 \times 10^{-4}$, with 95% of the decay rates being in the interval of 0.7×10^{-4} to 3.4×10^{-4} . The sampling time of Δt for the UUB traces is 8.3 ns . For each station and PMT, the mode of $r\Delta t$ is determined, if at least 30 fits were performed. In Fig. 3-right the histogram and the mean is shown for a specific station and PMT.

Fig. 4 shows the estimated decay rates for a total of 451 PMTs from 183 different UUB stations. The mean of the distribution is at $r\Delta t = 1.85 \times 10^{-4}$ with a standard deviation of $\sigma = 0.23 \times 10^{-4}$. Since this distribution is rather narrow, a global decay rate $r\Delta t = 1.88 \times 10^{-4}$, based on the mean of the distribution, is assumed. This rate can be converted to a decay time τ , which evaluates to

$$\tau = \frac{1}{r} = \frac{\Delta t}{r\Delta t} \approx 45 \mu\text{s}. \quad (2)$$

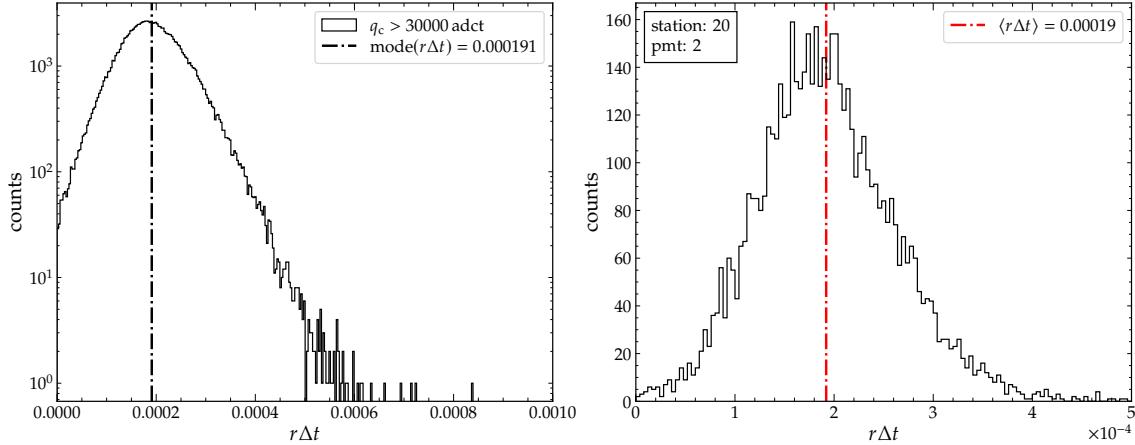


Figure 3: The distribution of $r\Delta t$ for all PMTs of all stations (left) and for a specific PMTs (right) covers a wide range of decay rates.

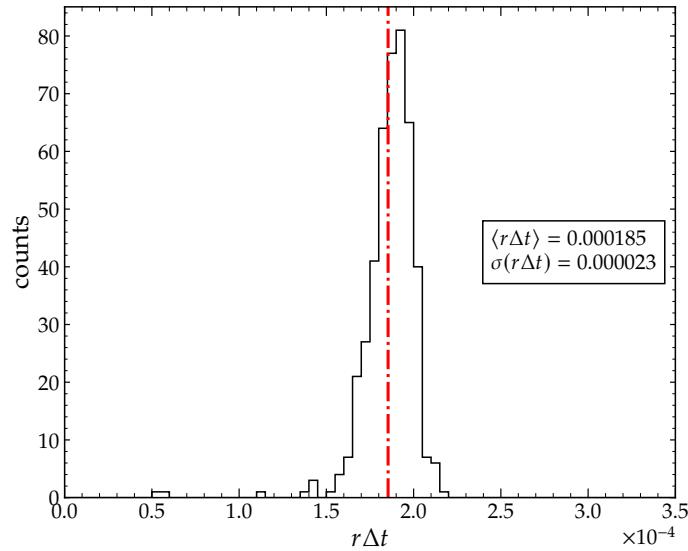


Figure 4: By calculating for each individual PMT of each station the most likely decay time, an universal constant, valid for all PMTs, can be estimated if the spread between the single PMTs is small enough. For 451 PMTs from 183 different UUB stations the average decay rate is $\langle r\Delta t \rangle = 1.85 \times 10^{-4}$ with a standard deviation of $\sigma = 0.23 \times 10^{-4}$, thus an universal and constant decay rate can be assumed for all PMTs.

3 Extension of the UB algorithm

A few changes have to be implemented in the KG baseline algorithm in order to correctly estimate the baselines of UUB traces. The full procedure of the KG algorithm for the UB baseline estimation has been explained in detail in Ref. [1]. While the total time length of the trace remains approximately the same, the sampling bin width is decreased from 25 ns down to 8.3 ns, which results in an increase of the number of bins from 768 to 2048. The sample window for the estimation of the front and end baseline is thus increased from 100 up to 300 bins. After both baseline parts have been determined, the undershoot ΔB and baseline error $\sigma_{\Delta B}$ are determined and an interpolation method is chosen. While for the UB algorithm a PMT is rejected if $\Delta B > +5\sigma_{\Delta B}$, for the UUB this threshold has to be raised due to an increase of the fluctuations in ΔB , as shown in Fig. 5. For

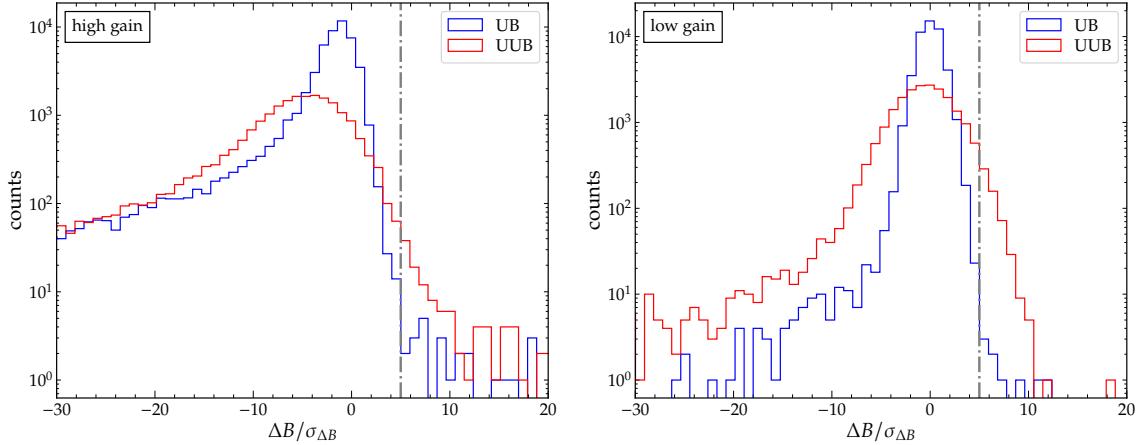


Figure 5: The distribution of $\Delta B/\sigma_{\Delta B}$ of 2 days of data for HG (left) and LG (right) traces varies for UB and UUB. For the UUB an increased fraction of traces have a ΔB larger than $+5\sigma_{\Delta B}$.

two days of data, two sets of 50 926 traces from UB and 21 657 traces from UUB were analyzed. In the UB 0.07% of the HG traces and 0.02% of the LG traces were with a ΔB larger than $+5\sigma_{\Delta B}$. For the UUB however 0.63% of the HG traces and 2.63% of the LG traces were with a ΔB larger than $+5\sigma_{\Delta B}$. This increase is a consequence of

Thus the upper threshold is chosen as $\Delta B > +10\sigma_{\Delta B}$. In the case of a significant undershoot with $0 > \Delta B \geq -1\sigma_{\Delta B}$, a minimum difference of 200 adc counts between the trace maximum and the front baseline estimate B_{front} is needed to execute the charge-linear interpolation of the baseline with the modification described below. After a time-linear interpolation of the baseline has been performed, yielding a baseline $b_i^{(0)}$, the baseline of the charge-linear interpolation can be calculated as

$$b_i^{(j)} = B_{\text{front}} + \epsilon_i^{(j)} \Delta B, \quad (3)$$

where ϵ_i is the fraction of the cumulative charge q_i in the total charge q_{ref}

$$\epsilon_i = \frac{q_i}{q_{\text{ref}}}, \quad (4)$$

where $q_{\text{ref}} \equiv q_{1898}$ is the integrated total charge with a decay kernel of the baseline-subtracted trace and runs all the way up to the center bin of the end baseline estimate, which is at the index $i = 1898$. Since there is a decay of the undershoot after the signal, an exponential kernel has to be added when calculating q_i so that

$$q_i^{(j)} = \sum_{k=0}^i (T_k - b_k^{(j-1)}) \exp\left(-\frac{\Delta t(i-k)}{\tau}\right) \quad (5)$$

where the previously determined universal decay time is $\tau = 45 \mu\text{s}$ and the indices run over the time bins of size Δt . In the case of large decay times, the exponential term becomes 1 and the equation is the same as for the UB baseline estimation. The UB algorithm can thus be reproduced by setting $r = 0$. The calculation of the baseline can be made more efficient with the following

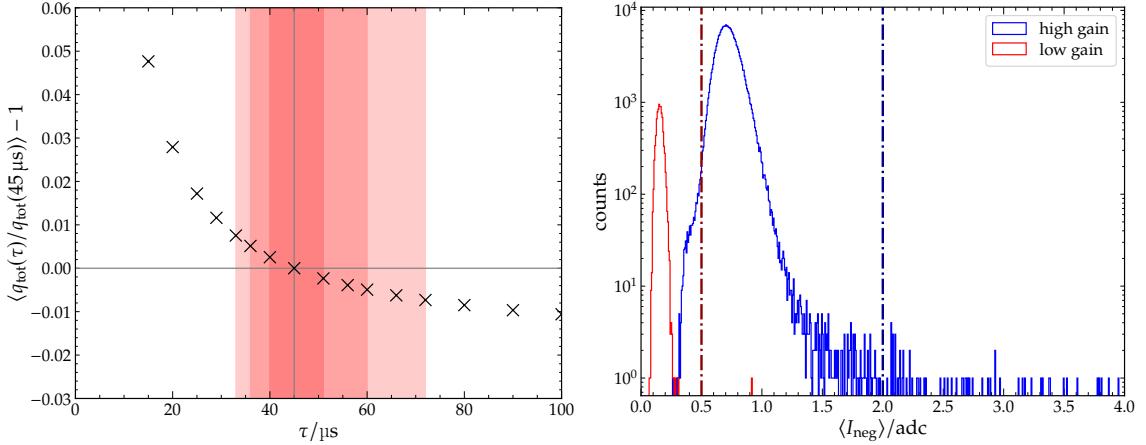


Figure 6: *Left:* Bias $\langle q_{\text{tot}}(\tau) / q_{\text{tot}}(45 \mu\text{s}) \rangle - 1$ of the total charge relative to the total charge for $\tau = 45 \mu\text{s}$. In red 1, 2, and 3σ deviations from $\tau = 45 \mu\text{s}$ are shown. *Right:* Mean negative amplitude $\langle I_{\text{neg}} \rangle$ of the baseline-subtracted trace for HG (blue) and LG (red). The threshold of the maximum allowed $\langle I_{\text{neg}} \rangle$ is plotted in dash-dotted lines. If $\langle I_{\text{neg}} \rangle$ should exceed the threshold, the PMT is rejected.

relation:

$$q_{i+1}^{(j)} = \sum_{k=0}^{i+1} (T_k - b_k^{(j-1)}) \exp\left(-\frac{\Delta t((i+1)-k)}{\tau}\right) \quad (6)$$

$$= (T_{i+1} - b_{i+1}^{(j-1)}) + \exp\left(-\frac{\Delta t}{\tau}\right) \sum_{k=0}^i (T_k - b_k^{(j-1)}) \exp\left(-\frac{\Delta t(i-k)}{\tau}\right) \quad (7)$$

$$= (T_{i+1} - b_{i+1}^{(j-1)}) + \exp\left(-\frac{\Delta t}{\tau}\right) q_i^{(j)}. \quad (8)$$

After a baseline has been determined, a check on the mean negative amplitude $\langle I_{\text{neg}} \rangle$ of the baseline-subtracted trace is performed [1]. For the UBS the thresholds of the maximum $\langle I_{\text{neg}} \rangle$ are set to the standard deviation of the traces $\sigma \approx 0.5$. For the UUBs the standard deviation of the LG traces does not change. However, the HG traces have a higher standard deviation of $\sigma \approx 2$. If $\langle I_{\text{neg}} \rangle$ exceeds the standard deviation, the PMT is rejected as bad PMT.

4 Evaluation

As shown in Section 2, the decay rate r , and thus the decay time τ , varies on the PMT level only to a small degree, justifying using only one universal decay rate for all PMTs for all stations. Nevertheless we have to check the systematic errors. For a set of 114 484 traces the baseline is estimated multiple times for a varying τ . The total charges $q_{\text{tot}}(\tau)$ of the traces are then calculated and compared to the trace with the baseline estimate that is performed with the universal decay time $\tau = 45 \mu\text{s}$. The resulting bias relative to $q_{\text{tot}}(45 \mu\text{s})$ is shown in Fig. 6-left. For decay times that are larger than the estimated universal decay time, the observed bias asymptotically approaches the limit of -2% . If the decay times are smaller the bias increases exponentially. At approximately $\tau = 20 \mu\text{s}$, the bias relative to the the universal decay time is at around $+2.7\%$. In Fig. 6-right the distribution of the mean negative amplitude $\langle I_{\text{neg}} \rangle$ of the baseline-subtracted trace is shown for HG and LG traces. More than 99% of the traces are below the threshold of 2 and 0.5 adc respectively.

Fig. 7 shows a comparison of estimated baselines between the OG algorithm, the KG algorithm, and the updated version of the KG algorithm. While previously the KG algorithm could not take

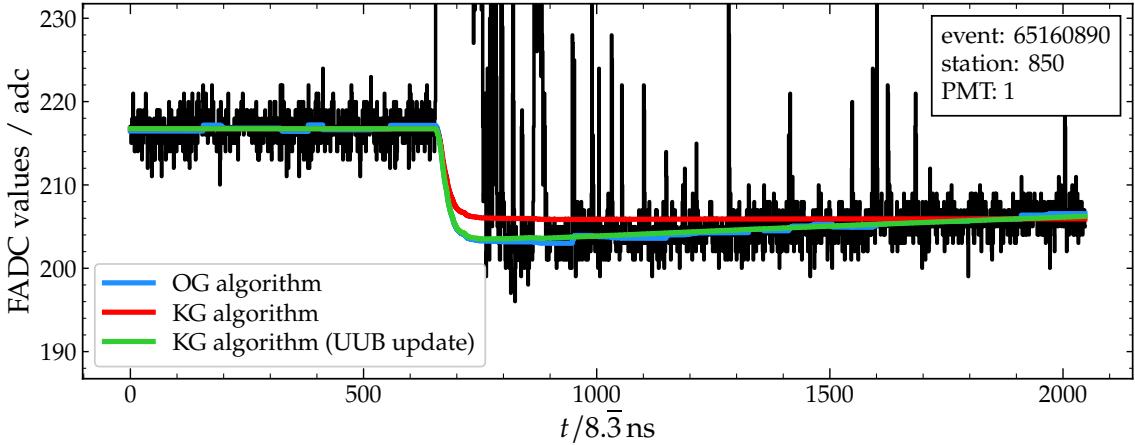


Figure 7: Comparison of the estimated baselines of OG algorithm (blue), KG algorithm (red), and the updated KG algorithm for UUBs (green). With the extension of the algorithm the recovery of the undershoot of the trace is now estimated correctly.

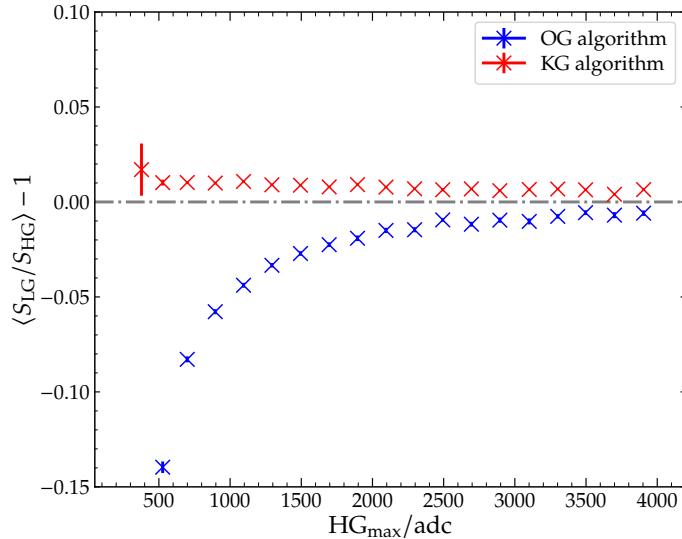


Figure 8: The ratio between signals from LG and HG traces as a function of the maximum bin entry in the HG channel HG_{max} , which provides an indication of how close the HG channel is to saturation. At the transition from HG to LG in the last bin 4095, the OG and KG algorithm for UUB both result in a similar bias of about 0.6%.

the undershoot recovery into account, the extension now allows for an undershoot recovery, using a constant universal decay time.

In Fig. 8 the LG to HG signal ratio is shown as a function of the HG trace maximum HG_{max} , using 41 065 traces from 6507 events between October 2021 and July 2022. At $HG_{max} = 4096$ the HG saturates and the LG is used instead. The bias in the last bin, and thus the bias at the transition from HG to LG, is for the OG algorithm at around -0.6% and for the KG algorithm at around 0.6% .

5 Conclusions

The KG baseline algorithm has been updated to properly estimate baselines for UUBs. An additional exponential kernel has been added to the charge integration to account for the recovery of the undershoot UUB traces. A universal decay constant τ that fits most PMTs has been determined to be $\tau = 45 \mu\text{s}$. An upper limit on the negative bias, relative to the determined baseline with $\tau = 45 \mu\text{s}$, is reached in the case of large decay times (as for UBs) at around 2%. For shorter decay times this bias increases exponentially with a positive bias of 2.7% at around $\tau = 20 \mu\text{s}$. We have shown that this updated algorithm reconstructs the trace shape and its baseline well and that the gain ratio $\langle S_{\text{LG}}/S_{\text{HG}} \rangle$ is on average biased only by 0.6% at the transition point from HG to LG.

References

- [1] T. Schulz, M. Roth, D. Schmidt and D. Veberič, *New Baseline Algorithm for SD Traces of the UB*, Auger internal note GAP-2022-045, 2022.
- [2] B. Genolini, T. Nguyen Trung and J. Pouthas, *Base line stability of the Surface Detector PMT base*, Auger internal note GAP-2003-051, 2003.
- [3] D. Nitz, Private communications.
- [4] I. Lhenry-yvon, *About bad PMTs : inventory of spurious data remaining after the two first levels of PMT quality cuts*, Auger internal note GAP-2018-022, 2003.
- [5] I. Lhenry-yvon and P. Papenbreer, *Bad PMTs: the third levels of quality cuts, tuned for PMT traces analysis*, Auger internal note GAP-2018-029, 2018.

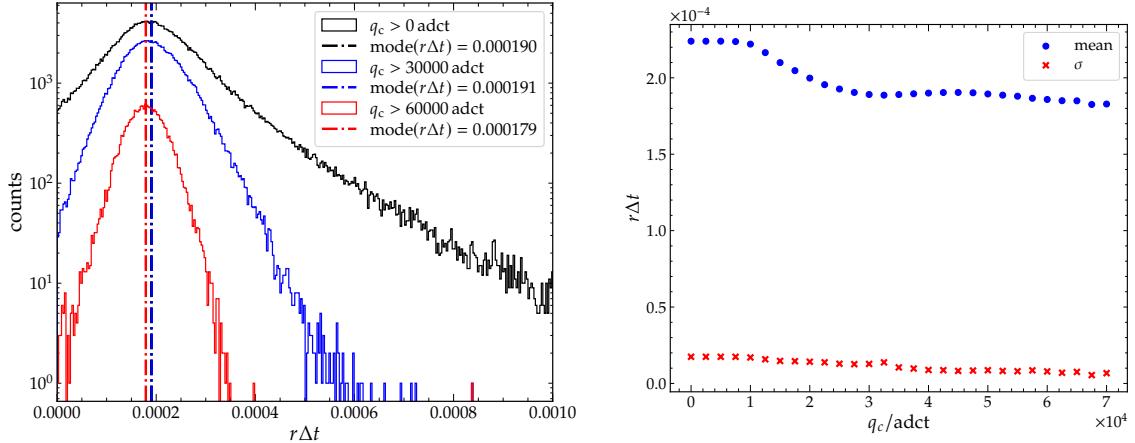


Figure 9: *Left:* Binned distribution of the decay rates $r\Delta t$ for various cuts on the signal charge q_c of the trace. The cut removes cases where the undershoot is not well developed and thus the decay is hard to observe. This cut effectively cleans the data set so that only well resolved decay of the undershoot is fitted. *Right:* The mean and standard deviation σ of the determined constant universal decay rate $r\Delta t$ for various cuts on the signal charge q_c of the trace. The mean of the universal decay rate decreases from $r\Delta t = 2.2 \times 10^{-4}$ down to approximately $r\Delta t = 1.8 \times 10^{-4}$ with an increasing cut on q_c .

A Signal Cut

Since the amplitude of the undershoot is proportional to the size of the signal that precedes them, a small signal will not produce a significant undershoot and thus the decay time can not be determined reliably. In order to determine a reasonable cut on the preceding signal, the full procedure of Section 2 is repeated. At first, for all traces the signal charge q_c of the trace is approximately determined as

$$q_c = \sum_{i=0}^{t_1} (T_i - B_{\text{front}}). \quad (9)$$

The end of the integration t_1 is chosen as the first bin after the trace maximum, where the trace becomes negative with $0.5 \mu\text{s}$ added to exclude potential signal contributions of the shower. A cut value for q_c is then chosen to study the development of the decay times. In Fig. 9-left the distribution of all fitted decay times is shown for three different cuts on q_c . With increasing cut on minimal q_c , the amount of fits with very long decay times seems to be reduced. Since the amplitude of the undershoot correlates with the signal size, the decay of the undershoot becomes more visible and easier to fit. The most probable decay time is estimated for each distribution of decay rates of individual PMTs. In the last step the average of the individual PMT decay times is calculated. Fig. 9-right shows the calculated average decay times as well as standard deviation for increasing cut on q_c . With increasing signal size the mean decay rate decreases from $r\Delta t = 2.2 \times 10^{-4}$ down to approximately $r\Delta t = 1.8 \times 10^{-4}$.

B Identification of “bad” PMTs

The two parameters I_u and $r\Delta t$ of the trace fits from Section 2 can be used not only to check if the fit was successful, but also to identify “bad” PMTs. Fig. 10-top left shows the distribution of both fit parameters I_u and $r\Delta t$ for each individual trace in a scatter plot. Most of the fits are in the lower right quadrant, which is the expected region for the two fit parameters when the trace has a normal shape and the fit was successful. Due to the expected undershoot, the parameter I_u is negative and the recovery of the undershoot is given by a positive rate $r\Delta t$. Outliers can be read off

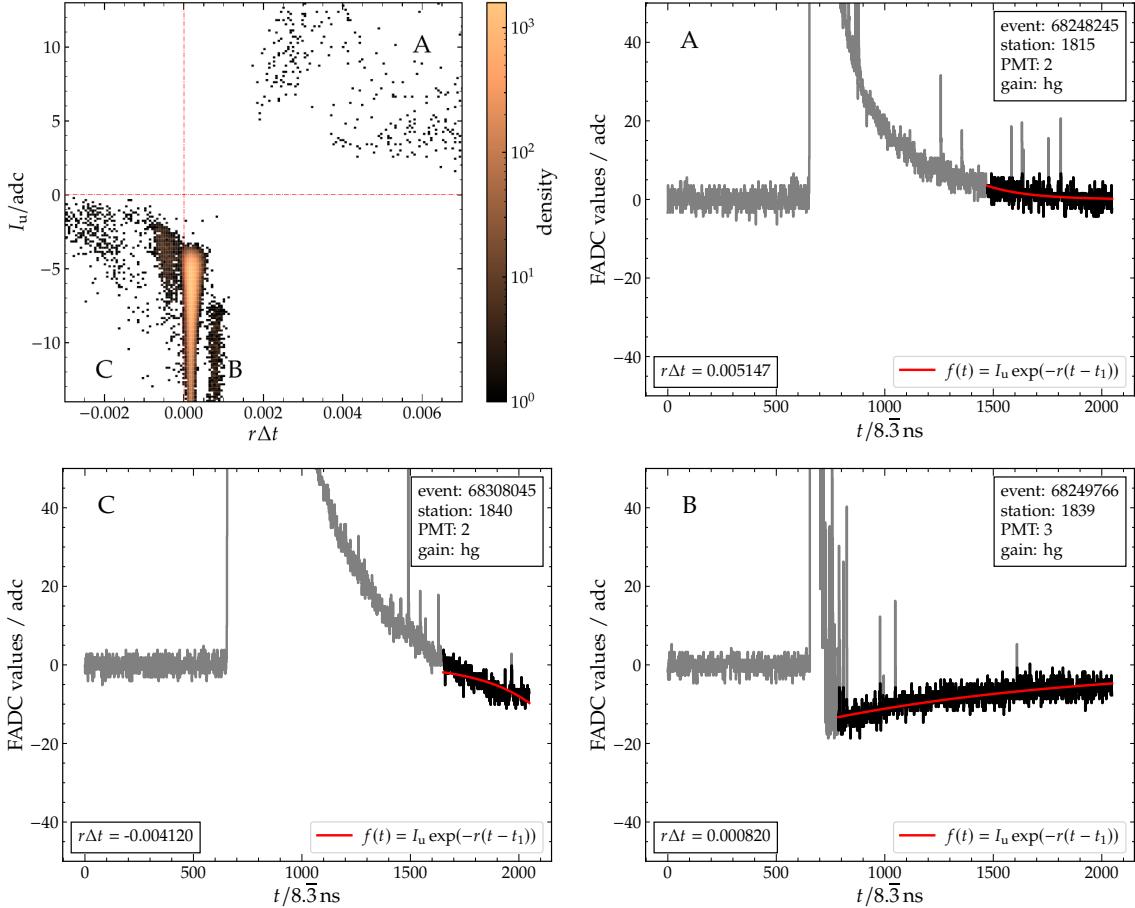


Figure 10: *Top left:* Distribution of the fit parameters I_u and $r\Delta t$ of the traces. The main group is located in the lower right quadrant with a negative I_u and a positive decay rate $r\Delta t$. Outliers can be easily identified due to their deviation from the main group. *Top right:* Anomalous trace with a fit of a long tail. The fit parameters are found in this case in the upper right quadrant of the scatter plot. *Bottom left:* Anomalous trace with an unsuccessful fit of a long tail. These fit parameters are found in the lower left quadrant of the scatter plot. *Bottom right:* Trace with an anomalously-fast recovery. These fit parameters can be seen as second distribution, right to the main distribution in the lower right quadrant.

of the plot as deviations from the most dense accumulation of data points in the upper right and lower left quadrant, as well as a second distribution to the right of the main group. The outliers in the upper right and lower left quadrant can be attributed to traces with a long, anomalous tail after the signal. Two example traces for both quadrants are shown in Fig. 10-top left and Fig. 10-bottom right, where either a successful fit of the long tail is performed, or the fit fails due to the long tail of the trace. For these two cases two PMTs of two separate stations (PMT 2 of station 1815 and PMT 2 of station 1840) have been identified to repeatedly show such anomalous behavior and have been removed from the data set of this analysis. A procedure to identify these PMTs is explained in [4] and [5]. In the lower right quadrant a smaller, second distribution can be seen to the right of the main group. This distribution has been identified to originate as well from one single PMT (PMT 3 of station 1839). The decay time of these traces is about $\tau = 10\mu s$, which is over 4 times smaller than expected decay $\tau = 45\mu s$, which was determined for most of the PMTs.