Charge reconstruction in large-area photomultipliers

M. Grassi^{a,b,1} **M. Montuschi**^{c,d,2} **M. Baldoncini**^{c,d} **F. Mantovani**^{c,d} **B. Ricci**^{c,d} **G. Andronico**^{e}
V. Antonelli^{a} **M. Bellato**^{f,g} **E. Bernieri**^{h,i} **A.Brigatti**^{a,k} **R. Brugnera**^{f,g} **A. Budano**^{h,i} **M.
Buscemi**^{e,j} **S. Bussino**^{h,i} **R. Caruso**^{e,j} **D. Chiesa**^{n,o} **D. Corti**^{h,i} **F. Dal Corso**^{f,g} **X. F. Ding**^{a,p} **S.
Dusini**^{f,g} **A.Fabbri**^{h,i} **G.Fiorentini**^{c,d} **R. Ford**^{q,a} **A. Formozov**^{a,k,l,m} **G. Galet**^{f} **A. Garfagnini**^{f,g}
M. Giammarchi^{a,k} **A. Giaz**^{f,g} **A. Insolia**^{e,j} **R. Isocrate**^{f,g} **I. Lippi**^{f,g} **F. Longhitano**^{e} **D. Lo
Presti**^{e,j} **P. Lombardi**^{a,k} **F. Marini**^{f,g} **S. M. Mari**^{h,i} **C. Martellini**^{h,i} **E. Meroni**^{a,k} **M. Mezzetto**^{f,g}
L. Miramonti^{a,k} **S. Monforte**^{e} **M. Nastasi**^{n,o} **F. Ortica**^{r,s} **A. Paoloni**^{t} **S. Parmeggiano**^{a,k} **D.
Pedretti**^{f,g} **N. Pelliccia**^{r,s} **R. Pompilio**^{a,k} **E. Previtali**^{n,o} **G. Ranucci**^{a,k} **A. C. Re**^{a,k} **A. Romani**^{r,s}
P. Saggese^{a,k} **G. Salamanna**^{h,i} **F. H. Sawy**^{f,g} **G. Settanta**^{h,i} **M. Sisti**^{n,o} **C. Sirignano**^{f,g} **M.
Spinetti**^{t} **L. Stanco**^{f,g} **V. Strati**^{c,u} **G. Verde**^{e} **L. Votano**^{t}

^{a}*Istituto Nazionale di Fisica Nucleare, Sezione di Milano,*
  Via Celoria 16, I-20133 Milano, Italy

^{b}*Institute of High Energy Physics,*
  19B YuquanLu, 1000049 Beijing, P.R.China

^{c}*Dipartimento di Fisica e Scienze della Terra, Università di Ferrara,*
  Via Saragat 1, I-44122 Ferrara, Italy

^{d}*Istituto Nazionale di Fisica Nucleare, Sezione di Ferrara,*
  Via Saragat 1, I-44122 Ferrara, Italy

^{e}*Istituto Nazionale di Fisica Nucleare, Sezione di Catania,*
  Via S. Sofia 62, I-95125 Catania, Italy

^{f}*Dipartimento di Fisica e Astronomia, Università di Padova,*
  Via Marzolo 8 , I-35131 Padova, Italy

^{g}*Istituto Nazionale di Fisica Nucleare, Sezione di Padova,*
  Via Marzolo 8 , I-35131 Padova, Italy

^{h}*Dipartimento di Matematica e Fisica, Università di Roma Tre,*
  Via della Vasca Navale 84 , I-00146 Roma, Italy

^{i}*Istituto Nazionale di Fisica Nucleare, Sezione di Roma Tre,*
  Via della Vasca Navale 84 , I-00146 Roma, Italy

^{j}*Dipartimento di Fisica e Astronomia, Università di Catania,*
  Piazza Università 2, I-95131 Catania, Italy

^{k}*Dipartimento di Fisica e Astronomia, Università di Milano,*
  Via Celoria 16, I-20133 Milano, Italy

^{l}*Joint Institute for Nuclear Research,*
  141980 Dubna, Russia

^{1}Now at APC Laboratory - IN2P3, Paris, France
^{2}Corresponding author

^{m}*Lomonosov Moscow State University Skobeltsyn Institute of Nuclear Physics,*
  119234 Moscow, Russia

^{n}*Dipartimento di Fisica, Università di Milano Bicocca,*
  P.zza della Scienza 3 , I-20126 Milano, Italy

^{o}*Istituto Nazionale di Fisica Nucleare, Sezione di Milano Bicocca,*
  P.zza della Scienza 3 , I-20126 Milano, Italy

^{p}*Gran Sasso Science Institute,*
  Via Crispi 7 , I-67100 L'Aquila, Italy

^{q}*SNOLAB,*
  Lively, ON, P3Y 1N2 Canada

^{r}*Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, via Elce di Sotto 8 ,*
  I-06123 Perugia, Italy

^{s}*Istituto Nazionale di Fisica Nucleare, Sezione di Perugia,*
  Via Pascoli , I-06123 Perugia, Italy

^{t}*Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati,*
  Via Fermi 40 , I-00044 Frascati (RM), Italy

^{u}*Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Legnaro,*
  Viale dell'Università 2 , I-35020 Legnaro (PD), Italy

  E-mail: montuschi@fe.infn.it

ABSTRACT: Large-area PhotoMultiplier Tubes (PMT) allow to efficiently instrument Liquid Scintillator (LS) neutrino detectors, where large target masses are pivotal to compensate for neutrinos' extremely elusive nature. Depending on the detector light yield, several scintillation photons stemming from the same neutrino interaction are likely to hit a single PMT in a few tens/hundreds of nanoseconds, resulting in several photoelectrons (PES) to pile-up at the PMT anode. In such scenario, the signal generated by each PE is entangled to the others, and an accurate PMT charge reconstruction becomes challenging. This manuscript describes an experimental method able to address the PMT charge reconstruction in the case of large PE pile-up, providing an unbiased charge estimator at the permille level up to 15 detected PES. The method is based on a signal filtering technique (Wiener filter) which suppresses the noise due to both PMT and readout electronics, and on a Fourier-based deconvolution able to minimize the influence of signal distortions —such as an overshoot. The analysis of simulated PMT waveforms shows that the slope of a linear regression modeling the relation between reconstructed and true charge values improves from 0.769 ± 0.001 (without deconvolution) to 0.989 ± 0.001 (with deconvolution), where unitary slope implies perfect reconstruction. A C++ implementation of the charge reconstruction algorithm is available online at [1].

Contents

1 Introduction

The consolidated Liquid Scintillator (LS) technology is driving neutrino physics into the era of precision calorimetry. The unprecedented scientific achievements of Borexino [2], Daya Bay [3], Double Chooz [4], KamLAND [5] and RENO [6] experiments have been the trailblazer for a new generation of multi-kiloton detectors (JUNO [7], Jinping [8], RENO50 [9], SNO+ [10], ANDES [11]). Going beyond the open issues in particle physics, the perspective will be some spin-off in applied antineutrino physics [12].

The experimental challenges of neutrino calorimetry revolve around improving both the energy and the spatial resolution of (anti)neutrino interaction detection. Since both resolution terms scale with the number of detected scintillation photons, maximizing the detector photocoverage —namely the sensitive area of the detector— is pivotal to improve them. In many current and future detectors, technical and budget constraints make the use of large-area PhotoMultiplier Tubes (PMT) the only viable solution to achieve large photocoverage. In some cases, these constrains even justify a R&D program dedicated to develop a novel PMT technology [13]. The largest PMT bulbs built to date are 20-inch diameter. Because of their large acceptance, they typically detect many photoelectrons (PE) per scintillation event, which are likely to pile up at the readout level. That is, the spacing between the PEs' time of arrival is lower or comparable to the width of a single-photoelectrons (SPE) pulse.

When PE pulses overlap, their identification becomes challenging. Especially if the pulses are affected by an overshoot (a distortion in the PMT output signal described in section 2), two subsequent PEs could easily be mis-reconstructed as a single one. A biased PMT charge reconstruction not only compromises the linearity —and therefore the resolution— of the detector-wise energy estimator, but also threatens the time-based reconstruction of the event vertex. In a large detector, a precise knowledge of the event vertex is crucial to define a fiducial volume meant to reject background energy depositions arising from natural radioactivity. The relevance of this issue can be further appreciated by noting that the Daya Bay experiment, after 4 years of smooth data taking, developed a new readout system based on fast digitizers, and its associated waveform reconstruction, to better assess the linearity of the detector energy response [14].
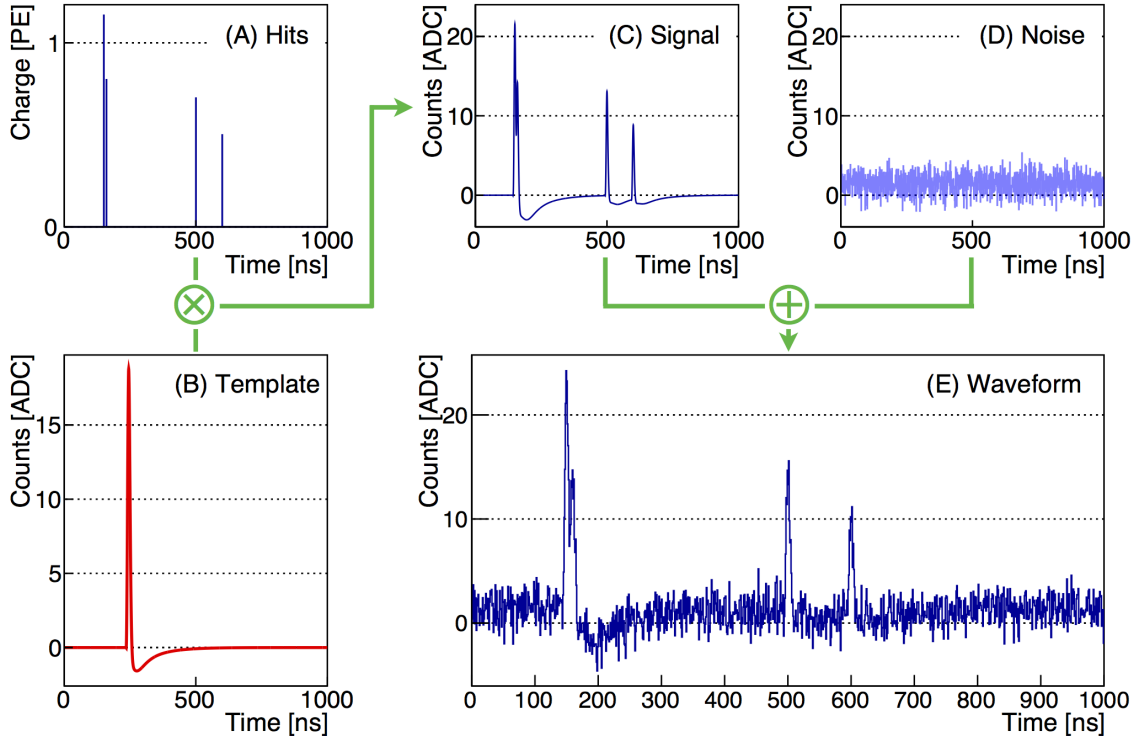
Figure 1. Building blocks of the waveform simulation. Panel A shows the hit generation, where for each detected PE a random charge value is generated. Panel B shows the analytical shape of a single PE template. Panel C is the result of the convolution between the SPE template and the hits. Panel D contains the PMT simulated noise. Panel E shows the final PMT waveform resulting from the sum of the signal and of the noise components.

The goal of this study is twofold: (I) to propose an open-source detector-independent charge reconstruction algorithm [1] processing the output of a generic fast digitizer (FADC) connected to a PMT, and (II) to define a procedure to assess the accuracy of the reconstructed charge, especially in the case of large PE pile-up. While (I) is based on realistic signal and noise assumptions, and particular care was devoted to model most of the PMT peculiarities, (II) is meant to allow a comparison between any charge reconstruction algorithm. It is worth mentioning that our algorithm is specifically designed to minimize those charge reconstruction biases introduced by the presence of an overshoot in the PE pulses, and by the noise fluctuations embedded in the PMT output pulse.

2   PMT Waveform Simulation

We build a PMT waveform simulation with the aim to develop and validate the charge reconstruction algorithm with a known input signal. To avoid to rely on the specification of a given manufacturer and/or model, we implement a parametric simulation of a generic PMT response. The emission of a PE from the photocathode, its collection by the anode, and its subsequent amplification are simulated as an instantaneous PMT charge output. In accordance to experimental data [15], the charge value q

Table 1. Parameters used in eq. 2.1 together with their nominal values.

Parameter	Description	Value
q_o	SPE calibrated gain	1 PE
q_p	Pedestal cutoff	0.3 PE
σ	SPE Gauss width	0.3 PE
ω	Under-amplified PE fraction	0.2
τ	Exponential decay constant	0.5 PE

is generated randomly according to the distribution $f(q)$ in eq. 2.1, adapted from [16]

$$f(q) = \begin{cases} (1-\omega)\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(q-q_0)^2}{2\sigma^2}\right] + \frac{\omega}{\tau}\exp-\frac{q}{\tau} & q \geq q_p \\ 0 & q < q_p \end{cases} \tag{2.1}$$

The weight ω determines the relative contribution of a Gaussian distribution with respect to an exponential tail, modeling the faction of under-amplified PEs. The width of the former (σ) is set to 30% to describe the typical uncertainty induced by the first dynode amplification ($\sim 1/\sqrt{10}$). The Gaussian mean value (q_0) is set to 1, as in the case of a perfect gain calibration. The exponential tail accounts for low amplitude hits due to elastically scattered and backscattered electrons from the first dynode [16]. The parameter q_p sets the threshold for the minimum charge to be considered in order to avoid to simulate the pedestal. The values of all the parameters used in eq. 2.1 are listed in table 1, and the analysis of their variability is described in section 4. The time at which a charge output occurs is randomly sampled from a flat distribution spanning a 1 μs-long time window, and each charge-time pair is defined as a *hit*.

A sequence of hits, resulting from several PEs impinging on the same PMT, are shown as instantaneous pulses in figure 1A. The pulse position on the horizontal axis represents the hit time with respect to a reference t_0 —for example a global detector trigger—, and the pulse amplitude represents the PMT output charge. We assume the PMT to be connected to a fast electronics able to sample its output at 1 GSample/s, and we describe the voltage drop resulting from the detection of each hit with a log-normal function (eq. 2.2) [15, 17]. The AC coupling often used to split the PMT high-voltage from the PMT output signal might induce a distortion in the PMT output waveform. We refer to the component of the signal with a polarity opposite to the log-normal as the overshoot, and we follow [15] to describe it using the sum of a Gaussian (eq. 2.3) and an exponential tail (eq. 2.4). The values of the parameters used in eq. 2.2, 2.3 and 2.4 are listed in table 2. The values related to the SPE amplitude are chosen to obtain an approximate 0.5 mV/count resolution, resulting in a dynamic range of ~50PE when using a commercial 10-bit digitizer. The overshoot values are adapted from the Daya Bay experience [15]. The analytical shape of a SPE-waveform is reported in eq. 2.5 and shown in figure 1B. We name this shape the SPE template. The signal-only waveform resulting from the hits shown is figure 1A is built by convolving each hit with the SPE template, and it is shown in figure 1C.

$$U_{\text{peak}}(t) = U_0 \cdot \exp\left(-\frac{1}{2}\left(\frac{\ln(t/\tau_0)}{\sigma_0}\right)^2\right) \tag{2.2}$$

Table 2. Parameters involved in the signal (top rows) and noise (bottom rows) simulation of the PMT waveform. Parameters for which the "Range" column is filled are generated randomly according to a flat probability density function at the beginning of the simulation.

Parameter	Value	Range	Description	Function
U_0	20 ADC			
σ_0	0.15		SPE template	eq. 2.2
τ_0	30 ns			
U_1	-1.2 ADC			
σ_1	55 ns		Overshoot	eq. 2.3
t_1	-4 ns			
U_2	-2.8 ADC		Overshoot	eq. 2.4
τ_2	80 ns			
μ_N	1.5 ADC		Baseline	Gaussian
σ_N	1 ADC		White Noise	
n_{FF}	5		N(Components)	Fixed Frequency
f_{FF}		[1, 480] MHz	Frequency	
A_{FF}		[0.1, 0.3] ADC	Amplitude	

$$U_{OS1}(t) = U_1 \cdot \exp\left(-\frac{1}{2}\left(\frac{t - t_1}{\sigma_1}\right)^2\right) \tag{2.3}$$

$$U_{OS2}(t) = U_2 \cdot \frac{1}{\exp\left(\frac{50\,\text{ns} - t}{10\,\text{ns}}\right) + 1} \cdot \exp\left(-\frac{t}{\tau_2}\right) \tag{2.4}$$

$$U(t) = U_{\text{peak}} + U_{OS1} + U_{OS2} \tag{2.5}$$

To make the simulation more realistic, we add a noise waveform (figure 1D) to the waveform built using only signal hits. The simulated noise includes a Gaussian component and some periodic components with fixed frequency. The former describes an overall baseline offset and time-uncorrelated baseline fluctuations. The Gaussian mean (μ_N) is set to 1.5 ADC counts, and the Gaussian width (σ_N) to 1 ADC count. The fixed-frequency components describe potential noise sources embedded in, or due to, the readout circuit. We simulate 5 such components (n_{FF}), with a random amplitude generated flat in the 0.1-0.3 ADC counts range, and with a random frequency in the 1-480 MHz range. The ultimate waveform comprising signal hits and all noise components is shown in figure 1E.

3 Charge Reconstruction

The simplest way to determine the PMT output charge is to sum up all the waveform samples collected in a readout window. However, this approach would embed both the noise and the overshoot into the charge measurement, resulting in a rough and biased charge estimator. The charge reconstruction algorithm described in this section and sketched in figure 2 is meant to mitigate the role of both noise and overshoot. It comprises two steps: (I) the Deconvolution Algorithm (DA), which filters
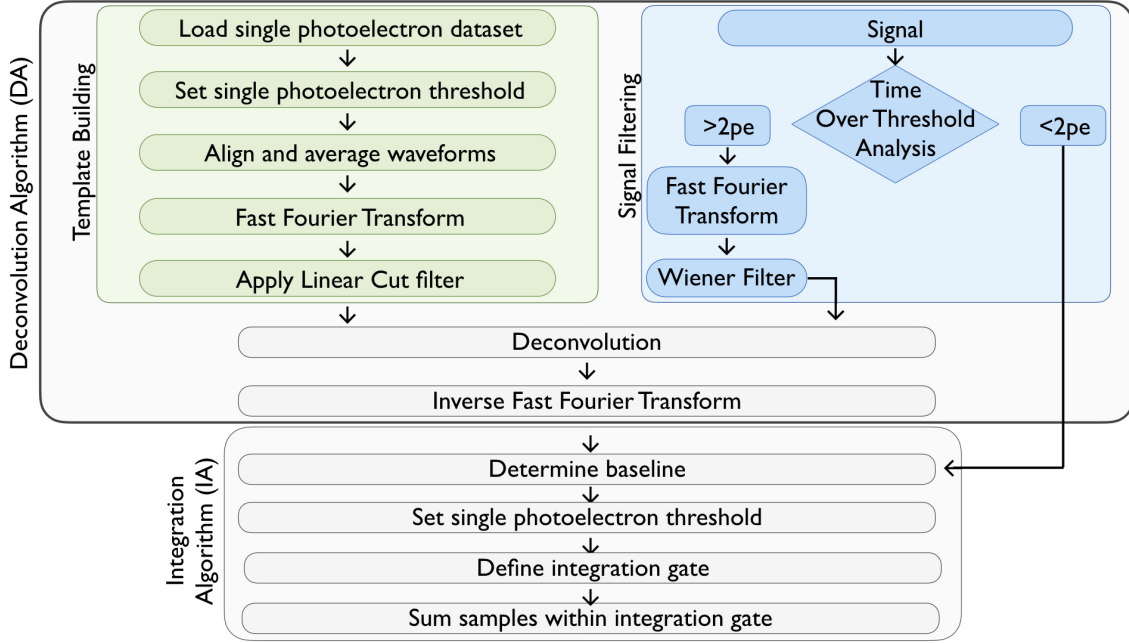
Figure 2. Schematic diagram of the charge reconstruction algorithm. The SPE template in the time domain is computed by averaging 10^4 SPE waveforms that simulate a LED-based calibration dataset. A Fast Fourier Transform (FFT) is applied to both the SPE template and the PMT waveform to be reconstructed. The latter is processed using the Wiener Filter to minimize the noise and suppress the overshoot. The SPE template in the frequency domain is then used as a benchmark pattern to deconvolve the filtered waveform in the frequency domain. We eventually process the deconvolved waveform with an Inverse FFT, so that the waveform in the time domain could be integrated to compute the PMT output charge.

the raw waveforms reducing the noise, and deconvolves SPE templates out of the filtered waveforms; and (II) the Integration Algorithm (IA), which integrates the deconvolved waveforms with the aim to determine the overall PMT output charge.

Both filtering and deconvolution are data-driven methods that need to be trained and tested. To this end, we rely on waveforms generated using the simulation package described in section 2. We produce two large datasets, one comprising mostly SPE waveforms (S dataset), and one with waveforms containing multiple hits (M dataset). In order to emulate a real experimental setup, the charge reconstruction algorithm is designed not to use the true hit information (time and charge). We unblind such parameters only when assessing its performance —as described in section 4. The first dataset emulates the behavior of a low-intensity LED placed in front of a PMT to measure its SPE response, where one typically gets either 0 hits or a single hit in the resulting waveform (p(0)≈0.3, p(1)≈0.6, and p(>1)≈0.1). In the second dataset, waveforms are generated such that the number of true hits follows a flat distribution in the [0-15] range. This distribution is chosen to investigate the performance of the reconstruction algorithm even in those events experiencing high pile-up.

The noise pattern present in simulated waveforms consists mostly of bin-to-bin uncorrelated fluctuations. On the contrary, a SPE pulse lasts for a few tens of ns. The difference between the two time scales is highlighted in figure 3, where the frequency spectra of a waveform with no hits (pure noise, red curve) and a waveform with 4 hits (blue curve) are compared. The former is rather
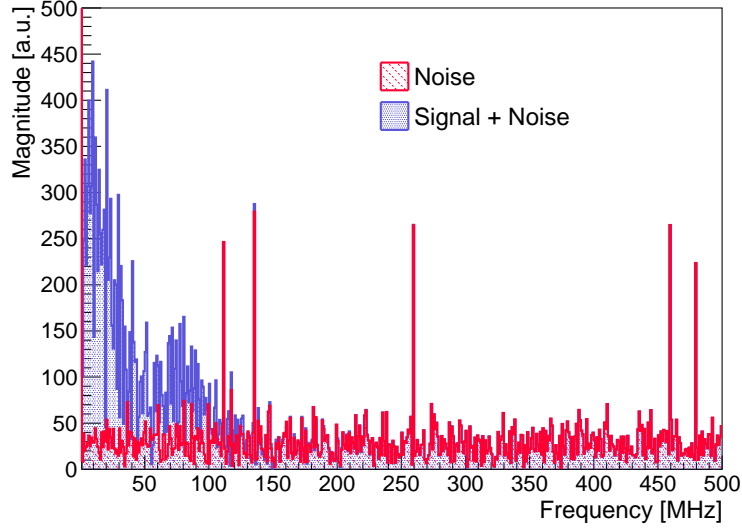
Figure 3. Frequency power spectra of a pure-noise waveform (red) and a complete waveform comprising noise and PE pulses (blue). The two spectra are used in eq. 3.1 to derive the Wiener Filter. The spikes in the noise spectrum are due to the fixed-frequency noise components.
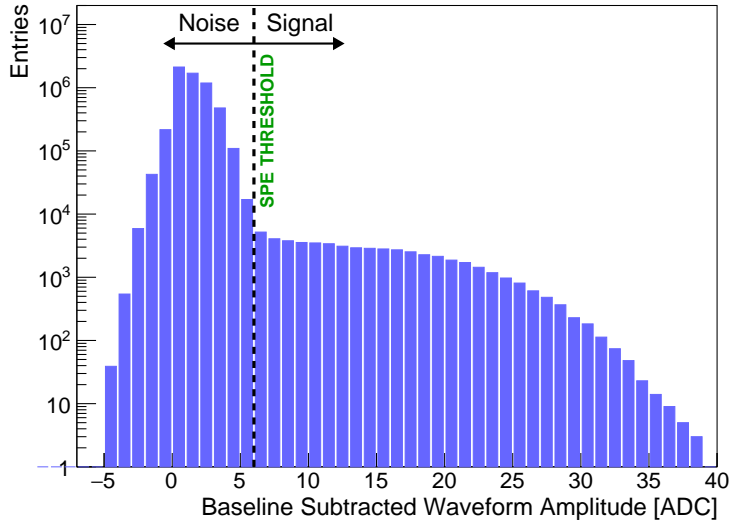

Figure 4. Distribution of all the baseline-subtracted FADC samples in a waveform, built using 10^4 waveforms containing mostly 0 or 1 hit (dataset S). Noise fluctuations are responsible for the peak centered at zero. FADC samples related to the fraction of the waveform where SPE pulses occurred are responsible for the tail extending to large ADC values. A 6 ADC trigger threshold is chosen to identify waveforms where at least 1 hit is present.

flat, with the exception of the fixed-frequency noise components, which manifest as spikes in the frequency domain. The latter, on the contrary, is peaked at low frequencies, as a result of SPE pulses being much slower than the noise fluctuations. We exploit such a difference to suppress the noise by

means of a Wiener Filter —a technique commonly employed in signal processing [18]. The filter is defined by the following kernel equation

$$H[f] = \frac{|S[f]|^2}{|S[f]|^2 + |N[f]|^2} \tag{3.1}$$

where $S[f]$ and $N[f]$ are the frequency spectra of the signal and of the noise. In literature, this filter is referred to as the *optimal* linear filter for the removal of additive noise [19]. Namely, the coefficients $H[f]$ of the Wiener Filter are calculated to minimize the average squared distance between the filter output and the desired signal [20]. Such coefficients are then used to weight the frequency components of the waveform being processed.

Eq. 3.1 can be rearranged by dividing both the numerator and the denominator by the noise power spectrum $|N[f]|^2$, and by substituting the variable $\text{SNR}[f] = |S[f]|^2/|N[f]|^2$. This manipulation yields

$$H[f] = \frac{\text{SNR}[f]}{\text{SNR}[f] + 1} \tag{3.2}$$

where SNR is a ratio of the signal power to the noise power. Eq. 3.2 makes the frequency response of the filter more intuitive, being a real positive number in the range $0 \leq H[f] \leq 1$. Frequencies that are barely affected by noise ($\text{SNR}[f] \to \infty$) result in the filter being close to unity, hence applying little or no attenuation to the input components. On the contrary, frequencies that are severely affected by noise ($\text{SNR} \sim 0$) result in the filter to heavily attenuate them ($H[f] \sim 0$). In summary, the Wiener Filter attenuates each frequency component in proportion to an estimate of its signal-to-noise ratio.

The deconvolution is a technique meant to identify and resolve the presence of SPE pulses within a filtered waveform, with the aim to perform an unbiased measurement of the PMT output charge. It is implemented as a division in the frequency domain between the waveform to be reconstructed and the template of a SPE pulse. We build the template by time-aligning and averaging 10^4 SPE waveforms selected from the S dataset, which emulates calibration data collected by illuminating the PMT with a low intensity LED. The selection determining which waveforms are to be used in the template building is based on a threshold-crossing criterion: only waveforms whose baseline-subtracted amplitude exceeds 30% of the mean SPE amplitude are retained. The distribution of all the digitized samples in the S dataset (1000 samples per waveform) is shown in figure 4, where the selection threshold clearly marks the transition between the noise region and the signal region. To further clean the SPE template from those noise contributions not suppressed by the averaging process, all frequencies above 120 MHz are stripped.

The outcome of the deconvolution performed on the waveform in figure 1E is shown in figure 5. Here the deconvolved waveform is brought back to the time domain by means of an Inverse Fast Fourier transform, and it shows narrow pulses whose amplitude is proportional to the original hit charge. However, rather than using the pulse amplitude, we find a better estimator of the hit charge to be the integral of the pulse.

To minimize the contribution of residual noise to the pulse integral, and the possible biases arising from the ringing visible in the vicinity of the pulses (Gibbs effect), only waveform samples falling within a Region Of Interest (ROI) are summed up to yield the final charge value. A ROI is opened any time the deconvolved waveform crosses a threshold corresponding to 30% of the
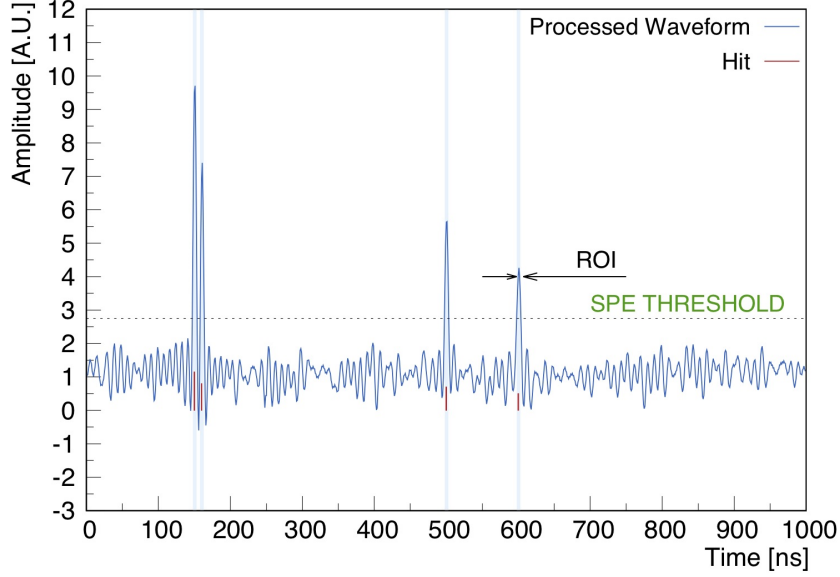
Figure 5. Waveform arising from filtering and deconvolving figure 1E. True hits are shown for reference in red. The charge of each PE is reconstructed by integrating the bins falling within a ROI (shaded areas) defined starting from the time at which the waveform crosses a trigger threshold (dashed line). The threshold is set to be 30% of the amplitude of a SPE pulse.

amplitude of a deconvolved SPE pulse. We define the minimal ROI to be 6 ns wide, and we extend it if at its end the waveform is still above the threshold.

Following an extensive optimization procedure, we determine that our reconstruction algorithm is better suited to process waveforms with more than two hits. Below this value, the CPU power required to filter and deconvolve the waveform is not paying off in terms of charge reconstruction performance. As a consequence, we implement a waveform preselection to determine how to process each waveform. We define Δt to be the time during which the SPE template stays over threshold (6 ns). When a waveform is found to be over threshold for at least $2\,\Delta t$, it undergoes the full charge reconstruction comprising filtering and deconvolution. Otherwise, a simple integral of the bins over threshold is used to compute the reconstructed charge.

4 Results and discussion

The performance of the charge reconstruction algorithm is assessed by processing a set of 10^4 simulated waveforms containing a random number of PE in the range [0,15], namely the M dataset. In particular we aim to show how the algorithm improves the precision and accuracy of the charge estimate with respect to a simple integration.

To quantify the performance of the algorithm, we process the M dataset waveforms twice, once using the IA alone, and once using both the DA and the IA. For each of the two reconstruction

approaches we build a correlation plot using pairs of reconstructed and true charge values, as shown in figure 6. An ideal reconstruction algorithm would result in the two quantities to be maximally correlated, and a linear regression would yield unitary slope and null intercept. Any deviation from such behavior is therefore to be interpreted as a bias in the reconstruction. In particular, a non-null intercept (q) models any bias that does not depend on the number of reconstructed PE, while a non-unitary slope (m) models any PE-dependent bias effectively compromising the linearity of the reconstruction algorithm. To further measure how scattered the reconstructed charge values are with respect to the true charge values, we use the Pearson correlation coefficient (ρ). We interpret the latter as an estimator of the charge reconstruction precision, while we interpret the linear regression coefficients as a measurement of its accuracy. By comparing the two plots shown in figure 6, it can be noticed that the algorithm significantly improves both the precision and the accuracy of the charge reconstruction. Indeed, m improves from 0.769 to 0.989, q improves from 0.540 to 0.053, and ρ improves from 0.979 to 0.988. To make the meaning of these numbers more evident, we report that if waveforms with 5 and 10 true PEs are processed with the IA alone, the average reconstructed charge is biased by 7% and 12% respectively. While, in the case of DA+IA, the bias becomes negligible (at permille level) in both cases. Figure 7 additionally shows the distribution of the true and reconstructed charge values for events with a number of PEs ranging between 1 and 15.

We further assess the resilience of the reconstruction algorithm to possible distortions in the input waveform. The data taking of a typical neutrino experiment spans indeed several years —if not decades—, during which several aging issues might compromise the initial PMT performance. Such issues often affect both the noise level and the shape of SPE pulses. To evaluate how the precision and the accuracy of the charge reconstruction degrades due to variations in the PMT waveforms with varying shape, we produce new datasets in which all the PMT input parameters are smeared by 10%, one at a time. In analogy to the procedure described above, (I) we process all these datasets using the full reconstruction algorithm (IA+DA); (II) for each dataset we build the reco-true charge correlation plots; (III) we use these plots to perform a linear regression and to compute the Pearson correlation coefficient. The discrepancy between the resulting values and the nominal values are plotted in terms of residuals in figure 8. Some considerations follow.

- Given the different role that ρ, m and q play in assessing the algorithm performance, only residuals within the same panel can be compared. That is, residuals here are meant to draw a hierarchy among the parameters describing the SPE shape, with the aim to show which of them affects the reconstruction algorithm the most.

- From the left panel it is evident that injecting SPEs with varying amplitude (U_0) compromises the algorithm precision. This is expected, since the SPE template now fails to describe the SPE pulses present in the PMT output waveforms.

- The parameters defining the width of the SPE shape, namely τ_0 and σ_0, heavily affect the accuracy of the linear regression parameters.

- The parameters shaping the overshoot (U_1, σ_1, t_1, U_2, τ_2) play a negligible role, as a consequence of the successful overshoot stripping by the DA.
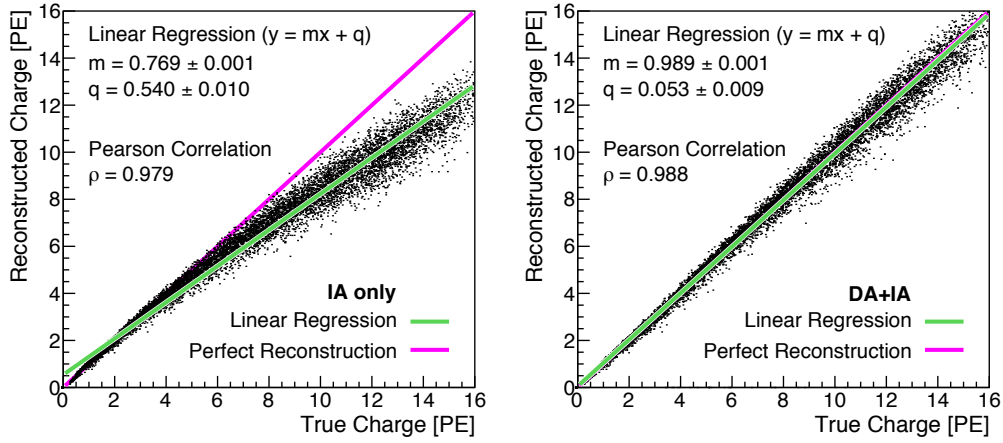
Figure 6. Reconstructed charge versus true charge for 10^4 PMT waveforms. The reconstructed charge is computed using the IA alone (left) and DA+IA (right). In both panels the green line is the result of a linear regression performed on the 10^4 data points.

- The white noise amplitude (σ_N) is able to introduce a non-negligible charge bias, suggesting that the implementation of the filter has room to be improved.
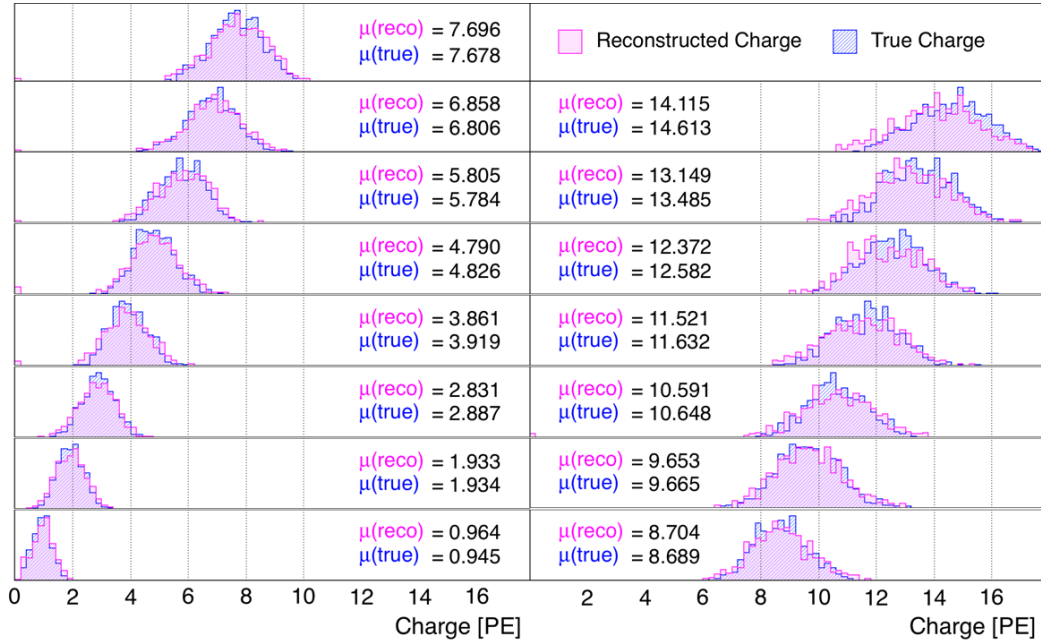

Figure 7. Distribution of the true charge (blue) and of the reconstructed charge (pink) using the full reconstruction algorithm (DA + IA). Each plot is built using waveforms with a defined number of PES ranging from 1 (bottom left) to 15 (top right). The mean true charge is consistently lower than the true number of PES because of the exponential tail used to simulate the SPE distribution in eq. 2.1.
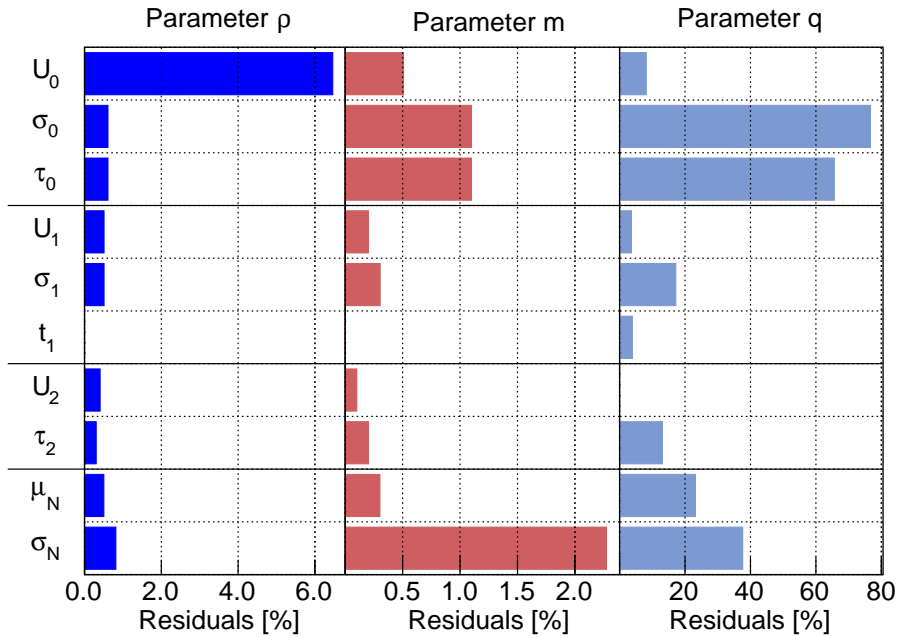
Figure 8. Residuals of the three performance parameters (linear regression coefficients and correlation coefficient) assessing how the charge reconstruction worsens when the input waveforms contain pulses that are different from the SPE template. The input waveforms here are produced by applying a 10% smearing to each of the parameters listed on the vertical axis of the plot.

5 Conclusions

This manuscript describes a new method to reconstruct the output charge of a PMT when sampled with a fast digitizer. Its originality stands in putting together several well established signal processing techniques with the aim to improve the charge reconstruction accuracy over a large dynamic range, in terms of both mean value of the reconstructed charge and of dispersion around the mean.

The algorithm comprises a filtering step, a deconvolution step, and an integration step, which aim to reduce the PMT noise, to compensate for predictable distortions in the PMT waveform, and to infer the number of PES detected by the PMT. Some of its features are summarized here below.

- The filter allows to mitigate not only white noise, but also noise at fixed frequencies often introduced at the level of the readout electronics (e.g. due to grounding issues).

- The deconvolution is based on a SPE template, which can be automatically derived from dark count events, effectively introducing a zero dead-time calibration procedure. In the case of an experiment using different PMT types, the algorithm can handle all of them naturally by computing a different template for each type. Moreover, the capability to build new templates continuously allows to account for any variation in the PMT performance over time.

- The deconvolution allows an a-priori estimate of the time needed to reconstruct each waveform. The time needed to perform the most CPU-intensive operation (Fast Fourier Transform) depends only on the number of samples within a waveform, and not on its complexity. On the

contrary, an approach based on reconstructing the charge by means of an analytic fit would become slow and unreliable in case of large pile-up. Such consideration becomes even more relevant in detectors instrumented with a large number of PMTs.

- The deconvolution effectively reduces the undesired SPE features from the PMT waveform (such as the overshoot), which can in principle bias the charge reconstruction, making the integration step much more robust.

We tested the reconstruction algorithm by analyzing waveform datasets produced with a custom-made PMT simulation. The latter allowed us to simulate a generic PMT readout starting from an analytic SPE signal template.

We investigated the effect of processing waveforms with a shape different from the SPE template. We did it by producing a sample of waveforms where the parameters describing the SPE pulses were smeared randomly by 10% around their nominal value. The aim was to test the resilience of the algorithm to any change in the PMT waveform due, for instance, to aging issues. We determined that the reconstruction algorithm is differently sensitive to the input parameters. In particular, those describing the amplitude and the duration of a SPE pulse are the most likely to bias the reconstruction performance.

The overall charge reconstruction algorithm here described is geared to play a key role in improving the energy resolution of LS detectors with large photocoverage. Such detectors are indeed expected to be severely affected by energy-related systematic uncertainties stemming from the charge reconstruction of PMT waveforms where several PEs pile up. To provide the community with the possibility to test our reconstruction algorithm on different inputs, and to compare its performance to different reconstruction tools, we made a C++ implementation available at [1].

Acknowledgments

References

[1] C++ implementation of the charge reconstruction algorithm described in this manuscript available at
http://www.fe.infn.it/CRA

[2] Borexino Collaboration, G. Alimonti et al., *Neutrinos from the primary proton-proton fusion process in the Sun*, *Nature* **512** (2014) 383-386.
http://dx.doi.org/10.1038/nature13702 .

[3] Daya Bay Collaboration, F. P. An et al., *Observation of Electron-Antineutrino Disappearance at Daya Bay*, 108, 171803 *Phys. Rev. Lett.* **108** (2012) 171803.
http://dx.doi.org/10.1103/PhysRevLett.108.171803 .

[4] Double Chooz Collaboration, Y. Abe et al., *Improved measurements of the neutrino mixing angle θ_{13} with the Double Chooz detector*, *JHEP* **10** (2014) 086 [*Erratum ibid* **02** (2015) 074].
http://dx.doi.org/10.1007/JHEP10(2014)086 .

[5] KamLAND Collaboration, A. Gando et al., *Reactor on-off antineutrino measurement with KamLAND*, *Phys. Rev. D* **88** (2013) 033001.
https://doi.org/10.1103/PhysRevD.88.033001 .

[6] RENO Collaboration, J.H. Choi et al., *Observation of Energy and Baseline Dependent Reactor Antineutrino Disappearance in the RENO Experiment*, *Phys. Rev. Lett.* **116** (2016) 211801.
https://doi.org/10.1103/PhysRevLett.116.211801 .

[7] JUNO Collaboration, Fengpeng An et al., *Neutrino Physics with JUNO*, *J. Phys. G: Nucl. Part. Phys.* **43** (2016) 030401.
http://dx.doi.org/10.1088/0954-3899/43/3/030401 .

[8] Wu, Yu-Cheng et al., *Measurement of cosmic ray flux in the China JinPing underground laboratory*, *CPC* **37** (2013) 086001.
https://doi.org/10.1088/1674-1137/37/8/086001 .

[9] Soo-Bong Kim, *New results from RENO and prospects with RENO-50*, *Nucl. Part. Phys. Proc.* **265-266** (2015) 93-98 .
http://www.sciencedirect.com/science/article/pii/S2405601415003661?via%3Dihub .

[10] SNO+ Collaboration, S. Andringa et al., *Current Status and Future Prospects of the SNO+ Experiment*, *High Energy Phys.* **2016** (2016) 6194250.
http://dx.doi.org/10.1155/2016/6194250 .

[11] P. A. N. Machado, T. Muhlbeier, H. Nunokawa, and R. Zukanovich Funchal, *Potential of a neutrino detector in the ANDES underground laboratory for geophysics and astrophysics of neutrinos*, *Phys. Rev. D* **86** (2012) 125001.
https://doi.org/10.1103/PhysRevD.86.125001 .

[12] N.S. Bowden, K.M. Heeger, P. Huber, C. Mariani, R.B. Vogelaar, *Applied Antineutrino Physics 2015 - Conference Summary*, arXiv:1602.04759 (2016).

[13] Y. Wang, *Large Area MCP-PMT and its Application at JUNO*, Talk at XVII International Workshop on Neutrinos Telescopes (NeuTel 2017) , Venice, Italy, 2017, available at https://agenda.infn.it/contributionDisplay.py?sessionId=15&contribId=55&confId=11857 .

[14] Y. Huang et al, *The Flash ADC system and PMT waveform reconstruction for the Daya Bay Experiment*, arXiv:1707.03699 (2017).

[15] S. Jetter et al, *PMT waveform modeling at the Daya Bay experiment*, *CPC* **36** (2012) 93-98.
https://doi.org/10.1088/1674-1137/36/8/009

[16] R. Dossi, A. Ianni, G. Ranucci, O.Ju. Smirnov, *Methods for precise photoelectron counting with photomultipliers*, *Nucl. Instrum. Methods Phys. Res. A* **451** (2000) 623-637
http://dx.doi.org/10.1016/S0168-9002(00)00337-5 .

[17] J. Caravaca et al, *Experiment to demonstrate separation of Cherenkov and scintillation signals*, *Phys. Rev. C* **95** (2017) 055801 https://doi.org/10.1103/PhysRevC.95.055801

[18] Steven W. Smith , *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, San Diego, U.S.A. (1998).

[19] S.V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, John Wiley & Sons Ltd., Chichester U.K. (2000)

[20] A.V Oppenheim, G.C. Verghese, *Signals, Systems, and Inference*, *Class notes "Introduction to Communication, Control and Signal Processing"*, MIT 2010
https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/
6-011-introduction-to-communication-control-and-signal-processing-spring-2010/
readings/MIT6_011S10_chap11.pdf