The Cockcroft Institute
of Accelerator Science and Technology

Cockcroft-12-27

A non-linear algorithm for current signal filtering and peak detection in SiPM

M Putignano, A Intermite and C P Welsch
Cockcroft Institute, Daresbury, U.K.
University of Liverpool, Oxford Street, Liverpool, U.K.

A non-linear algorithm for current signal filtering and peak detection in SiPM

2012 JINST 7 P08014

(http://iopscience.iop.org/1748-0221/7/08/P08014)

PUBLISHED BY IOP PUBLISHING FOR SISSA MEDIALAB

RECEIVED: *June 8, 2012*
ACCEPTED: *July 19, 2012*
PUBLISHED: *August 16, 2012*

A non-linear algorithm for current signal filtering and peak detection in SiPM

M. Putignano,[1] **A. Intermite and C.P. Welsch**

Cockcroft Institute,
Daresbury, U.K.
University of Liverpool,
Oxford Street, Liverpool, U.K.

E-mail: m.putignano@liverpool.ac.uk

ABSTRACT: Read-out of Silicon Photomultipliers is commonly achieved by means of charge integration, a method particularly susceptible to after-pulsing noise and not efficient for low level light signals. Current signal monitoring, characterized by easier electronic implementation and intrinsically faster than charge integration, is also more suitable for low level light signals and can potentially result in much decreased after-pulsing noise effects. However, its use is to date limited by the need of developing a suitable read-out algorithm for signal analysis and filtering able to achieve current peak detection and measurement with the needed precision and accuracy.

In this paper we present an original algorithm, based on a piecewise linear-fitting approach, to filter the noise of the current signal and hence efficiently identifying and measuring current peaks. The proposed algorithm is then compared with the optimal linear filtering algorithm for time-encoded peak detection, based on a moving average routine, and assessed in terms of accuracy, precision, and peak detection efficiency, demonstrating improvements of $1 \div 2$ orders of magnitude in all these quality factors.

KEYWORDS: Pattern recognition, cluster finding, calibration and fitting methods; Data processing methods; Data reduction methods

[1]Corresponding author.

doi:10.1088/1748-0221/7/08/P08014

Contents

2012 JINST 7 P08014

1 Introduction

1.1 Silicon Photomultipliers

Sensors capable of detecting single photons have found different applications in diverse fields such as astronomy [1, 2], laser ranging [3], biomedical imaging [4], Positron Emission Tomography (PET) [5–7], optical time domain reflectometry [8] and beam loss detection in particle accelerators [9] to name a few, replacing in such applications the former use of photomultiplier tubes [10].

These developments call for low cost, small dimensions and high amplification photon counting detectors, and have triggered a growing interest in recent years on the development of solid state technologies, in particular related to Avalanche Photo Diodes (APDs). Amongst these detectors, the highest amplification, sufficient to discriminate single photons, is achieved by APDs operated well above the breakdown voltage, in the so called *Geiger mode*, known as Single Photon Avalanche Diodes (SPADs) [11] .

In a SPAD, a single impinging photon which creates an electron hole pair in the depletion region initiates a self sustaining avalanche process fed by the large reverse voltage, creating a signal large enough to be detected. The avalanche is then switched off, or *quenched*, by suitable circuitry, until no current is flowing anymore, the junction capacitance recharges, and the device

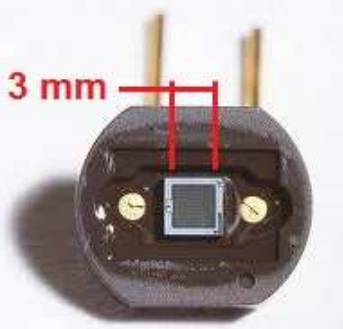

Figure 1. Photograph of a SiPM. The individual cells forming the arrays are visible.

is ready for another detection. However, SPADs are intrinsically digital devices, as the number of photons creating electron-holes pairs within the same avalanche event does not influence the amplitude of the signal, thus preventing their use for single shot light flux measurements.

Increase in the dynamic range of these sensors has been demonstrated by replicating the SPAD structure into many cells connected in parallel to a single output channel, resulting in the so called Silicon Photo Multiplier (SiPM). Bi-dimensional, closely packed arrays of up to 500 independent SPADs per square millimeter are now available commercially, an example of a SPAD array is shown in figure 1. Such arrays provide an output signal proportional to the incident light, responding with macroscopic currents of some mA to each photon detected. The SiPM retains the photon counting ability of the SPAD while outperforming it dramatically for dynamic range and recovery time, hence allowing use of this technology in the fields requiring single shot light flux measurements. Nowadays SiPMs are widely used in diverse applications including fluorescence analysis [12, 13], fluorescence lifetime measurement [14], single molecule detection [15] and PET [16]. Furthermore, due to their small dimensions and rather low intrinsic radioactivity, SiPM are also considered to be promising candidates for light detectors in rare event searches applications.

1.2 SiPM noise

One important limit to the successful application of SiPM, however, is constituted by noise. Use of high-quality materials and state of the art device fabrication techniques minimize the number of impurities and defects that contribute to noise; however, it still stays significant, especially in low light conditions, making noise an important branch of research in the field [17, 18]. Indeed, high levels of noise make it impossible, when light levels are modest or measurement precision is a priority, to reliably measure light fluxes, and limit SiPM applicability in these cases to on-off photon detection [19].

Noise which significantly contributes to counting errors is SiPMs is caused by three phenomena:

- electron-hole pairs created in the depletion layer by random thermal ionization (*dark count*).
- parasitic avalanche triggering by photons created during a primary avalanche and migrated to a neighboring cell (*optical cross-talk*). Optical cross talk has been reported to be sensibly

2012 JINST 7 P08014

reduced for SiPM featuring optical trenches: strips of material with different refraction index placed between neighboring cells, which deflect photons away from the active area [18].

- time delayed release of a hot carrier by a trap level due to imperfections in the lattice, leading to a time delayed second avalanche phenomenon (*after-pulsing*).

One important effect of these noise sources is that dark noise and afterpulsing signals in particular are indistinguishable from a *real* event, and can therefore lead to erroneous counting. This is not immediately applicable to cross-talk noise, whose contribution is to overestimate a given event amplitude, rather than creating spurious counts. However, by virtue of their high dynamic range, SiPMs are virtually insensitive to dark noise when used in presence of medium to high intensity short light pulses of a few ns duration, as the combined signal coming from several cells firing at once due to inbound photons easily overshadows the signal from dark noise random firing cells. The other two noise sources however cannot be ignored, and can be corrected, by adjusting the expected signal distribution of the SiPM, only within the statistical effects of shot noise [18, 20–22].

Reduction of cross-talk noise can be obtained by modifying the structure of the SiPM at the manufacturing stage with the inclusion of optical trenches; after-pulsing is instead more difficult to deal with, and it also increases when lowering the temperature in the junction, which results otherwise in the positive effect of reducing the dark noise. Nevertheless, the error associated to after-pulsing is mainly due to the readout procedure commonly used for SiPM, which relies on charge integration over a defined gate time: all events happening within this gate time are hence superimposed, so that afterpulsing and signal are no more distinguishable.

An alternative read-out procedure, based on direct detection of the signal peak current in the time domain, would instead allow discrimination of after-pulsing noise from signal, and ultimately more reliable operation of SiPM for light flux measurement. However, such procedure faces the challenge of reliably extracting the signal information from the complex waveform constituting a SiPM current signal (see e.g. the signal shown in figure 11). Standard linear noise filtering algorithms, indeed, fail to provide accurate results especially in the high sensitivity operation mode, characterized by high overvoltage, in which dark noise peaks are more closely packed and superimpose severely.

In this paper, after a basic description of SiPMs and their typical output signal, we analyze more in depth the noise reduction advantages linked to the use of peak height monitoring over charge integration and describe the difficulties met by commonly used analysis procedures. We then describe an original, non-linear algorithm for noise filtering and SiPM current signal analysis, which overcomes the limitations of standard linear alternatives, and extracts current peak height and rise time accurately even in presence of high intensity, high frequency noise. Finally, we empirically compare the two approaches, showing a significantly better performance of the non-linear analysis proposed.

2 SiPM structure and working principle

2.1 Individual SPAD cells

The SiPM signal results from the superposition of the signals due to the individual SPAD cells in the array; thus, analysis of the SiPM signal rests on the operation principle of the individual SPADs.

2012 JINST 7 P08014

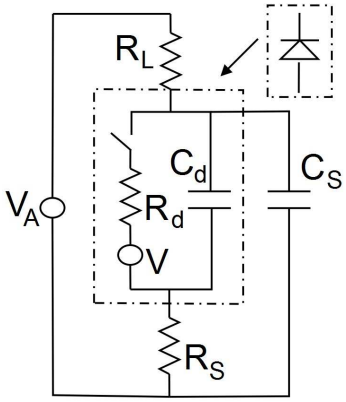

Figure 2. Equivalent circuit of a SPAD in the current-mode output configuration [23].

If individual SPAD cells are biased above the breakdown voltage V_b, the avalanches triggered are self sustaining, and need a suitable control circuitry in order not to persist indefinitely. The process of turning off the avalanche, readying the device for another detection, is known as *quenching* and can be achieved in two different ways: actively and passively.

Passive quenching is the most elementary way of quenching, and the one still used in SiPM. It makes use of a large resistance, of about 100 kΩ, in series to the SPAD: when the current rises due to an avalanche event, a large voltage drop develops across the resistance, which reduces the external voltage across the SPAD below V_b, quenching the avalanche. Passively quenched SPAD cells are well described and reduced to an equivalent electric circuit in the work of Cova [23]. When the avalanche is triggered and the voltage drop develops on the quenching resistor R_L, the voltage across the diode is kept above breakdown by the charge stored in two capacitances in parallel with the diode: the junction capacitance and the capacitance of the diode with respect to ground, with total effective capacitance C. As the capacitances discharge with a time constant R_dC, with R_d being the internal resistance of the diode, the voltage across the diode drops, together with the current, until the latter eventually falls below the *latching current*, at which level the probability of the avalanche turning off out of pure statistical effects is not negligible, and the avalanche is quenched.

The leading and restoring edges of the current peak are hence due to two different phenomena and present timescales which differ by several orders of magnitude: the rise time of the peak is driven by the velocity of the charge carriers avalanche multiplication process and the immediate effects of opening the switch in the equivalent circuit shown in figure 2. This results in a theoretical rise time of the order of tens of picoseconds, and is thus much faster than the restoring time, which is instead linked to the time constant R_dC, and is about 3 orders of magnitude longer.

From the discussion presented in [23], it is shown that the current passing through the diode shows a peak whose maximum value is given by:

$$I_p = \Delta V / R_d = (V_{\text{ext}} - V_b)/R_d \tag{2.1}$$

where ΔV, also known as "overvoltage", is the difference between the applied bias voltage and V_{ext} the breakdown voltage, and R_d ranges from a few hundred Ω for wide area devices with thick depletion layers up to several thousands Ω for small area devices with thin depletion layers, which

2012 JINST 7 P08014

is the case for most SiPMs. Equation (2.1) is thus relevant insofar as the current peak signal is monitored, as suggested in this paper. On the other hand, charge integration is usually preferred instead, and the relevant quantity becomes the total charge in the pulse Q_{pc}. Q_{pc} can be obtained by integrating the current signal or, from the equivalent circuit analysis, estimate as the total charge stored in the total capacitance in parallel with the diode for the given overvoltage:

$$Q_{\mathrm{pc}} = \Delta V \cdot C \tag{2.2}$$

2.2 Signal form model

Full analysis of the equivalent circuit shows that the current across the diode after an avalanche has been triggered is described by a sum of two exponentials with differing time constants:

$$I_d = I_0 \left(e^{(-\alpha t)} - e^{(-\beta t)} \right) \tag{2.3}$$

As mentioned above, the time constant of the rising exponential is of the order of tens of picoseconds, whilst the restoring edge time constant is up to 3 order of magnitude larger. Therefore, the fine structure of the peak rise time is inevitably lost in the ≈10 GHz bandwidth commonly available for the electronic read-out. However, the peak height is correctly sensed, as the decrease in current due to the falling exponential is usually negligible in the time scale of electronic signal acquisition. Therefore the SPAD signal, and hence, by superposition, the SiPM signal, can be approximated by a rising straight line and a falling exponential:

$$I(t) = \begin{cases} P_H \cdot \frac{t-t_0}{t_R} & t_0 \leq t \leq t_p \\ P_H \cdot exp\left(-\frac{t-t_p}{t_{\mathrm{rec}}} \cdot ln(R_c) \right) & t_p \leq t \leq t_p + t_{\mathrm{rec}} \end{cases} \tag{2.4}$$

where P_H is the peak height; t_0 the time when the avalanche is triggered; t_R is the rise time; t_{rec} the recovery time defined as the time which the current takes to drop below $1/R_c$ of its peak value and t_p is the time when the peak occurs, equal to $t_0 + t_R$. The time constants t_R and $ln(R_c)/t_{\mathrm{rec}}$ are found empirically and are related to the amplifier rise time and the product $R_d C$ of diode resistance and total capacitance respectively.

Figure 3 shows the superposition of the approximation described in equation (2.4) on a real SiPM dark noise signal. As expected from the analysis of the equivalent circuit shown in figure 2, the approximation follows the signal very closely, apart from random deviations due to noise. In this work, the signal form shown in equation (2.4) has thus been used to approximate the SPAD signal.

2.3 Read-out techniques

It is established since early literature in the subject [18, 23], that there exist 3 main ways to read the output signal of a SiPM following an avalanche:

1. Measuring the total charge displaced by the SPADs.
2. Measuring the voltage drop across the SPADs.
3. Measuring the current signal across the SPADs.

2012 JINST 7 P08014

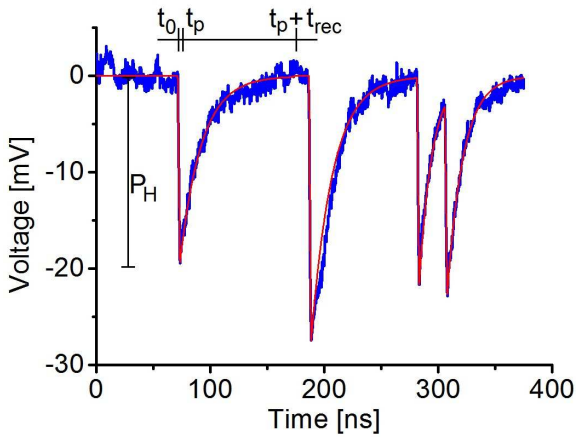

Figure 3. Superposition of the approximated signal form described in the text on a real SiPM dark noise signal. The approximation is shown in red, as a smooth solid line; the real signal is instead shown in blue, and is recognizable from the random noise oscillations.

Measuring the total charge displaced by the SPADs through a time gated charge to digital converter is the most common method of analysis, and is based on equation (2.2). The advantage of this method lies in it being less sensitive than the others to higher frequency electrical noise, as the gate time for charge integration needs be in the order of tens of nanoseconds to achieve acceptable signal to noise ratios (SNR), thus providing an in-built time averaging low pass noise filter. However, it does present the shortcoming that any charge fluctuations in the gate time due to phenomena other than the signal to be measured are erroneously added up. In particular, the timescale of tens of ns is compatible with large afterpulsing contributions in commonly used SiPM devices [22], marking this readout method as particularly susceptible to afterpulsing noise errors.

From the point of view of the applications, problems arise both when the SiPM are operated in presence of medium to intense light pulses and when less intense light or dark noise alone is measured. Indeed, in the first case, intense light pulses will cause several cells to fire at once, linearly increasing the probability of registering afterpulsing effects within the gate time. On the other hand the gate time also becomes very relevant when dark noise is measured: in this case, such error effectively prevents the technique from being applied in the high overvoltage regime, when the probability of a second dark count happening within the gate time becomes not negligible (dark count rate $>$ 10 MHz).

Measuring the voltage drop across the SPAD is usually not chosen as a read-out method, as it shares all the pros and contras of measuring the current signal, but the equivalent circuit analysis reported in [23] shows it to be intrinsically slower and hence less precise.

Monitoring of the current signal across the SPAD, based on equation (2.1), is instead also used, although more seldom, due to it being more sensitive to electric noise and requiring more involved data processing if signals overlap: therefore, it only becomes interesting as an easier to implement alternative to charge integration when large SNR is present, e.g. in the case of medium to high brightness light pulses, and avalanche events are well spaced. However, not requiring a gate time for integration, the method is intrinsically faster than charge integration and thus couples significant advantages on afterpulsing noise reduction with better performance in presence of high frequency dark counts such as those registered in the high overvoltage regime.

2012 JINST 7 P08014

Application of current signal monitoring to lower SNR and high event rate is withheld by the need of developing a suitable data processing algorithm allowing filtering of electrical noise, fast response, and high precision. The following sections present a novel such algorithm, allowing extension of current signal monitoring to low SNR and high event rate experiments.

3 Algorithm description

3.1 Challenges posed by the SiPM signal

To effectively use the peak height readout option, an algorithm optimized for high precision in measuring peak heights and rise times is needed. The algorithm described in this paper is designed to analyze a continuous signal from a SiPM, composed of several different avalanche events occurring at random times. This signal will be referred to in the following as *trace*. The difficulties in achieving the goals listed above with existing algorithms lie in the following issues:

- *Changing baseline of the signal*: for high overvoltages, when the dark count rate increases substantially, peaks are often superimposed on each other with a time delay, so that the algorithm needs to identify and measure peaks starting at different heights.
- *Electrical noise*: when mainly first order peaks are measured, the signal to noise ratio can be as low as 5 in the worst cases (low overvoltages and high electronic noise), giving an uncertainty of 20% on the peak height if it is measured directly, leading to the need of a filtering and fitting algorithm.
- *Very different peak heights and widths*: these features change drastically with overvoltage and SiPM model, hence the code has to be able to adapt itself to measuring different features without the user having to vary and optimize the parameters for every measurement.
- *Random event distribution*: which presents problems if a filtering fully based on the frequency domain is used.

To identify the peaks starting and end points, a standard first derivative threshold method such as the one described in section 3.2 can be used, insofar as a suitable filtering algorithm is used beforehand to decrease the signal to noise ratio to a level low enough for no peaks to be either lost or mistakenly counted. However, ordinary linear filters and analysis codes, based either on the time or frequency domain, have serious difficulties in meeting the requirements listed above, therefore the algorithm we developed uses instead a non-linear approach.

3.2 Linear fitting algorithm

The non-linear algorithm presented in this paper and described in detail in the next section was compared with a standard linear filtering and peak finding algorithm based on a moving average routine, which is the optimal solution for linear filtering of time encoded, discrete events signals, such as SiPM peaks [24].

This linear code first filters the trace by calculating for each point the average value of N points around it, and assigning to it this value: N is thus known as the averaging window width. This filtering ensures, if it is done with a large enough value of N, that noise spikes are flattened and not

2012 JINST 7 P08014

mistaken for peaks. Then, the derivative of the filtered trace is calculated by the ratio between the vertical and horizontal distance of two points N_d points apart in the filtered trace, where N_d is the derivative window width: the derivative trace shows at this point spikes in correspondence of every peak. Finally, any derivative spike that overcomes a certain threshold D_{th} identifies a peak. The stationary points of this filtered trace for each identified peaks are taken as its starting and end points.

In this work, in order to choose optimized values for the parameters N, N_d and D_{th}, we generated artificial SiPM signals following eq. (2.4), with white noise being simulated with a 20 GHz bandwidth. The artificial signal homogeneously spanned through the range of event frequencies and SNR commonly met in the regions of interests of SiPM: the SNR, defined as the ratio between the rms amplitudes of signal and noise, was varied in the interval $2 \div 20$ and the frequency, expressed as the expected number of events happening in one detector recovery time, varied in the interval $3 \div 0.1$. The values of N, N_d and D_{th} have been varied and the optimum set which optimizes peak detection efficiency (i.e. the number of peaks detected / total number of peaks) has been chosen for use in the optimized linear filtering algorithm.

An example of the results of the application of such procedure to an actual experimental trace are shown in figure 11.

3.3 Non-linear fitting algorithm

The non-linear algorithm reported in this work is based on a piecewise linear fitting approach [25–27], i.e. on the idea of approximating the investigated curve with straight lines of different lengths and gradients and optimized endpoint locations, depending on the features of the analyzed signal. The algorithm leading to the choice of each segment endpoints, referred to as *breakpoints*, is what differentiates different piecewise linear fitting approaches. To do this, our program uses 3 peculiar features:

1. Preliminary analysis for problem scaling
2. Tapered limit condition for insertion of breakpoints
3. Back-tracing optimization

The preliminary analysis allows the program to tune the parameters needed, listed later in this section, to the particular trace which is analyzed. It is performed only on a smaller sample of the trace, and is carried out before the actual trace is analyzed. This step calculates an estimate for signal qualities such as SNR and expected peak height, hence fulfilling the requirements detailed at the point "*very different peak heights and widths*" in section 3.1.

In this phase, the code makes use of an efficiency optimized, existing time-domain filtering algorithm (based on a moving average routine with large averaging width compared to the expected peak rise time) to obtain a first estimate of the peak characteristics (estimating the average peak height and rise time with a $< 30\%$ error), and to set the needed thresholds, which we will refer to as *preliminary parameters*.

These consist in the average peak height and rise time, h_μ and $t_{r-\mu}$, and in the standard deviation of their distributions, h_σ and $t_{r-\sigma}$. From these numbers, it is possible for the program to have a scale of the peak heights, and also an estimate of the average value of derivative to be expected

2012 JINST 7 P08014

in presence of a peak (from the ratio between peak height and rise time). A third preliminary parameter ε_{lim} is taken from the maximum deviation of the oscilloscope trace from the filtered trace, and represents an estimate of the maximum noise amplitude. Finally, σ_{lim} is computed as the average standard deviation of the experimental points around the filtered line in the regions of higher derivative values (corresponding to the peaks). This is needed as the noise contribution to the signal is usually much lower in the regions of the peaks as compared to the flat regions in between them, due to the rise time of the avalanche event being much smaller than the amplifier rise time, thus saturating its frequency domain response in those regions.

Preliminary parameters are only computed for a smaller part of the full trace, through a fully automatic routine not requiring any input from the user, and take a computing time per analyzed point slightly longer (on average about 35% more) than the computing time taken by the analysis of the trace itself. The influence of the preliminary analysis on the overall computing time is thus chiefly regulated by the percentage of the trace which is included in it, and can thus be made negligible if the experimental conditions do not vary significantly during the experiment and only a small portion of the trace can be used in the preliminary analysis.

For applications where small signals are expected, resulting in only a few cells firing, the preliminary analysis can be performed on a rather short time scale: it is only required that a few dark noise peaks fall into the part of the trace included in the preliminary analysis. For applications where large signals are expected, it is still possible to perform the preliminary analysis on dark noise peaks alone: events resulting in taller peaks will still be discerned. A simple amplitude threshold can then be used to discriminate dark noise peaks from the sought signal. As an alternative, this threshold procedure can be directly built into the code by requiring that at least one large signal peak is included in the preliminary analysis; this will result in dark noise peaks being classed as noise and ignored by the fitting routine.

Once the preliminary parameters have been computed, the code analyses the trace from the beginning and draws a best fit line through the first two points. It then iteratively includes the following point in the fit, and accordingly updates the two parameters of the best fit line together with the new point deviation from the best fit ε and the overall goodness of fit σ (measured as the standard deviation of the data distribution around the best fit). A quality test is then carried out to see if the new point assigned to the line is likely to actually belong to it. This test can be failed by failing any of the following two conditions:

1. ε exceeds the preliminary parameter ε_{lim} measuring the average noise amplitude.

2. σ exceeds the parameter σ'_{lim} measuring the expected goodness of fit.

Point 1 improves the definition of the peak start point, avoiding that the initial points of the peaks are mistakenly accounted for noise; moreover, the peak start point location is sharpened in a later stage of analysis described in the following.

Point 2 serves instead to obtain precise measurement of the peak top point, and is where the second feature of the non-linear algorithm, i.e. the tapered limit condition for insertion of breakpoints, plays its role. Setting $\sigma'_{\text{lim}} = \sigma_{\text{lim}}$ as calculated in the preliminary analysis would be too stringent a condition in flat regions between peaks, resulting in too many tests to be failed. On the other hand, as discussed in section 2.1, peaks develop on a very short timescale, only limited in

2012 JINST 7 P08014

practical applications by the amplifier risetime, hence white noise oscillations are not critical on the leading edge of a peak, and the error condition on $\sigma'_{\rm lim}$ can be more stringent. Therefore, $\sigma'_{\rm lim}$ is a limit condition tapered from $\varepsilon_{\rm lim}$ to $\sigma_{\rm lim}$ as a function of the vertical length of the line under analysis.

In particular, $\sigma'_{\rm lim} = \varepsilon_{\rm lim}$ for null vertical distance of the first point assigned to the line to the last one assigned; $\sigma'_{\rm lim}$ increases then linearly until $\sigma'_{\rm lim} = \sigma_{\rm lim}$ when the vertical distance equals the average peak height measured in the preliminary analysis. This solution guarantees very stringent breakpoint condition in the final region of the peaks, when additional points are likely to be erroneously included in the peak as the derivative only gradually flattens, and results thus in very sharp definition of the peaks.

When any of the above two tests is failed, the new point is not included in the fit and a breakpoint is inserted. This is the point when two different best fit lines merge: i.e. the starting point of a new best fit line. When this happens, the code performs the back-tracing optimization mentioned above, in order to obtain a sharper definition of peak start points. The line just ended is traced back again with the same procedure used to fit forward: the point just before the first which was assigned to the current line (and belongs, therefore, to the previous line) is considered for inclusion in the current line by being tested against the conditions of $\varepsilon < \varepsilon_{\rm lim}$ and $\sigma < \sigma'_{\rm lim}$. If the tests are passed, the best fit error σ calculated for this back-traced line is then compared with the best fit error σ of the previous line and the configuration which minimizes the error is chosen. If this results in the starting point of the current line being changed, the same is repeated for another point backwards, until the tests for inclusion are failed and the program can move on to the next line. The addition of this backward tracing process to the algorithm significantly increases the precision in localizing the breakpoints. The presented algorithm as a whole produces then results well compatible with the much longer procedure of finding the optimum breakpoint positions in the whole trace by making a standard piecewise linear fitting, e.g. by trying all possible breakpoints, calculating the goodness of fit σ and then choosing the breakpoint position which minimizes σ.

Once this filtering procedure has been carried out, peaks are identified with a standard derivative threshold procedure, analogue to the one described for the linear code used for comparison, with the difference that the derivative value is not taken by difference of points across a derivative window width, but is simply the gradient of each segment found.

An example of the results of the application of such procedure to an actual experimental trace are shown in figure 10.

4 Performance measurement

A full characterization of linear digital filters, useful for comparing different solutions, can be usually given by quoting their kernel: i.e. the response to an ideal Dirac's Delta pulse. However, when this procedure is applied to the non linear filter discussed in this paper, the filters follows the signal perfectly, and the kernel turns out to be ideal. Therefore, to provide a quantitative statement on the increase in peak determination performance obtained when using the non-linear filter against the optimal linear technique, both techniques have been applied to a set of test traces with different characteristics, and the results compared.

2012 JINST 7 P08014

4.1 Characteristics of the test signal

The most challenging scenario for a filter intended to identify peaks from a SiPM signal is the detection of randomly occurring single cell peaks, as random time distribution implies that no external trigger can be used, and single cell peaks are the smallest signals produced by a SiPM, and are hence associated with the smallest SNR. A suitable such measurement is hence the measurement of dark noise alone, for which the event time distribution is random, and mostly single cell peaks appear, due to dark noise being dominated by a Poisson statistics [28].

The test signal used for algorithm comparison consists thus of a continuous trace filled with randomly placed, unit height peaks ($P_H = 1$), as described by equation (2.4). When analyzing such trace to identify and characterize its peaks, two parameters of the signal have been tested and identified to have the most influence: the SNR and the peak frequency. As mentioned in section 3.2, the peak frequency is defined as the average number of peaks occurring in one recovery time t_{rec}, as this gives a direct measure of how many peaks are likely to superimpose.

The remaining peak characteristics are set to resemble common experimental values (see e.g. [29–31]. The peak rise time t_R is thus set to 1 ns, R_c is set to e to simplify equation (2.4), and the recovery time constant t_{rec} is set to 30 ns.

The performance of the algorithms is measured by determining the distributions of the errors in the determination of peak height and width. The results of the algorithms are compared with the true values, known for each generated trace, and an histogram of the difference between true and measured peak height and width for a sample of 100k peaks for any given SNR and frequency configuration is plotted. This procedure results in the establishment of a bell shaped curve, described by a mean and standard deviation.

The percent difference between the true value and the mean of each distribution gives an indication of the systematic error introduced by the algorithm, such as e.g. the decrease in peak height due to smoothing of the peaks for the linear algorithm. The standard deviation of the distribution is instead linked to the algorithm precision, and is due to statistical error, which increases with decreasing SNR, and expresses the uncertainty in the measurement. It is also expressed as a percent of the true value.

To complement these two observables, the probabilities of measuring a non-existent peak (artifact counts) or missing an existent one (undetected peaks) are also measured as respectively the number of peaks mistakenly detected and the number of non detected peaks over the total number of true peaks present in the sample, i.e. 100k. These probabilities provide information on the reliability of the algorithm at the given SNR and frequency regimes.

4.2 Results

Figure 4 shows the mean percentage error on peak heights determination of both the linear and non-linear algorithms as a function of SNR (x axis) and frequency (curves parameter).

The expected peak height underestimation of the linear algorithm, due to smoothing, is reflected in the negative values of the corresponding curves in the figure. On the other hand, the non linear algorithm is instead characterized by a systematic peak height overestimation, reflected in the positive values of the corresponding curves. This is due to noise being included in the peak line until the error control is flagged: as it is confirmed by the effect getting less severe for smaller

2012 JINST 7 P08014

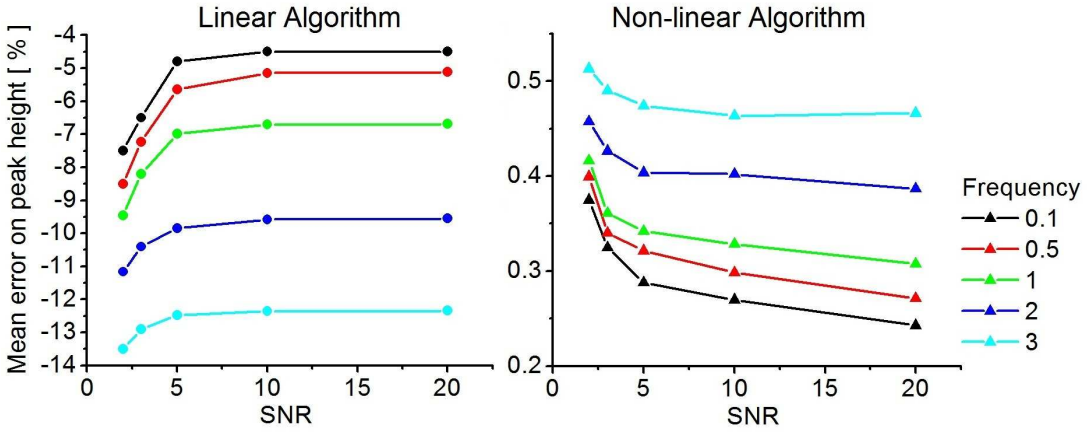

Figure 4. Mean percentage error on peak heights determination of both the linear and non-linear algorithms as a function of SNR (x axis) and frequency (curve parameter). Frequency defined as the average number of peaks occurring in one recovery time t_{rec}.

SNR. However, in the range of SNR and frequency spanned, the absolute value of the non-linear algorithm overestimation, $\approx 0.2 \div 0.5\%$, is smaller than the absolute value of the linear algorithm underestimation by a factor of ≈ 25, confirming the better performance of the non-linear algorithm.

The response to variation in SNR and frequency is similar in both cases: increasing frequency or decreasing SNR leads to larger errors. In particular, varying the frequency by a factor of 3, between 3 and 1, leads to the peak height errors decreasing by a factor of about 40% in the case of the linear code and 30% in the case of the non-linear code, when averaged over all the SNR values.

Similarly, decreasing SNR leads to increasing errors, however only moderate differences, within a factor of 2, correspond to variations in SNR between 2 and 20. Whilst the trend in figure 4 seems to suggest a sharp increase in mean error for lower values of SNR, it should be noted that in practical applications it is unlikely to have a SNR$<$2, hence this part of the curve has no consequences in practical applications.

Figure 5 shows instead the spread of the peak heights around the mean values, i.e. the precision of the corresponding algorithm.

The non-linear algorithm performs sensibly better than the linear alternative, especially for SNR larger than 5, relevant for most applications: with an improvement in precision of about 1 order of magnitude. However, the non linear algorithm is more affected by increased frequency, and the precision improvement factor moves from an average over SNR of 12 to an average of only 8 when moving from frequency 1 to 3.

Figure 6 and 7 illustrate the same analysis shown above applied to the determination of the peak rise time.

With reference to figure 6, differently from the case of peak height determination, the mean error contribution to the linear algorithm is strongly dominated by the averaging window chosen for the moving average routine, of 25 ns for the data presented. Accordingly, the non linear algorithm features a much greater accuracy, achieving sub ns values for SNR larger than 5. In both cases, the dependence on frequency is not critical, and only becomes relevant in the case of the

2012 JINST 7 P08014

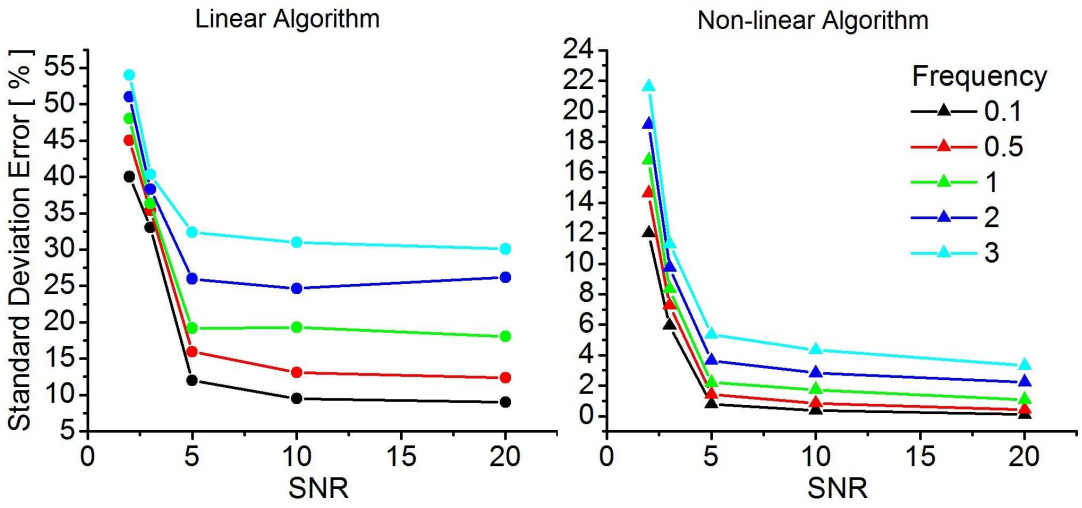

Figure 5. Standard deviation of the peak height distribution of both the linear and non-linear algorithms as a function of SNR (x axis) and frequency (curve parameter). Frequency defined as the average number of peaks occurring in one recovery time t_{rec}.

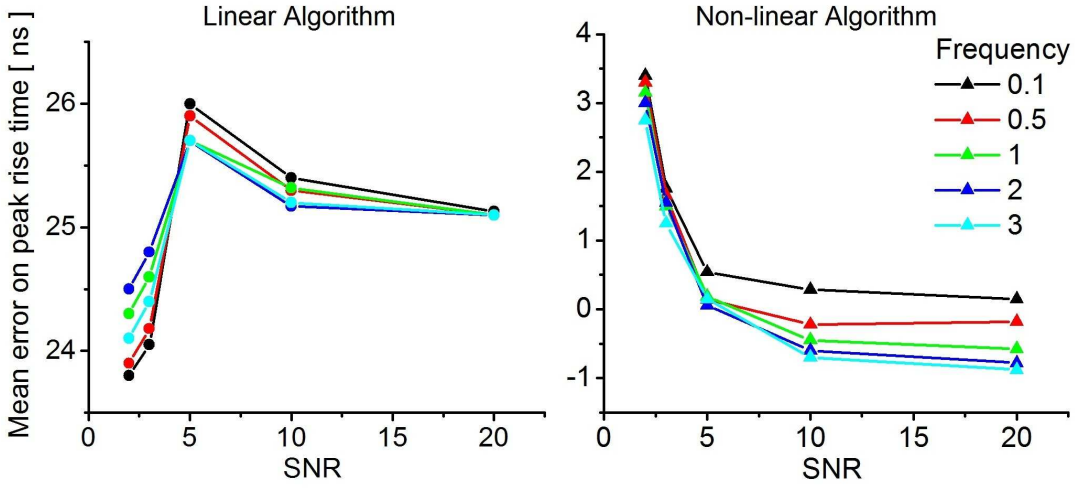

Figure 6. Mean percentage error on peak rise time determination of both the linear and non-linear algorithms as a function of SNR (x axis) and frequency (curve parameter). Frequency defined as the average number of peaks occurring in one recovery time t_{rec}.

non-linear algorithm at high SNR. In these conditions, increasing frequency corresponds to lower and lower values of the peak rise time, leading eventually to underestimation of it. This is due to the increasing superpositions of signals for higher frequencies, resulting in the leading edge of the rise being shadowed by the previous peak.

Figure 7 shows instead that, for high SNR, the precision of the non-linear algorithm is in fact worse than the linear alternative. This is again due to the averaging window dominating the process in the linear case: for high SNR this effect is even more pronounced, and hence the measured peak width varies only slightly from the value of 25 ns, leading to better precision.

2012 JINST 7 P08014

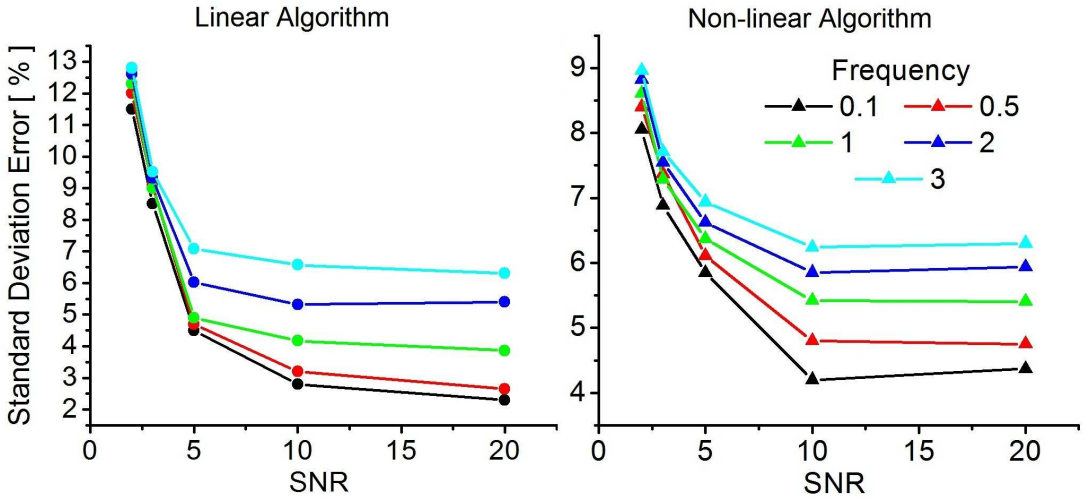

Figure 7. Standard deviation of the peak rise time distribution of both the linear and non-linear algorithms as a function of SNR (x axis) and frequency (curve parameter). Frequency defined as the average number of peaks occurring in one recovery time t_{rec}.

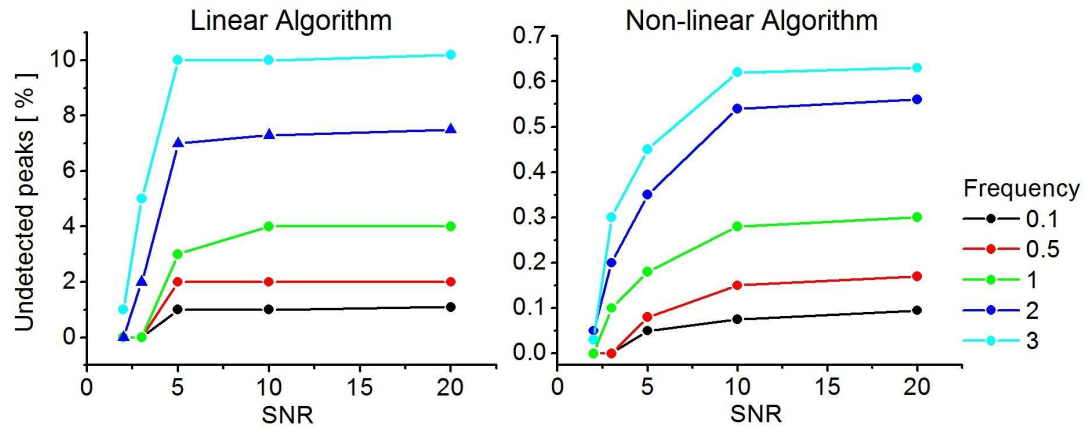

Figure 8. Percentage of undetected peaks for both the linear and non-linear algorithms as a function of SNR (x axis) and frequency (curve parameter). Frequency defined as the average number of peaks occurring in one recovery time t_{rec}.

Finally, the information on peak height and rise time errors is complemented by the data reported in figure 8 and 9, showing respectively the percentage of undetected and artifact peaks (where noise spikes have been wrongly classified as peaks), and hence characterize the reliability of each algorithm. As expected, both algorithms improve their performance dramatically depending on the level of noise, for low SNR, whilst leveling off at higher values to a plateau representing the maximum reliability of each algorithm. The results concerning the frequency dependence are also immediately understandable, as higher frequencies correspond to more undetected, and less artifact peaks. However, the impact of frequency is only significant for the undetected peaks, whilst it has a smaller impact, of about 50% the average value, for artifact peak detection at high SNR.

2012 JINST 7 P08014

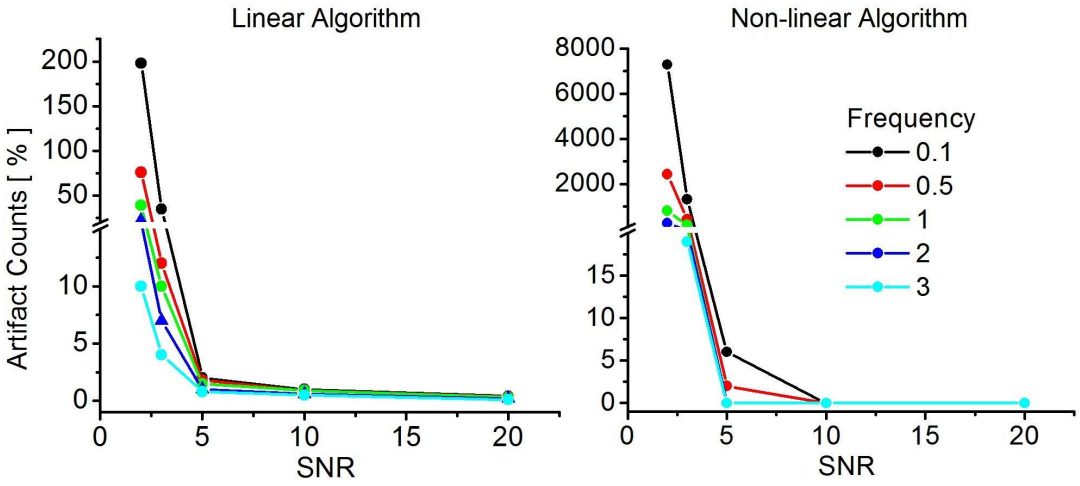

Figure 9. Percentage of artifact peaks for both the linear and non-linear algorithms as a function of SNR (x axis) and frequency (curve parameter). Frequency defined as the average number of peaks occurring in one recovery time t_{rec}.

As it was the case for the analysis of the precision and accuracy, the qualitative behaviors of the two algorithms are remarkably similar, only differing in minor details, such as e.g. the more gradual dependence on SNR of the non-linear algorithm and, similarly to the previous analysis, a slightly reduced dependence on frequency. However the absolute values show a considerable performance increase for the non-linear algorithm in excess of 1 order of magnitude.

4.3 Comparison with experimental data

Figure 10 and 11 present an example of both the standard linear procedure and the proposed non-linear procedure applied to an actual experimental trace obtained from a SiPM dark noise signal in the high overvoltage regime. It is seen how the non-linear analysis results in a series of points along the curve, representing the breakpoints of the piecewise linear fit. Drawing a straight line between each pair of points would result in the curve as fitted by the code. The linear analysis results instead in a continuous, smooth curve. The results of the post-processing, which identifies the peak start and end points are also shown in each graph, in the form of peak start and end points. It is easily seen from the peak height points how the linear code consistently underestimates the peak height, especially for very sharp peaks. It is noted that the parameters of the linear analysis shown in figure 11, and in particular the averaging window width, have been optimized for the particular experimental signal shown, and would therefore perform very poorly in different conditions, e.g. in low overvoltage regime. Nevertheless, it is still possible to see how the linear analysis consistently overestimates the peak rise time as well, as the start points always precedes the real peak onset and the end point always follows in time the point corresponding to maximum peak height.

Finally, inspection of the tall, second order peaks at time approximately 750 and 1100 ns, shows how the linear analysis, differently from the non-linear algorithm, fails to identify these peaks as coming from the superposition of two different, closely spaced signals and counts them instead as higher single peaks. That this is not the case is visible from the sharp gradient sign

2012 JINST 7 P08014

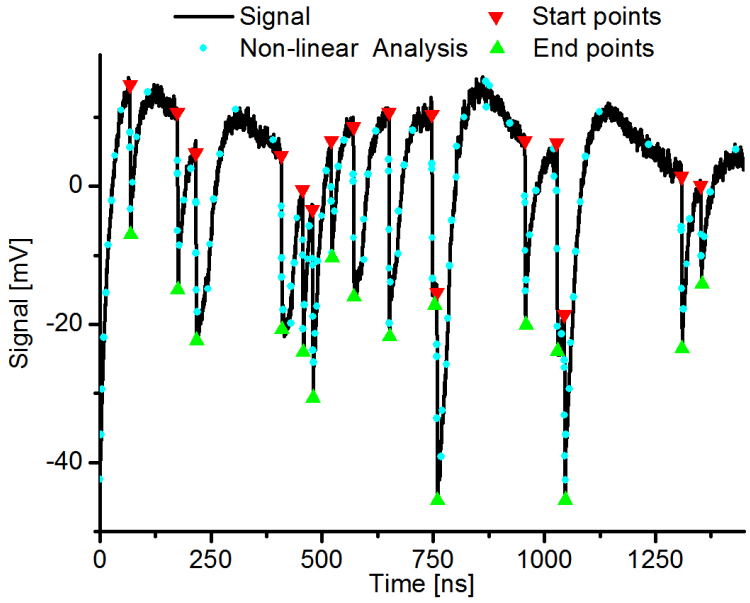

Figure 10. Filtering and fitting analysis of an experimental waveform carried out with the non linear algorithm presented. The waveform has been acquired from a SiPM in the high overvoltage regime without any impinging light signal (dark noise only). Start and end points of each peak shown as calculated by the post-processing, threshold based process described in section 3.2.

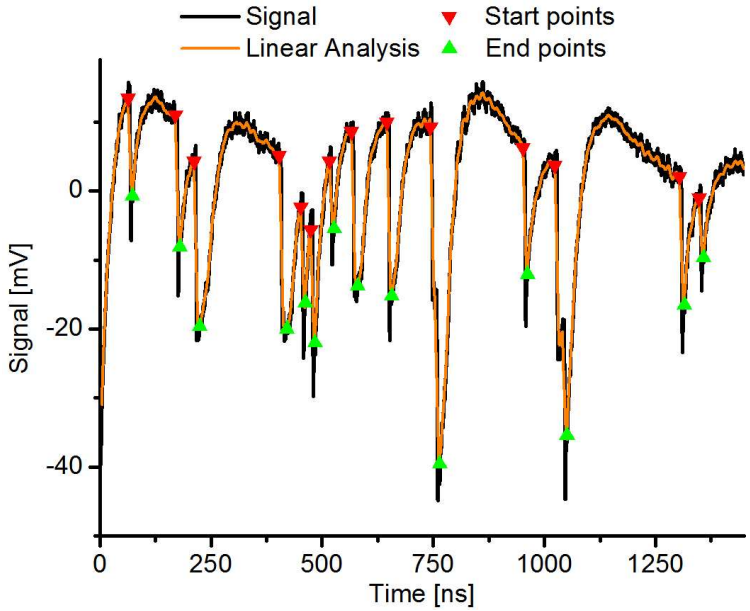

Figure 11. Same waveform shown in figure 10, but analyzed with the linear algorithm.

inversion happening at about half height in both cases, and is clearly identified by the non linear algorithm.

It should be also added that it would be erroneous to consider these peaks as second order peaks, which would indicate quasi-simultaneous firing of more cells together due to, e.g. cross talk

2012 JINST 7 P08014

or a light signal. Indeed, the distance between them can be measured on an oscilloscope to be in the order of few ns: about 2 orders of magnitude longer than required for the avalanche forming process, and even longer in fact than the amplifier rise time. Therefore, the peaks are instead best interpreted as randomly occurring closely spaced peak as correctly analysed with the non-linear algorithm.

Such improvements in performance prove can prove crucial, especially in applications where dark noise alone or small amplitude signals are analyzed. An example of the use of this algorithm can be found in [32]. In this paper, the improvement in performance allows a more precise dark noise analysis also for high overvoltages, and leads to the possibility of explaining the Gaussian spread observed in SiPM peak height spectra and assessing the manufacturing quality of SiPM.

5 Conclusions

In this paper, we have presented a novel non-linear algorithm optimized for current peak detection and measurement in SiPM signals. The algorithm is based on a custom piecewise linear fitting approach, and allows noise filtering in high noise signals typical of low light levels and dark noise measurements.

Empirical comparison of this algorithm with an optimal linear algorithm, based on a moving average routine, shows a considerable improvement in performance. In particular, when it comes to peak height determination, the non-linear alternative improves accuracy by a factor of 25 and precision by a factor of $8 \div 12$. Similar results are obtained for determination of peak widths, where the non-linear algorithm improves accuracy by a factor of $50 \div 100$, even though it features a precision comparable, and in some cases worse by a factor of up to 2 when compared to the linear algorithm. However, the advantage of higher precision for the linear algorithm is shadowed by the dramatic poorer performance in terms of accuracy.

Finally, the detection efficiency was assessed in terms of probability of undetected and artifact peaks. The tests show that also in this case the non-linear algorithm performs significantly better, demonstrating improvement factors in excess of 1 order of magnitude for both undetected and artifact peaks.

These improvements in filtering and peak detection algorithm performance pave the way for a more efficient read-out technique for SiPM, based on the analysis of the current signal, rather than the charge. Such read-out technique, as described in section 2.3, is intrinsically faster than charge integration, and can be used to reduce the effects of after-pulsing noise, for which to date there is no significant measure of control at the manufacturing stage, such as in the case of cross-talk noise.

Acknowledgments

Work supported by the EU under contract PITN-GA-2008-215080 and the STFC Cockcroft Institute Core Grant No. ST/G008248/1.

References

[1] B. Dravins, D. Faria and D. Nilsson, *Avalanche diodes as photon-counting detectors in astronomical photometry*, *Optical and IR Telescope Instrumentation and Detectors*, *Proc. SPIE* **4008** (2000) 298.

2012 JINST 7 P08014

[2] N. Nightingale, *A new silicon avalanche photodiode counting detector module for astronomy*, *Exper. Astron.* **1** (1991) 407.

[3] C. Zappa, F. Ripamonti, G. Lacaita, A. Cova and S. Samori, *Tracking capabilities of SPADs for laser ranging*, in proceedings of 8th *International Workshop on Laser Ranging Instrumentation*, J.J. Degnan ed., NASA, Greenbelt U.S.A. *NASA Conf. Publ.* **3214** (1992) 25.

[4] G.-C. Wang and J. Gordon, *Solid state photomultipliers for biomedical imaging applications*, *Proc. SPIE* **7182** (2009) 718218.

[5] N. Otte et al., *New results from a test of Silicon Photomultiplier as readout for PET*, *IEEE Nucl. Sci. Symp. Conf. Rec.* **6** (2004) 3738.

[6] Y. Musienko et al., *SSPM readout of LSO, (Lu-Y)AP:Ce and PWO-II pixels for PET detector modules*, *IEEE Trans. Nucl. Sci.* **55** (2008) 1352.

[7] D.J. Herbert et al., *The Silicon Photomultiplier for application to high-resolution positron emission tomography*, in proceedings of the 7th *International Conference on Position-Sensitive Detectors*, *Nucl. Instrum. Meth.* **A 573** (2007) 84.

[8] S. Ripamonti, G. Ghioni and M. Vanoli, *Photon timing OTDR: a multiphoton backscattered pulse approach*, *Electron. Lett.* **26** (1990) 1569.

[9] A. Intermite, M. Putignano and C.P. Welsch, *Feasibility study of an optical fibre sensor for beam loss detection based on a SPAD array*, in proceedings of *DIPAC*09, Basel Switzerland (2009), pg. 228.

[10] V. Kovalchuk, G. Lolos, Z. Papandreou and K. Wolbaum, *Comparison of a Silicon Photomultiplier to a traditional vacuum photomultiplier*, *Nucl. Instrum. Meth.* **A 538** (2005) 408.

[11] F. Zappa, S. Tisa, A. Tosi and S. Cova, *Principles and features of single-photon avalanche diode arrays*, *Sens. Actuators* **A 140** (2007) 103.

[12] J.R. Lakowicz, *Topics in fluorescence spectroscopy. Volume* 1*: techniques*, Springer, Germany (1992), pg. 1.

[13] A. Muller, N. Gisin and J.-P. Pellaux, *Decay time measurements of fluorescente fibers with photon counting*, *Rev. Sci. Instrum.* **64** (1993) 1158.

[14] J.C. Jackson, D. Phelan, A.P. Morrison, R.M. Redfern and A. Mathewson, *Characterization of Geiger mode avalanche photodiodes for fluorescence decay measurements*, *Proc. SPIE* **4650** (2002) 55.

[15] L.-Q. Li and and L.M. Davis, *Single photon avalanche diode for single molecule detection*, *Rev. Sci. Instrum.* **64** (2003) 1524.

[16] A. Del Guerra et al., *Silicon Photomultipliers (SiPM) as novel photodetectors for PET*, *Nucl. Instrum. Meth.* **A 648** (2011) S232.

[17] A. Campisi at al., *Multipixel Geiger-mode photon detectors for ultra-weak light sources*, *Nucl. Instrum. Meth.* **A 571** (2007) 350.

[18] P. Finocchiaro at al., *Features of Silicon Photomultipliers: precision measurements of noise, cross-talk, afterpulsing*, *IEEE Trans. Nucl. Sci.* **56** (2009) 1033.

[19] A. Balling et al., *Machine-mode aware beam loss monitoring*, in proceedings of *PCaPAC*08, Ljubljana Slovenia (2008).

[20] R.J. McIntyre, *Multiplication noise in uniform avalanche diodes*, *IEEE Trans. Electron Dev.* **13** (1966) 164.

2012 JINST 7 P08014

[21] W.G. Oldham, R.R. Samuelson and P. Antognetti, *Triggering phenomena in avalanche diodes*, *IEEE Trans. Electron Dev.* **19** (1972) 1056.

[22] P. Finocchiaro et al., *Characterization of a novel* 100*-channel Silicon Photomultiplier — part I: noise*, *IEEE Trans. Nucl. Sci.* **55** (2008) 2757.

[23] S. Cova, M. Ghioni, A. Lacaita, C. Samori and T. Zappa, *Avalanche photodiodes and quenching circuit for single-photon detection*, *Appl. Opt.* **35** (1996) 1956.

[24] S.W. Smith, *The scientist and engineer's guide to Digital Signal Processing*, California Technical Pub., U.S.A. (1997).

[25] G. Ferrari-Trecate and M. Muselli, *A new learning method for piecewise linear regression*, in proceedings of *Artificial Neural Networks — ICANN* 2002, *International Conference*, Madrid Spain August 28–30 2002.

[26] A. Magnani and S.P. Boyd, *Convex piecewise-linear fitting*, *Optim. Eng.* **10** (2009) 1.

[27] G.F. Malash and M.I. El-Khaiary, *Piecewise linear regression: a statistical method for the analysis of experimental adsorption data by the intraparticle-diffusion models*, *Chem. Eng. J.* **163** (2010) 256.

[28] I.F. Castro, A.J. Soares and J.F. Veloso, *Impact of dark counts in low-light level Silicon Photomultiplier multi-readout applications*, *IEEE Nucl. Sci. Symp. Conf. Rec.* (2009) 1592.

[29] A. Intermite, M. Putignano and C.P. Welsch, *Influence of dark count on the performance of Silicon Photomultipliers*, in proceedings of *BIW*10, Santa Fe New Mexico U.S.A. (2010).

[30] D. Renker, *New trends on photodetectors*, *Nucl. Instrum. Meth.* **A 571** (2007) 1.

[31] H.T. van Dam et al., *A comprehensive model of the response of Silicon Photomultipliers*, *IEEE Trans. Nucl. Sci.* **57** (2010) 2254.

[32] A. Intermite et al., *Study of SiPM manufacturing quality based on dark noise analysis for varying bias voltages*, *JINST* in preparation.

2012 JINST 7 P08014