EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH
CERN — BEAMS DEPARTMENT

CERN-BE-2014-005 BI

Near-Saturation Single-Photon Avalanche Diode Afterpulse and Sensitivity Correction Scheme for the LHC Longitudinal Density Monitor

E. Bravin; S. Mazzoni; M. Palm

CERN – Geneva/CH

Abstract

Single-Photon Avalanche Diodes (SPADs) monitor the longitudinal density of the LHC beams by measuring the temporal distribution of synchrotron radiation. The relative population of nominally empty RF-buckets (satellites or ghosts) with respect to filled bunches is a key figure for the luminosity calibration of the LHC experiments. Since afterpulsing from a main bunch avalanche can be as high as, or higher than, the signal from satellites or ghosts, an accurate correction algorithm is needed. Furthermore, to reduce the integration time, the amount of light sent to the SPAD is enough so that pile-up effects and afterpulsing cannot be neglected. The SPAD sensitivity has also been found to vary at the end of the active quenching phase. We present a method to characterize and correct for SPAD deadtime, afterpulsing and sensitivity variation near saturation, together with laboratory benchmarking.

Presented at the IBIC 2014 Conference – Monterey-CA/USA from 14 to 18 September 2014

Geneva, Switzerland
November, 2014

Near-Saturation Single-Photon Avalanche Diode Afterpulse and Sensitivity Correction Scheme for the LHC Longitudinal Density Monitor

M. Palm*, E. Bravin, S. Mazzoni, CERN, Geneva, Switzerland

Abstract

Single-Photon Avalanche Diodes (SPADs) monitor the longitudinal density of the LHC beams by measuring the temporal distribution of synchrotron radiation. The relative population of nominally empty RF-buckets (satellites or ghosts) with respect to filled bunches is a key figure for the luminosity calibration of the LHC experiments. Since afterpulsing from a main bunch avalanche can be as high as, or higher than, the signal from satellites or ghosts, an accurate correction algorithm is needed. Furthermore, to reduce the integration time, the amount of light sent to the SPAD is enough so that pile-up effects and afterpulsing cannot be neglected. The SPAD sensitivity has also been found to vary at the end of the active quenching phase. We present a method to characterize and correct for SPAD deadtime, afterpulsing and sensitivity variation near saturation, together with laboratory benchmarking.

INTRODUCTION

The LHC RF cavities operate at about 400 MHz (2.5 ns RF buckets), with a distance of at least 10 buckets between nominally filled buckets (main bunches). A fraction of the beam can be found in the nominally empty buckets, where they are called satellite or ghost bunches [1]. The number of particles in satellites and ghost bunches is typically less than four or five orders of magnitude lower than the main bunches. However, they can create background noise at the interaction points and, due to their large number, cause problems for luminosity calibration [1].

The purpose of the LDM is to measure the relative number of particles in the different buckets, via the synchrotron radiation emitted by the beam.

A schematic overview of the LHC LDM system (one per beam) that was installed during Run 1 is shown schematically in Fig. 1: an SPAD registers individual synchrotron radiation photons emitted from an undulator (below 2 TeV) or separation dipole (above 2 TeV). Their arrival times relative to the LHC turn clock are stored with 50 ps resolution, allowing a histogram over the longitudinal beam profile, with sufficient statistics, to be generated in a few minutes. The light intensity on the SPAD sensor can be adjusted by moving individual neutral density (ND) filters in or out of the light path. With little filtering, the average number of *photons* per ghosts is high, but the *counts* per ghost that are within one deadtime of a main bunch is reduced due to the deadtime of the detector. For heavy filtering, the availability of the detector is higher, but the number of photons per ghost lower. The filtering strength should therefore be chosen to maximize the number of counts per ghost. A more detailed description of the system can be found in [1].

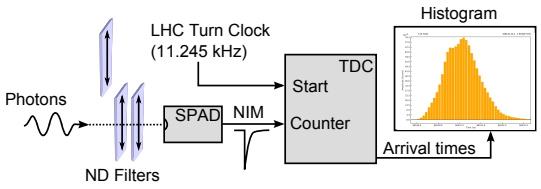

Figure 1: Schematic overview of the LDM system.

An example of part of the beam profile histogram measured by the LDM during Fill 3005 in 2012 is shown in Fig. 2 (integrating over 5 minutes, or 3.3 million turns). Two features must be taken into account when resolving the relative bunch population:

1. Reduced detector availability due to the detector deadtime.
2. Afterpulsing from main bunches, which is comparable to, or greater than, the signal from ghost and satellite bunches.

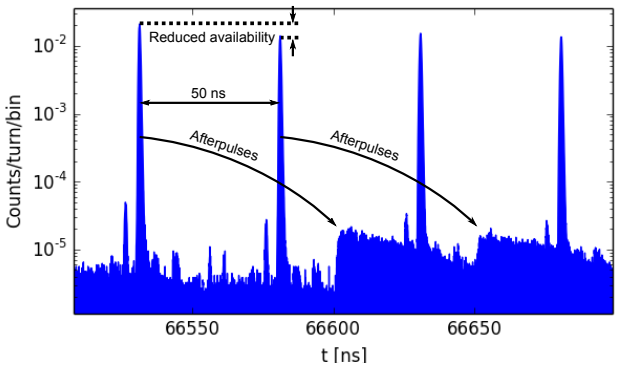

Figure 2: Part of the uncorrected profile histogram from fill 3005 (Beam 1). Main bunch separation is 50 ns. Satellites and ghosts are visible around the main bunches as smaller peaks.

The LDM sensors used in the LHC are of the PDM-series from Micro Photon Devices[1], with Peltier-cooling and integrated active quenching circuits. An identical spare detector was used for the characterization described in this paper, with the key figures summarized in Tab. 1.

* marcus.palm@cern.ch

[1] http://www.micro-photon-devices.com/

Table 1: Characteristics of the spare SPAD used in laboratory tests. (*= As measured in laboratory setup.)

Photon detection efficiency	49% (550 nm)
Deadtime, τ	≈72 ns
Afterpulse probability	4%
Dark count rate	300 Hz*
Timing resolution (FWHM)	50 ps
Sensor diameter	50 μm

METHODS

Detector characterization

The principle of active quenching is well described in e.g. [2]. After an avalanche occurs, the voltage over the SPAD is lowered below breakdown to quench the avalanche and allow most of the charge carriers that are trapped during the avalanche to be released. After about τ seconds, the deadtime, the voltage is restored.

Sensitivity It has been observed that as the voltage is restored after an avalanche, the relative probability for a photon or trapped charge carrier to trigger a new avalanche, $S[t]$, oscillates around its ideal, steady-state value of 1 during a few nanoseconds. To characterize this behavior, the SPAD was exposed to a laser pulse (<1 ns) and illuminated with a weak LED pulse (∼100 ns) during the time where the voltage is restored. By comparing the response to the LED light with and without the initial laser pulse, $S[t]$ could be measured. The result is shown in Fig. 3. $T_{s,min}$ and $T_{s,max}$ indicate where $S[t]$ begins to increase from 0 and settles at 1. $S[t]$ is referred to as the sensitivity response profile.

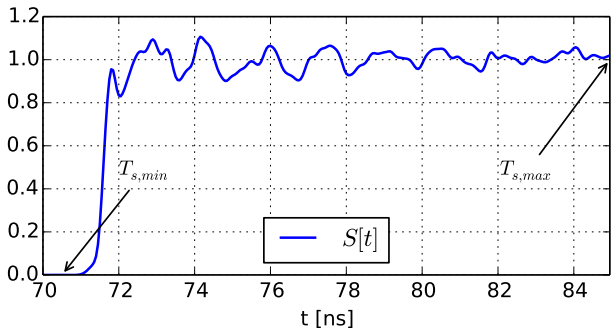

Figure 3: Sensitivity response profile after an avalanche at $t = 0$.

Afterpulsing We define $R[t]$ as the the probability that an avalanche at $t = 0$ will cause an afterpulse avalanche at time bin t, assuming there are no intermediate avalanches reducing the availability. To measure $R[t]$, the SPAD was exposed to a single laser pulse and the resulting afterpulse tail measured. When correcting the afterpulse tail for reduced availability and removing the constant background, the result is essentially an infinite series of convolutions between the initial laser profile $\delta[t]$ and progressively higher order afterpulses:

$$R_{tail}[t] = \delta[t] * (\underbrace{R[t]}_{\text{Primary}} + \underbrace{R[t] * R[t]}_{\text{Secondary}} + \ldots) \quad (1)$$

$R[t]$ is resolved by a deconvolution, and the result for the spare SPAD is shown in Fig. 4. Note that the fluctuations around t =72 ns corresponds to the same fluctuations seen in $S[t]$. The afterpulses were still measured several 100 μs after the initial laser pulse (inset, Fig. 4), albeit reduced by 4-5 orders of magnitude.

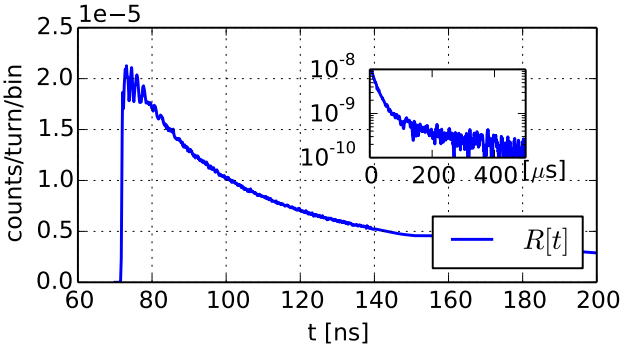

Figure 4: Single pulse afterpulse profile, $R[t]$.

Retrieving the photon rate

The goal of the correction algorithm is to retrieve the relative rate of photons striking the sensor. Avalanches registered in the histogram are assumed to have three causes: (1) Synchrotron radiation photons, $\gamma[t]$, (2) Afterpulses, $r[t]$, and (3) Background events (ambient photons, dark counts), $b[t] \ll 1$. Square brackets indicate an average number per time bin (50 ps) per turn. For a sufficiently long integration time, the measured histogram (divided by the number of turns), $a[t]$, contains the probability that an avalanche will be triggered in an individual time bin. The probability that no avalanche is measured (subscript m) in time bin t can then be written as:

$$1 - a[t] = P(\Gamma_m[t] + B_m[t] + R_m[t] = 0) \quad (2)$$

Here, $\Gamma_m[t]$, $R_m[t]$ and $B_m[t]$ are independent random variables from Poisson distributions with mean values $\gamma_m[t]$, $r_m[t]$ and $b_m[t]$. Note that these means are in general lower than $\gamma[t]$, $r[t]$ and $b[t]$, due to the detector deadtime. We introduce the *average sensitivity*, $h[t]$, as a scaling factor for the probability to measure an avalanche at t, as compared to an ideal detector that does not suffer from deadtime effects (similarly to what was done for deadtime correction in [1]). Setting $\Phi = \Gamma + B$, $h[t]$ relates Φ_m to Φ as (R_m is elaborated later):

$$P(\Phi_m > 0) = h[t]P(\Phi > 0) \quad (3)$$

For example if the first main bunch triggers an avalanche with 10% probability, $h[t]$ will be reduced to 0.9 immediately after the main bunch, reducing the number of avalanches from following ghosts and satellites by 10%. The probability that a Poisson-variable with mean λ is zero is $e^{-\lambda}$, from which Eq. 2 can be reshuffled (omitting the time-dependence of $b[t]$) to give:

$$\gamma[t] = -\ln\left[1 - \frac{1-(1-a[t])e^{r_m[t]}}{h[t]}\right] - b \qquad (4)$$

Average sensitivity

In previous analyses of LDM data, the sensitivity response profile $S[t]$ has been modeled as a binary function switching from 0 to 1 after a constant number of time bins. With $S[t]$ in fact behaving as shown in Fig. 3, a more careful approach is needed. To estimate h in a particular time bin t_2, only avalanches in the interval $[t_2 - T_{s,max}, t_2 - 1]$ will have an impact ($S[t]$ is restored to 1 for earlier avalanches). This interval can be split into three regions as shown in Fig. 5. Since the minimum time between two avalanches is $T_{s,min}$, there are only five possible configurations, **a**-**e**, of preceding avalanches in the three regions, as summarized in Tab. 2. Their respective probabilities and resulting single-turn sensitivity at t_2 are denoted p_a-p_e and h_a-h_e. The average sensitivity $h[t_2]$ is the weighted sum $\sum p_i h_i$. Note that with an avalanche in I_2 or I_3, $h[t_2]$ is 0.

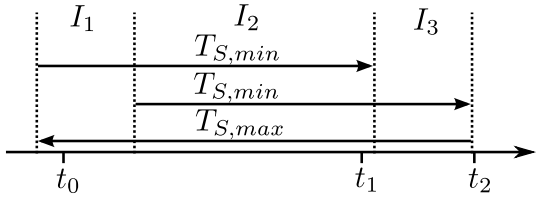

Figure 5: Regions considered to evaluate $h[t_2]$.

Table 2: Possible avalanche configurations in I_1-I_3 during a single turn (x=avalanche in region I_1, I_2 or I_3).

I_1	I_2	I_3	P	$h[t_2]$
-	x	-	p_a	$h_a[t_2] = 0$
x	-	-	p_b	$h_b[t_2] \neq 0$
x	-	x	p_c	$h_c[t_2] = 0$
-	-	x	p_d	$h_d[t_2] = 0$
-	-	-	p_e	$h_e[t_2] = 1$
			$\sum p_i = 1$	$h[t_2] = \sum p_i h_i[t_2]$

***a**:* If there was an avalanche in I_2, there could not have been one in I_1 and there can be none in I_3. Since avalanches in the I_2-bins never occur in the same turn, the probability for this configuration, p_a, is given by a sum of $a[t]$ over I_2: $p_a = \sum_{t' \in I_2} a[t']$.

***b**:* The probability of an avalanche at $t_0 \in I_1$ is $a[t_0]$. For this case to occur, there must be no subsequent avalanches up to $t_2 - 1$. Assuming that $h[t]$ is known for all $t < t_2$, we use the *corrected* avalanche rate, $a'[t]$:

$$a'[t] \equiv a[t]/h[t] \qquad (5)$$

to estimate p_b. Given an avalanche at t_0, the sensitivity at $t_1 > t_0$ is $S[t_1 - t_0]$ and subsequent avalanche probabilities are estimated as $S(t_1 - t_0)a'[t_1]$.

The probability of no avalanche in I_3, given an avalanche at t_0, is denoted $P(!I_3|t_0)$ and estimated to:

$$P(!I_3|t_0) = \prod_{t_1 = t_0 + T_{s,min}}^{t_2 - 1} (1 - S[t_1 - t_0]a'[t_1]) \qquad (6)$$

p_b is given by a sum over all $t_0 \in I_1$:

$$p_b = \sum_{t_0 \in I_1} a[t_0] P(!I_3|t_0) \qquad (7)$$

Finally, h_b is the weighted sum of $S[t_2 - t_0]$

$$h_b = \sum_{t_0 \in I_1} a[t_0] P(!I_3|t_0) S(t_2 - t_0)/p_b \qquad (8)$$

***c**:* Given an avalanche in I_1, this is the complement of case **b**

$$p_c = \sum_{t_0 \in I_1} a[t_0] P(I_3|t_0) = \qquad (9)$$
$$= \sum_{t_0 \in I_1} a[t_0](1 - P(!I_3|t_0)) = \left(\sum_{t_0 \in I_1} a[t_0]\right) - p_b$$

***d**:* The probability of an avalanche in I_3 is: $P(I_3) = \sum_{t_0 \in I_3} a[t_0] = p_c + p_d$. Therefore:

$$p_d = P(I_3) - p_c \qquad (10)$$

***e**:* This is the last case, i.e:

$$p_e = 1 - (p_a + p_b + p_c + p_d) \qquad (11)$$

To evaluate $h[t_2]$ for all t_2 in the entire histogram, $h[t_2]$ must be evaluated iteratively since, since $h[t_2 - 1]$ and $a'[t_2 - 1]$ are used.

Afterpulse rate

To estimate the rate of measured afterpulses, it is tempting to convolute $a[t]$ with $R[t]$ and then multiply with the average sensitivity $h[t]$. However, this would in general underestimate the afterpulsing rate where main bunches are closer than $T_{s,min}$ since two main bunches spaced by 25 or 50 ns will never cause an avalanche in the same turn. For example the rate of measured afterpulses from the first bunch in Fig. 2 is completely unaffected by the presence of the second main bunch, no matter how high its intensity. This single-turn correlation between a primary avalanche

and its afterpulses therefore needs to be addressed. To evaluate how an avalanche at t_0 affects the measured rate of afterpulses at t_2, we therefore distinguish between *recent* afterpulses, $r_r[t]$, and *late* afterpulses, $r_l[t]$. The total afterpulse rate $r_m[t]$ is the sum $r_m[t] = r_r[t] + r_l[t]$.

Recent afterpulses We define "recent" to mean $t_2 - t_0 \leq 2T_{s,min}$, i.e. about 140 ns in our case. Given an avalanche at $t_0 \in [t_2 - 2T_{s,min}, t_2 - T_{s,min}]$, an afterpulse at t_2 can occur only if there are no subsequent avalanches in the interval $[t_0 + 1, t_2 - 1]$. We estimate the probability for this to be:

$$P(![t_0+1, t_2-1]|t_0) = \prod_{t_1=t_0+1}^{t_2-1} 1 - S[t_1 - t_0]a'[t_1] \quad (12)$$

Here, the corrected avalanche rate $a'[t]$ is used. In the case of Fig. 2, the influence by the second main bunch on afterpulses from the first main bunch would be masked by the factor $S[t_1 - t_0]$.

Since no avalanches in $[t_2 - 2T_{s,min}, t_2 - T_{s,min}]$ can occur in the same turn, the total rate of recent afterpulses at t_2 is a straight sum over t_0:

$$r_r[t_2] = \sum_{t_0=t_2-2T_{s,min}}^{t_2-T_{s,min}} a[t_0]P(![t_0+1, t_2-1]|t_0)R[t_2-t_0] \quad (13)$$

Late afterpulses For late afterpulses, the assumption is made that the gap between avalanche occurrence t_0 and afterpulse time t_2 is sufficiently large that the average sensitivity $h[t_2]$ can be used to estimate the afterpulse:

$$\begin{aligned} r_l[t_2] &= \textstyle\sum_{t_0<t_2-2T_{s,min}-1} a[t_0]R[t_2 - t_0]h[t_2] = \\ &= h[t_2] \times (a * \tilde{R})[t_2] \end{aligned} \quad (14)$$

where $*$ denotes discrete convolution and $\tilde{R}[t]$ is identical to $R[t]$, but with the contribution from recent afterpulses masked:

$$\tilde{R}[t] = \begin{cases} 0 & \text{if} \quad t \leq 2T_{s,min} \\ R[t] & \text{if} \quad t > 2T_{s,min} \end{cases} \quad (15)$$

RESULTS

Afterpulse rate

In order verify the accuracy of the afterpulse rate, $r_m[t]$, the SPAD was illuminated with 300 laser pulses spaced at 25 ns with a repetition rate of 100 kHz to simulate the typical main bunch rate under operational conditions. The light was attenuated with neutral density (ND) filters of different strengths (0-2, corresponding to three orders of magnitude). Fig. 6 shows the measured histogram, $a[t]$ (solid), and the predicted afterpulse rate, $r_m[t]$ (dashed), at the end of the pulse train. $r_m[t]$ and $a[t]$ are in excellent agreement up to ND1, where the number of counts per main bunch is about 0.13 (=0.2 at full availability). At this count rate (5.2×10^6/s), the availability in the pulse train is in general reduced by about 40%.

With ND0, the SPAD is close to saturation, but the counts per bunch are only marginally higher (0.15). The discrepancy between $a[t]$ and $r_m[t]$ can be explained by an SPAD feature unaccounted for in the current model. At the end of the deadtime, there is a time window of a few ns where the bias voltage is above breakdown, but the signal output is still inactive. Avalanches occurring in this window will be unaccounted for, causing an overestimation of $h[t]$. The quenching of these avalanches is also delayed, so that the longer lasting avalanche current will trap more charge carriers and increase the afterpulse rate.

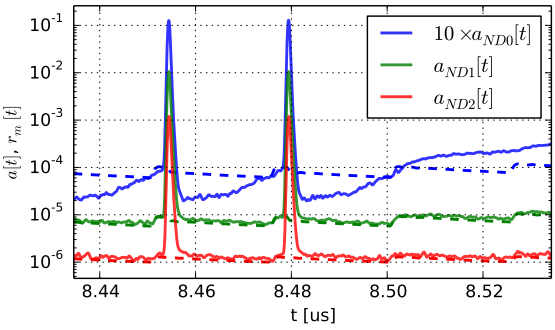

Figure 6: $a[t]$ (solid) and $r_m[t]$ (dashed) at the end of a 300 long 40 MHz pulse train. $a_{ND0}[t]$ is offset by a factor 10 for readability.

Photon rate

In order to verify the correction algorithm, the SPAD was exposed to two pulsed lasers with different intensity, simulating operational conditions with 40 MHz high intensity "main bunches" with intermediate "ghosts" of about 2.4×10^{-5} times lower intensity. $a_{H+L}[t]$ in Fig. 7 shows the initial part of the pulse train. The afterpulsing between main bunches shows a larger variation than the amplitude of the "ghosts", with almost discontinuous jumps one τ after each main bunch. When removing the afterpulses to estimate $\gamma[t]+b$ (Eq. 4), the baseline between main bunches is flattened, which greatly facilitates estimation of ghost and satellite population. As there are inevitably some unregistered avalanches, the estimated baseline is slightly higher than the "true" background rate (as a reference, $a_L[t]$ shows the signal with only the low intensity laser on). The higher baseline, which is fairly stable, can be explained by late afterpulses accumulated from unregistered avalanches.

LHC LDM data correction

The correction algorithm has also been tested on some data sets from LHC Run 1. One example is shown in Fig. 8.

Uncorrected, the afterpulses alone produce a baseline variation between main bunches that is several times larger than the ghost and satellite signal. After correction, the baseline is flattened and the satellite and ghost bunch profiles are more apparent.

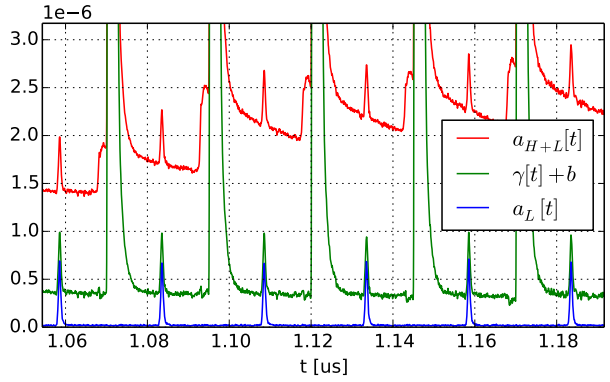

Figure 7: Initial part of LHC-like pulse train with intermediate ghosts (red: high+low intensity laser, blue: only low intensity laser, green: calculated $\gamma[t] + b$).

A gradual decrease of the corrected baseline between main bunches can be seen in the signal. This is an effect of the unregistered avalanches, which are more likely to happen where there are main bunches, and result in an overestimation of $h[t]$, with the effect that $r_m[t]$ is overestimated and $\gamma[t] + b$ consequently underestimated. The closer the window where avalanches can occur without being registered is to the main bunch separation (a multiple of 25 ns), the more enhanced this effect would be. It has been confirmed that the end of this window is very close to 75 ns. For Run 2, the deadtime of the SPADs has been adjusted to minimize this effect.

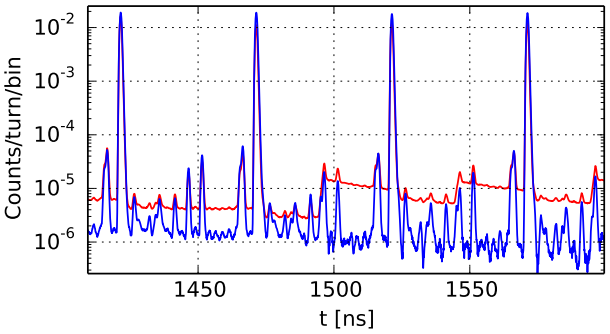

Figure 8: Raw (red) and corrected (blue) LDM data from fill 3005, Beam 2.

CONCLUSIONS

A model based characterizing the response of Single-Photon Avalanche Diodes has been developed to enable the removal of afterpulse signals to retrieve the underlying relative photon rate. Its accuracy and ability to distinguish between afterpulse and photon induced avalanches has been confirmed experimentally under LHC-like conditions at count rates over $5{\times}10^6$/s.

ACKNOWLEDGMENT

Many thanks to Georg Simmerle at MPD for his kind support and help in interpreting the data and behaviour of the SPADs, as well as for tuning the deadtime of the LDM detectors to our needs.

REFERENCES

[1] A. Jeff et.al, "Longitudinal density monitor for the LHC", Phys. Rev. ST-AB, p. 032803 (2012) 15(3)

[2] F. Zappa et.al, "Monolithic Active-Quenching and Active-Reset Circuit for Single-Photon Avalanche Detectors", IEEE Journal of solid-state circuits, p. 1298 (2003) 3(7)