

Characterisation of silicon photomultipliers in a dilution refrigerator down to 9.4 mK towards a cryogenic cosmic-ray muon veto system

The QUEST-DMC collaboration

A. Kemp^{a,*}, S. Autti^b, E. Bloomfield^c, A. Casey^d, N. Darvishi^d, D. Doling^d, N. Eng^d, P. Franchini^c, R.P. Haley^b, P. J. Heikkinen^d, A. Jennings^e, S. Koulosousas^{c,d}, E. Leason^{c,d}, L.V. Levitin^d, J. March-Russell^c, A. Mayer^b, J. Monroe^{a,c}, D. Münstermann^b, M.T. Noble^b, J.R. Prance^b, X. Rojas^d, T. Salmon^b, J. Saunders^d, J. Smirnov^f, R. Smith^a, M.D. Thompson^b, A. Thomson^b, A. Ting^d, V. Tsepelin^b, S.M. West^d, L. Whitehead^b and D.E. Zmeev^b

^aParticle Physics Department, UKRI STFC Rutherford Appleton Laboratory, Harwell, Didcot, OX11 0QX, U.K.

^bDepartment of Physics, Lancaster University, Bailrigg, Lancaster, LA1 4YB, U.K.

^cDepartment of Physics, University of Oxford, Keble Road, Oxford, OX1 3RH, U.K.

^dDepartment of Physics, Royal Holloway University of London, Egham Hill, Egham, TW20 0EX, U.K.

^eRIKEN Center for Quantum Computing, RIKEN, Wako, 351-0198, Japan

^fDepartment of Mathematical Sciences, University of Liverpool, Peach Street, Liverpool, L69 7ZL, U.K.

E-mail: ashlea.kemp@stfc.ac.uk

ABSTRACT: We report the characterisation of a FBK NUV-HD-cryo silicon photomultiplier (SiPM) sensor operated in a 9.4 ± 0.2 mK environment inside a dilution refrigerator, towards the development of a cryogenic cosmic-ray muon veto system to be operated internal to a dilution refrigerator required for low background experiments such as the QUEST-DMC dark matter search experiment. We characterise the single photon response and the gain (the charge produced per detected photon), the dark count noise rate, and correlated noise contributions as a function of operating voltage. This paper also reports first proof-of-concept measurements of using a SiPM coupled to scintillator internal to a dilution refrigerator, towards detecting high-energy events consistent with candidate cosmic-ray muon signals.

KEYWORDS: Cryogenic detectors; Photon detectors for UV, visible and IR photons (solid-state); Scintillators and scintillating fibres and light guides

ARXIV EPRINT: [2512.16769](https://arxiv.org/abs/2512.16769)

*Corresponding author.

Contents

1	Introduction	1
2	Experimental setup	2
2.1	Hardware	2
2.2	Data acquisition and processing	4
3	Characterisation analysis	7
3.1	Reverse I-V curves	7
3.2	Single photon response and gain	9
3.3	Breakdown voltage	10
3.4	Dark count rate	11
3.5	Direct crosstalk	14
3.6	Correlated delayed avalanches	16
4	Towards a cryogenic cosmic-ray muon veto system	19
5	Conclusions	21

1 Introduction

In recent years, silicon photomultipliers (SiPMs) have emerged as promising single photon detectors for use in particle physics, particularly in low background, rare-event search experiments such as direct dark matter detection experiments. Photomultiplier Tubes (PMTs) have long been the favoured technology for dark matter experiments, such as DEAP-3600 and LZ [1, 2], however compared to PMTs, SiPMs are cheaper to manufacture, require a lower operational voltage, and are insensitive to magnetic fields [3]. In addition to their high gain and high photon detection efficiency, they can have a lower intrinsic radioactivity compared to PMTs. One drawback of SiPMs is that they generally exhibit higher intrinsic thermal noise than PMTs. However, since thermal noise decreases proportionally with temperature, it can be suppressed by operating the SiPMs at cryogenic temperatures.

Whilst limited, there have been measurements showing that SiPMs can operate close to liquid helium (LHe) temperatures and below [4–6], illustrating that there is a growing interest to demonstrate the viability of operating these sensors in sub-Kelvin conditions. As such, SiPMs are being explored by the QUEST-DMC collaboration as a potential photon sensor for use in a cryogenic cosmic-ray muon veto system. The QUEST-DMC collaboration is developing, in one of its work packages, a direct dark matter search experiment projected to reach world-leading sensitivity to spin-dependent sub-GeV/ c^2 mass dark matter using a superfluid helium-3 (^3He) target enclosed in a $\sim 1\text{ cm}^3$ bolometer box [7–9]. The QUEST-DMC collaboration is operating bolometers at Royal Holloway, University of London and Lancaster University, in ultra-low temperature dilution refrigerators fitted with an adiabatic demagnetisation stage capable of cooling ^3He to operating temperatures below 300 μK .

Since QUEST-DMC is currently operated at sea level, and not underground like the majority of dark matter experiments, cosmic-ray induced background interactions will be a limiting factor for

QUEST-DMC’s dark matter search [10]. The cosmic-ray muon veto system is currently in development; the baseline design makes use of a scintillator volume surrounding the bolometer cell, where the exterior of this volume is coupled to the inner vacuum can which itself is coupled to the 10 mK stage of the dilution refrigerator, and is where the SiPM will be mounted. Therefore, 10 mK is the relevant temperature to characterise the SiPM in order to demonstrate the feasibility of this scheme. We are considering SiPMs over technologies traditionally deployed for photon detection in this temperature range due to the fact that they are large-area, commercially manufacturable, and simple to operate, compared with Superconducting Nanowire Single-Photon detectors (SNSPDs) which are much smaller, and Transition Edge Sensors (TESs) which have a relatively complex fabrication process.

SiPM technology has previously been used in cryogenic cosmic-ray muon veto systems, such as the NUCLEUS experiment [11], which makes use of a plastic scintillator deployed in a dilution refrigerator at 850 mK. In this configuration, the plastic scintillator is instrumented with wavelength shifting fibres, and is coupled to a SiPM that is thermally coupled to the 300 K part of the cryostat. However, we are exploring the option of directly coupling SiPMs to the plastic scintillator itself, which would require the SiPM technology to be able to operate in mK environments. Without amplifier electronics and in the absence of light, a single NUV-HD-cryo SiPM with a surface area of $12 \times 8 \text{ mm}^2$ dissipates $\mathcal{O}(\text{pW})$ of power whilst it is biased, based on the device’s properties measured at 77 K [12]. At this level of power dissipation, the heat load should be low enough such that a SiPM could be thermally coupled to a $\mathcal{O}(10^{\prime}\text{s})$ mK temperature stage of a dilution refrigerator without disrupting the overall cooling power of the cryostat.

In this paper, we show the first operation of a NUV-HD-cryo SiPM [12] in a dilution refrigerator, demonstrating the feasibility of deploying this technology in a cryogenic cosmic-ray muon system for the QUEST-DMC experiment. Section 2 outlines the experimental setup, the data acquisition, and data processing chain. Section 3 describes the characterisation analysis and results. Section 4 reports first proof-of-principle measurements of using a SiPM coupled to a scintillator and operated internal to a dilution refrigerator. To conclude, section 5 discusses whether, based on the characterisation measurements presented in section 3, the SiPM tested could meet the requirements of a cryogenic cosmic-ray muon veto system, and how this work informs the next-step R&D for the system.

2 Experimental setup

2.1 Hardware

We operate a single NUV-HD-cryo SiPM [12] of $12 \times 8 \text{ mm}^2$ surface area and $30 \mu\text{m}$ cell pitch inside a cryogen-free dilution refrigerator located at Royal Holloway, University of London. The dilution refrigerator is a modified Oxford Instruments Triton 200 [14]. A simplified schematic of the experimental setup is shown in figure 1. The SiPM is mounted onto a custom PCB that distributes bias voltage and is housed inside a $7 \text{ cm (L)} \times 3.8 \text{ cm (W)} \times 1.4 \text{ cm (D)}$ copper box, shown in figure 2. The copper box is mounted and thermally coupled to the mixing chamber plate via a copper stand attached perpendicular to the mixing chamber plate, also shown in figure 2.

The base temperature is measured using current sensing noise thermometry (CSNT) [13]. A CSNT sensor with resistance $2 \text{ m}\Omega$ is mounted onto the ultra-low temperature plate, which is suspended from the mixing chamber plate using alumina tubes. The mixing chamber plate is thermally coupled to the ultra-low temperature plate via a flexible copper link. Whilst the data was acquired for this work, the

base temperature as measured from the CSNT sensor on the ultra-low temperature plate, and thus, the mixing chamber plate, was $T_{\text{CSNT}} = 9.4 \pm 0.2$ mK. Further detail on the fridge design can be found in [14]. To ensure that the characterisation results we obtain actually reflect the SiPM operating in a 9.4 mK environment, we waited at least three days after cool-down to the mixing chamber base temperature before acquiring any data for this work. This allows for the plastic (FR-4) PCB on which the SiPM is mounted to thermalise to the same temperature as the copper.

We estimate the maximum temperature difference between the SiPM and the mixing chamber temperature measured by the CSNT thermometer assuming the worst possible thermal link, i.e., the SiPM is directly mounted onto the FR-4 PCB which itself is sat directly on the mixing chamber plate (SiPM \rightarrow FR-4 \rightarrow mixing chamber). In this scenario, all of the heat would flow vertically through the FR-4 bulk only. First, we estimate the power dissipation of the SiPM whilst operated in the dark, P_{SiPM} , as,

$$P_{\text{SiPM}} = \text{DCR} \cdot G \cdot q \cdot V_{\text{bias}}, \quad (2.1)$$

where DCR is the dark count rate, G is the gain, $q = 1.6 \times 10^{-19}$ C is the electron charge, and V_{bias} is the bias voltage. We take the upper DCR value reported at 77 K from [12] of 5 mHz/mm². We assume a gain of 3×10^6 , taken also from [12] for a SiPM with 30 μm pitch and at 5 V of overvoltage. We take $V_{\text{bias}} = 34$ V to correspond with the highest bias voltage we acquired characterisation data. With these values, we calculate $P_{\text{SiPM}} = 8.2$ pW, which we round up to 10 pW. We also add an additional 10 pW of power attributed to wiring within the dilution fridge, yielding a total $P_{\text{SiPM}} = 20$ pW. We calculate the temperature difference between the SiPM and mixing chamber, $\Delta T = P_{\text{SiPM}} \cdot R_{\text{th}}$, where R_{th} is the thermal resistance between the SiPM and mixing chamber. R_{th} is defined as:

$$R_{\text{th}} = \frac{L}{k \cdot A}, \quad (2.2)$$

where L is the PCB thickness = 1.6 mm, A is the SiPM surface area = 96 mm², and k is the thermal conductivity of FR-4. We take $k = 1.64$ mW/(m·K) as measured at 0.3 K in [15]. With this, we calculate a worst-case $R_{\text{th}} = 1.016 \times 10^4$ K/W. In reality, we anticipate R_{th} to be much lower owing to additional thermal pathways of: the soldering of the SiPM onto copper pads on the PCB; the metal screws attaching the PCB to the copper box; the electrical grounding to the copper box; and the direct mounting of the copper box onto the flexible copper stand attached to mixing chamber plate. With these values, we make an estimate of $\Delta T = 203$ nK. This temperature difference is negligible compared to $T_{\text{CSNT}} = 9.4 \pm 0.2$ mK, thus we are confident that the temperature of SiPM is essentially at the temperature of the mixing chamber. Furthermore, since we are operating at the base temperature of this fridge, there is no more cooling power available at the mixing chamber; even 1 μW of injected power would cause a measurable increase in the CSNT temperature ($\Delta T \sim 10$ mK assuming the worst-case thermal coupling), which we did not observe.

As figure 1 shows, all associated electronics are located outside of the dilution refrigerator at room temperature, including a custom-designed signal amplification board, providing a 130 V/V voltage gain. A Keysight E3649A provides the low-voltage supply required to power the board. A Keithley 2450 sourcemeter is used to deliver the bias voltage to the SiPM via the custom amplifier board and an external coaxial cable that connects to the top of the cryostat; an internal coaxial cable then delivers the bias voltage down to the SiPM mounted on the mixing chamber plate on

the cold side. In addition, an LED that emits in the red wavelength range is attached to the outside of the copper box, where there is a 1 mm clearance hole allowing a direct line of sight from the LED to the SiPM. The LED is connected to a signal pulse generator via a current-limiting resistor, both located outside of the dilution refrigerator, which is used to flash the LED whilst acquiring reverse I-V curves, discussed in section 3.1. Data is acquired with a Tektronix DPO72304DX 23 GHz bandwidth oscilloscope [16].

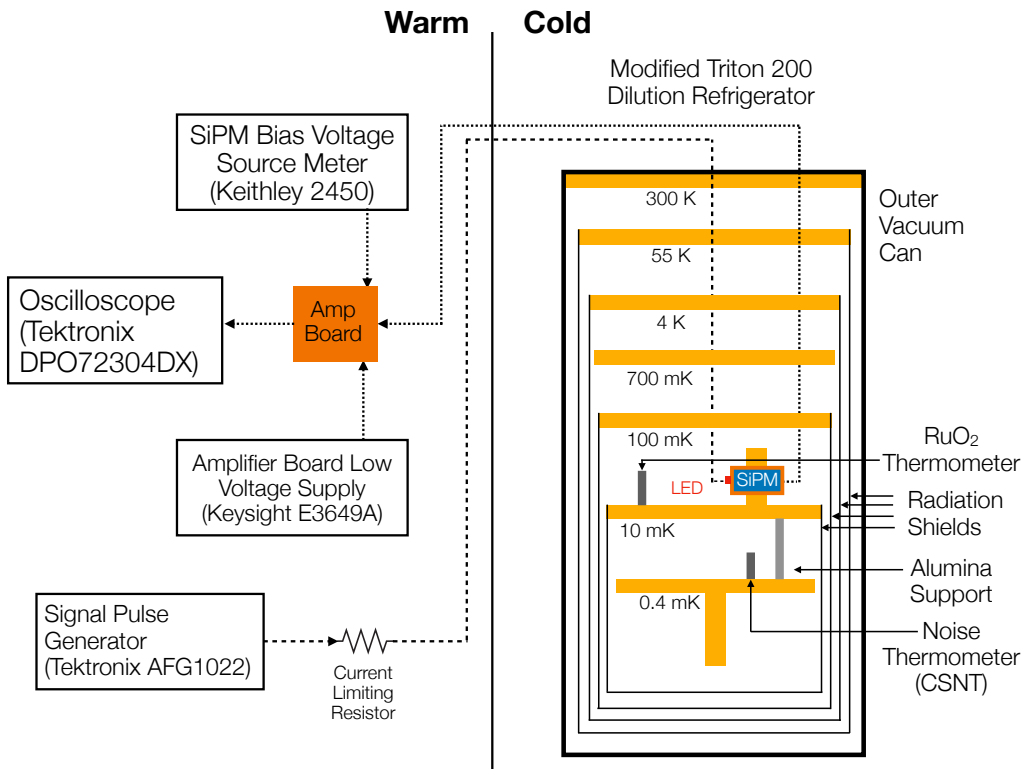


Figure 1. Simplified schematic diagram of the experimental setup used for the operation of a single NUV-HD-cryo SiPM in a cryogen-free dilution refrigerator.

2.2 Data acquisition and processing

We acquired three long datasets corresponding to three different operating bias voltages: $V_{\text{bias}} = 32 \text{ V}$, 33 V , and 34 V . Each dataset was acquired in dark conditions, i.e., the LED was switched off. In principle, there should be no other external light reaching the SiPM, as this fridge is inherently dark due to the significant radiation shielding implemented to enable reaching ultra-low temperatures. Therefore, the only signals we expect to observe in these datasets should originate from thermally-induced intrinsic noise associated with the SiPM known as the dark count, plus any additional correlated noise sources. The dark count rate of the NUV-HD-cryo technology is expected to be extremely low, as it is measured to be $< 5 \text{ mHz/mm}^2$ at 77 K [12].

Approximately 300,000 waveforms were acquired in self-trigger mode per voltage. Data were acquired using the Tektronix FastFrame Memory scheme [17], which segments the available memory in the oscilloscope such that each segment can be filled with a triggered acquisition at the desired

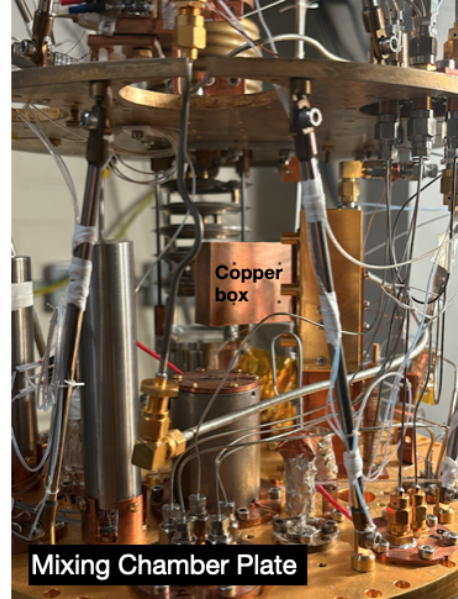
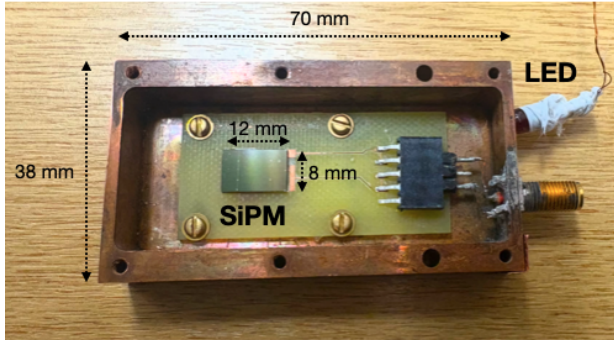


Figure 2. Left: picture of the single NUV-HD-cryo SiPM tested in this work, mounted onto a PCB and housed inside a 7 cm (L) \times 3.8 cm (W) \times 1.4 cm copper box. An LED that emits in the red wavelength range is glued to the outside of the copper box using araldite, over a 1 mm clearance hole allowing direct line of sight to the SiPM for acquiring reverse I-V curves in dark conditions. Right: picture of the copper box (shown in left figure) mounted inside the dilution refrigerator. The copper box is attached to a copper stand thermally coupled to the mixing chamber plate, which reaches a base temperature of $T_{\text{CSNT}} = 9.4 \pm 0.2$ mK.

sample rate with minimal dead time between the acquisitions. This is ideal for applications where pulses are separated by large intervals of inactivity in the signal [17], such as in our case. We acquire waveforms with a 250 MS/s sampling rate and a 100 μs acquisition window, requiring a 5 μs pre-trigger window. We set a 5 mV trigger level on the oscilloscope, which is terminated at 50 Ω , to trigger on the rising edge of the pulse. The trigger level was adjusted by eye in order to maximise triggers from real pulses and minimise triggers from electronics noise, and corresponds to approximately 70% of the raw mean height of a single photon pulse at $V_{\text{bias}} = 32$ V. Data is stored in a binary file format, saving the waveform data for each trigger, as well as the trigger time.

Each dataset is analysed using a simple pulse finding algorithm, in order to identify all candidate pulses in each 100 μs -long waveform, and not just the pulse which triggered the event. In each 100 μs -long waveform, a moving-average filter with a 80 ns window is applied to smooth out noise fluctuations. A baseline-subtracted waveform is computed using the baseline mean calculated in the first 3.2 μs of the waveform, in the pre-trigger region. We then search for regions of the waveform above a given threshold, which we set to 40% of the filtered mean amplitude of the single photon pulse for that voltage. We record the segment “start” time at which the signal first crosses the threshold, and the segment “end” time at which the signal drops below the threshold. This corresponds to the number of consecutive samples where the threshold condition was satisfied, known as the time-over-threshold (ToT). We set a lower threshold in the pulse finding algorithm to ensure that we are sensitive to pulses with amplitudes less than the single photon pulse, which could occur from correlated delayed avalanches occurring within the SiPM recharge time (discussed further in section 3.6).

Figure 3 (left) shows the ToT versus amplitude for candidate pulses at $V_{\text{bias}} = 32$ V. There are three distinct populations which correspond to the 1, 2, and 3 photon pulses. The 1 photon population is slightly asymmetric compared to the 2 and 3 photon populations, extending to lower amplitude and ToT; this likely originates from recharge-limited correlated delayed avalanches. In addition, there is a separate low amplitude-low ToT population, which is likely driven both by recharge-limited correlated delayed avalanches, and noise fluctuations occurring on the tail of a preceding pulse. We decide to apply an additional, but relaxed, requirement on the time-over-threshold (ToT) of ≥ 40 ns for pulse finding, to reduce contamination from noise fluctuations.

The consequence of requiring a lower pulse finding threshold is that we have less sensitivity to pulses occurring close together in time; if a secondary pulse occurs before the signal has dropped below the threshold condition from a primary pulse, the pulse finder will not have the sensitivity to resolve the secondary pulse from the primary. Figure 3 (right) shows the ToT projection of the single photon population illustrated by the red box in the left-hand figure. The mean ToT for the single photon pulse is 192 ns, which means that on average, secondary pulses that occur within 192 ns of a primary pulse will not be resolved as a separate pulse. The mean ToT value for the single photon population is consistent to within 20 ns across all three V_{bias} values we operated at. For cases where a secondary pulse occurs below the signal drops below threshold, the pulse finder will identify the pulse with the higher amplitude, which predominately originates from the primary pulse. This can introduce a small bias in the characterisation of the correlated delayed avalanche probability, characterised in section 3.6.

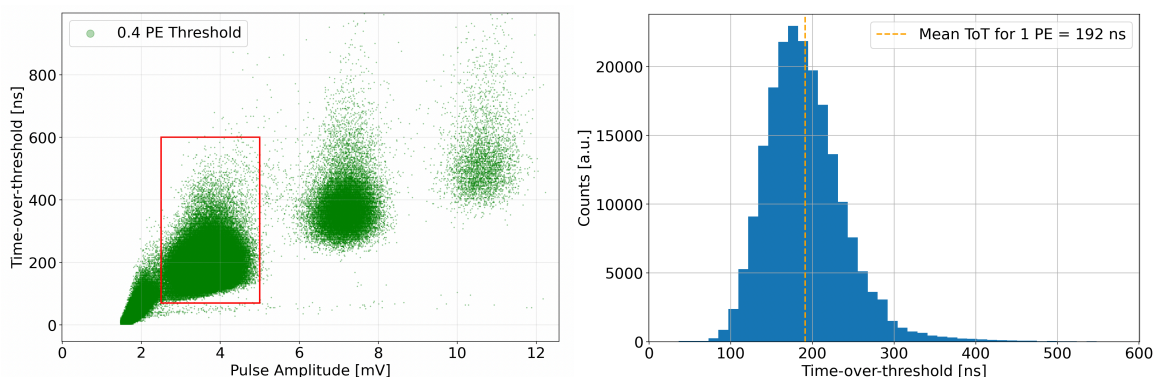


Figure 3. Left: distribution of the time-over-threshold (ToT) versus amplitude for candidate pulses that exceed the pulse finding threshold (40% of the filtered mean amplitude of the single photon pulse) at $V_{\text{bias}} = 32$ V. Right: ToT projection of the single photon population illustrated by the red box in the left-hand figure.

Figure 4 shows an example waveform acquired at $V_{\text{bias}} = 32$ V, corresponding to a single dark count rate signal, which is equivalent to the single photon signal. The raw waveform (WF), the baseline-subtracted (BL-sub) and moving-average filtered waveform (MA-filtered WF), the pulse finding threshold, and the identified pulse time (green) are shown. The pulse time, t_p , is defined as the time corresponding to the maximum amplitude. We also calculate a pulse area (proportional to charge) by integrating the waveform between $[t_p - 100 \text{ ns}; t_p + 500 \text{ ns}]$; this is illustrated by the shaded purple region. In addition to this information, we determine the absolute time difference between the current pulse and the preceding pulse, dt , calculated using the time of the pulses within their respective waveforms, the unix timestamp of the waveforms, and the fraction of a second past the unix timestamps of the waveforms. All of this information is then stored in a data file upon which the downstream characterisation analysis described in section 3 is performed.

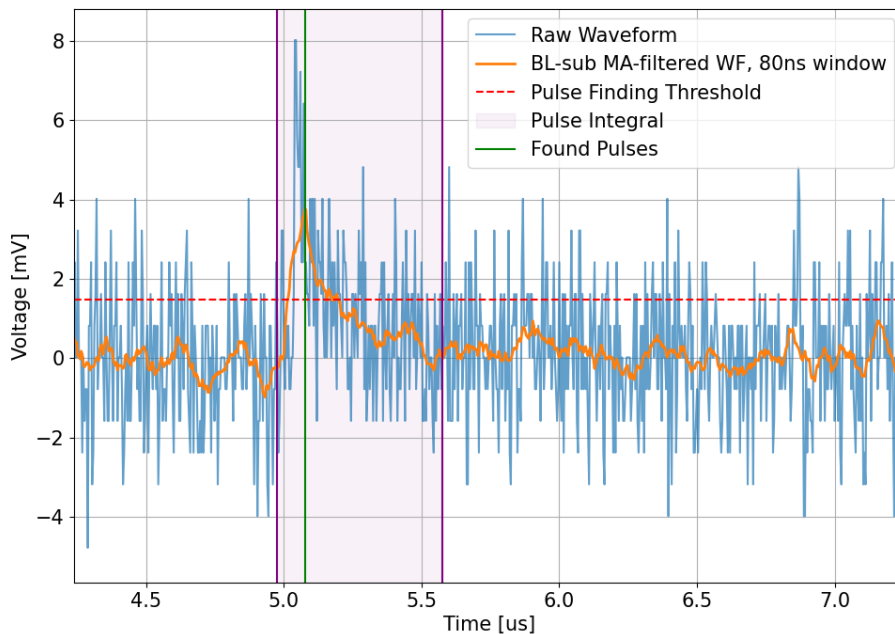


Figure 4. Example waveform (WF) acquired at $V_{\text{bias}} = 32$ V, corresponding to a single dark count rate signal, which is equivalent to the single photon signal. Raw WF (blue), baseline-subtracted (BL-sub) and moving-average (MA) filtered WF (orange), pulse finding threshold (red), and identified pulse time (green) are shown. The shaded purple region corresponds to the region $[t_p - 100 \text{ ns} : t_p + 500 \text{ ns}]$ where the pulse integral is computed to measure the charge.

3 Characterisation analysis

3.1 Reverse I-V curves

We first measured reverse I-V curves using the Keithley 2450 sourcemeter both at room temperature (prior to the cool-down), and after the cool-down to the mixing chamber base temperature, $T_{\text{CSNT}} = 9.4 \pm 0.2$ mK. The reverse I-V curve is used to characterise the relationship between the current flowing through the SiPM and the applied bias voltage, and can be used to determine its breakdown voltage, defined as the minimum bias voltage above which the SiPM can operate in Geiger mode and detect single photons. Our main motivation for acquiring reverse I-V curves was to ensure that the SiPM had survived the cool-down to base temperature before we attempted to acquire full characterisation data.

At room temperature, the dark count rate of the SiPM is large enough such that breakdown can be observed when the SiPM is in dark environment. At cryogenic temperatures however, an external light source is required to trigger breakdown. A signal pulse generator (1 kHz, 999 μs pulse width) and a current-limiting resistor were used to bias the LED whilst the I-V curve was acquired. Both the room temperature and mixing chamber base temperature reverse I-V curves were measured between 0–35 V, in 300 mV steps, with a 10 ms time delay between subsequent voltage steps during the sweep. The current limit on the Keithley 2450 was set to 60 μA to avoid damage to the SiPM. Figure 5 (left) shows a comparison of the reverse I-V curve taken at room temperature prior to cool-down, versus at the mixing chamber base temperature post cool-down. We clearly observe the SiPM breakdown both pre and post cool-down, demonstrating that the SiPM survived the cool-down to $T_{\text{CSNT}} = 9.4 \pm 0.2$ mK. In addition, we observe a clear shift downwards in the breakdown voltage

(i.e., the onset of self-sustaining avalanches) between the room temperature measurement and after cool-down, as expected with decreasing temperature.

We observed that after acquiring the reverse I-V curve, the temperature of the mixing chamber plate increased to ~ 80 mK. This was measured using a Ruthenium Dioxide (RuO_2) sensor mounted onto the mixing chamber plate, which measures the temperature accurately down to 20 mK; below 20 mK, the mixing chamber temperature is measured with the CSNT sensor. Given that the SiPM is thermally-coupled to the mixing chamber plate prior to commencing the I-V scans, it is likely that the temperature of the SiPM itself also increased during the sweeps. It is therefore very difficult to determine the breakdown voltage using the reverse I-V curve at a guaranteed temperature of 9.4 mK.

Furthermore, using follow up measurements of the SiPM submersed in a liquid helium (LHe) dewar at $T = 4$ K, we observed that the reverse I-V curve shape is strongly dependent on the delay time used between subsequent voltage steps during the sweep. Figure 5 (right) shows reverse I-V curves taken in the LHe dewar using a delay time of 10 ms (as was used for the reverse I-V curves taken at the mixing chamber base temperature and at room temperature), versus 1 s. Figure 5 shows that at $T = 4$ K, there is a clear difference between the reverse I-V curves using delay times of 10 ms and 1 s, both in the absolute current in the linear pre-breakdown regime, and in the apparent breakdown voltage.

We initially chose a 10 ms delay time between voltage steps as we did not want to inject unnecessary heat into the dilution refrigerator, however, we did not anticipate that there would be such an impact in the reverse I-V curve shape. This is likely driven by an increase in the SiPM quenching resistance at cryogenic temperatures, which lengthens the microcell recharge and thus requires a longer delay time during the sweep in order to get an accurate current measurement. We therefore cannot use the reverse I-V curves to get an accurate measurement of the breakdown voltage, which we also cannot guarantee to be the value of the breakdown at $T_{\text{CSNT}} = 9.4 \pm 0.2$ mK due to the observed heating during the sweep. As such, we estimate the breakdown voltage using an alternative method, described in section 3.3.

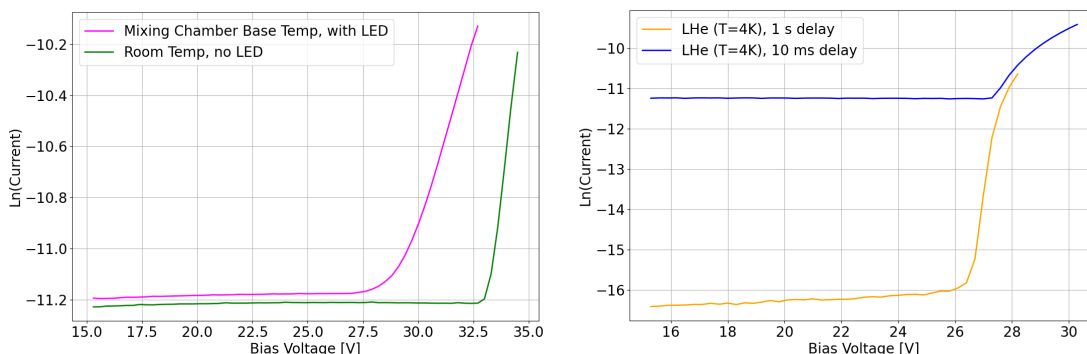


Figure 5. Left: reverse I-V curves taken at room temperature prior to cool-down (green), and at the mixing chamber base temperature post cool-down (magenta). Right: reverse I-V curves taken whilst the SiPM was submersed in a LHe dewar at $T = 4$ K, acquired using a delay time of 10 ms (blue), and 1 s (orange) between subsequent voltage steps, demonstrating the impact that the delay time has on both the absolute current value in the linear pre-breakdown regime, and on apparent breakdown voltage.

3.2 Single photon response and gain

We first characterise the single photoelectron (SPE) response of the SiPM. To do this, we plot two histograms corresponding to the pulse amplitude and pulse area for each dataset. As an additional data cleaning cut for the SPE analysis, we only consider pulses where the time difference between the pulse in question and the preceding pulse is $\geq 2 \mu\text{s}$. This is to reject pulses with amplitudes less than the single photon pulse, that can occur from correlated noise or from noise fluctuations on the tails of the previous pulse.

Figure 6 shows the pulse amplitude (left) and pulse area (right) distributions of the SiPM whilst operated at $V_{\text{bias}} = 32 \text{ V}$. In both histograms, there are distinct peaks for successive numbers of PE. The first peak is the 1 PE signal, i.e., observed output pulse following the detection of a single photon. Due to intrinsic noise, a SiPM still generates output pulses in the absence of light. The 1 PE peak we observe in our distributions is due to a combination of thermally-induced dark count noise, and noise from correlated delayed avalanches occurring at times later than the SiPM recharge time. These noise sources are discussed in detail in sections 3.4 and 3.6 respectively. The successive peaks corresponding to > 1 PE we observe in our distributions are a result of correlated noise from direct crosstalk, which manifest as superpositions of the 1 PE waveform. This noise source is discussed further in section 3.5.

The mean SPE response is extracted from a gaussian fit to the first peak in each spectrum. The SPE response at $V_{\text{bias}} = 32 \text{ V}$ is calculated to be $\mu_{\text{SPE,amp}} = 3.65 \text{ mV} \pm 0.42 \text{ mV}$ and $\mu_{\text{SPE,area}} = 0.15 \text{ V} \pm 0.02 \text{ V}$ in amplitude and area respectively. The mean SPE amplitude is smaller than the raw mean amplitude due to the moving-average filter which is applied to the data prior to running the pulse finding algorithm which reduces the height of the pulses; this can be seen in figure 4.

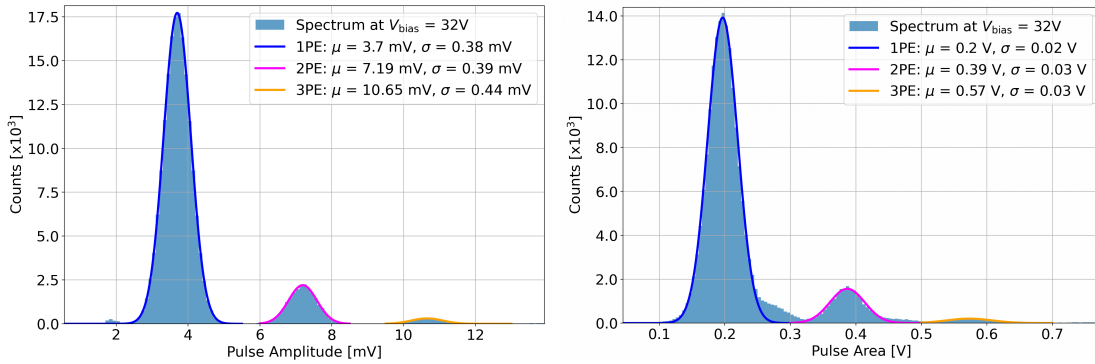


Figure 6. Pulse amplitude (left) and area (right) distributions of the NUV-HD-cryo SiPM operated at $V_{\text{bias}} = 32 \text{ V}$. Gaussian functions are fit to the first 3 peaks in each histogram. The first peak in each histogram corresponds to the 1 PE peak.

The separation between each pair of consecutive peaks in figure 6 (right) is constant and is proportional to the gain, G , defined as the amount of charge generated in an avalanche in response to each detected photon. Each microcell in the SiPM produces a highly quantised amount of charge when a photon is absorbed in the active volume, and this response is extremely uniform across all microcells. This unique property of SiPMs results in excellent charge resolution, and allows for a very precise measurement of the gain. The gain is expected to increase proportionally with higher ΔV .

To determine the separation, S , we use the results of the gaussian fits to the remaining peaks in the pulse area spectrum. We fit out the separation between peaks as a parameter in a linear polynomial fit to the fitted means of the 1, 2, and 3 PE peaks. The separation is calculated to be $S_{\text{area}} = 0.149 \pm 0.0003$ V at $V_{\text{bias}} = 32$ V. S_{area} is multiplied by the oscilloscope sampling period to convert the separation into units of charge, which is then divided by the electron charge to calculate the gain:

$$G = \frac{S_{\text{area}}[\text{V}] \cdot t_s[\text{s}]}{R_{\text{scope}}[\Omega] \cdot G_{\text{board}} \cdot q_e[\text{C}]}, \quad (3.1)$$

where $R_{\text{scope}} = 50 \Omega$ is the input resistance of the oscilloscope, $G_{\text{board}} = 130$ V/V is the gain of the custom amplifier board from the design specification, $q_e = 1.6 \times 10^{-19}$ C is the electron charge of an electron, and $t_s = 4$ ns is the sampling period.

Figure 7 shows the measured gain as a function of V_{bias} (ΔV). As expected, we observe a proportional increase in the gain with ΔV . The measured gain ranges between 7×10^5 and 1.2×10^6 for $32 \text{ V} < V_{\text{bias}} < 34 \text{ V}$. At 77 K, the gain is reported to be approximately $G = 3.2 \times 10^6$ for the NUV-HD-cryo technology [18], which is a factor of 3.7 higher compared to our value of $G = 8.6 \times 10^5$ for the same overvoltage value (7 V) based on an interpolation between our data points.¹ This difference could be driven by a ‘‘freeze-out’’ of free carriers at ultra-low temperatures, a mechanism which has previously been reported at LHe temperatures [19], however we do not explore this further in this manuscript.

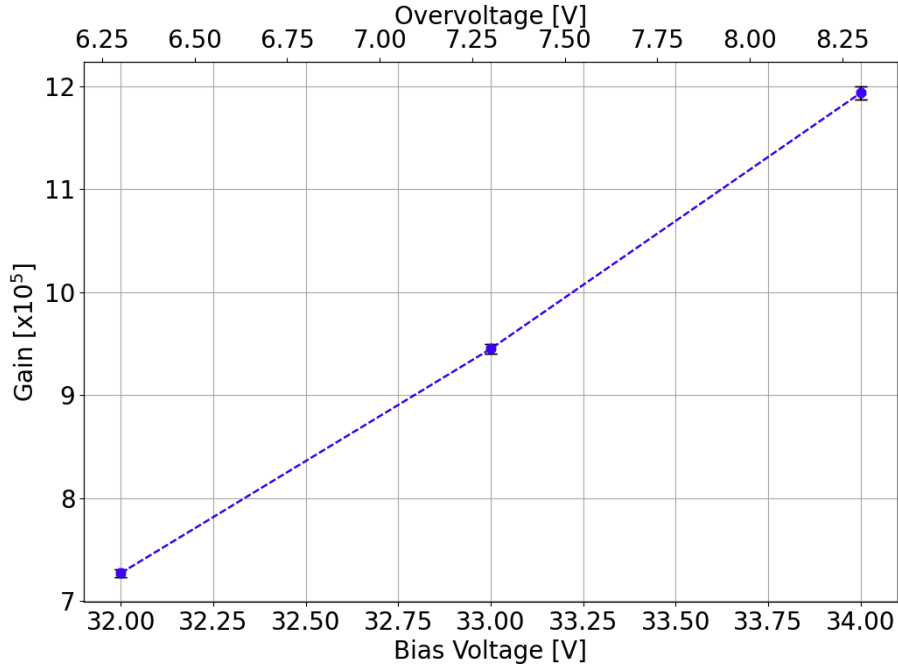


Figure 7. Gain of NUV-HD-cryo SiPM as a function of V_{bias} . Statistical error bars are shown, however are not necessarily visible as the errors are small. Blue dashed line is interpolated between the data points.

3.3 Breakdown voltage

To determine the breakdown voltage (V_{bd}) of the SiPM at $T_{\text{CSNT}} = 9.4 \pm 0.2$ mK, we linearly extrapolate the mean SPE amplitude $\mu_{\text{SPE,amp}}$ down to zero. Since the pulse amplitude is proportional to the

¹The overvoltage is defined in section 3.3.

prompt avalanche current, it is sensitive to the onset of self-sustaining avalanches and therefore can be used to estimate the breakdown voltage.

We first identify candidate 1 PE signals using the pulse finder algorithm described in section 2.2. Similarly to the SPE analysis described in the previous section, to ensure that we obtain a clean sample of 1 PE signals, i.e., excluding recharge-limited pulses with amplitudes < 1 PE or noise fluctuations on the tail, we require that the time between the pulse in question and the previous pulse $\geq 2 \mu\text{s}$. We then search for the maximum amplitude of the pulse in a second filtered version of the waveform, which uses a much shorter MA window of 8 ns (2 samples), within the ToT region identified by the pulse finder.

We do not use the 80 ns MA filtered pulse amplitude for this analysis since the filter reduces the 1 PE pulse amplitude; a longer MA window is intentionally chosen for the pulse finder to more effectively filter out baseline noise, however this does result in a decrease in the pulse height. Using this amplitude in the extrapolation would result in an overestimation of V_{bd} . We therefore use the 8 ns MA filtered amplitude, which is a much closer proxy to the actual height of the 1 PE amplitude. We also repeat the analysis using the raw waveform amplitude, however this results in a large uncertainty on the estimated value of V_{bd} due to the presence of additional noise in the raw waveforms. We instead use this result as a cross-check with the result obtained from the 8 ns MA filtered amplitude.

Similarly to the SPE characterisation, we fit a gaussian to the 1 PE peak amplitude distribution in order to determine $\mu_{\text{SPE,amp}}$ at each bias voltage. We fit these data points with a linear function and extrapolate down to the value of V_{bias} at which $\mu_{\text{SPE,amp}}$ goes to zero. Figure 8 shows $\mu_{\text{SPE,amp}}$ (using the 8 ns MA filtered amplitude) as a function of V_{bias} , including the linear fit and uncertainty band; we determine a breakdown voltage of $V_{\text{bd}} = 25.70 \pm 0.04 \text{ V}$ with this method. Repeating this analysis with the raw waveform amplitude yields a breakdown voltage of $V_{\text{bd}} = 25.47 \pm 0.58 \text{ V}$. Since our two measurements are consistent within uncertainties, we report the 8 ns filtered MA amplitude result. From hereafter, we report all characterisation measurements both in terms of the bias voltage and in the overvoltage, defined as $\Delta V = V_{\text{bias}} - V_{\text{bd}}$.

The breakdown voltage of the SiPM is expected to decrease with lower temperatures. At 77 K, the breakdown voltage of the NUV-HD-cryo technology is expected to be $V_{\text{bd}} = 27.1 \text{ V}$ [12]. As the selected photon sensors for the upcoming DarkSide-20k dark matter experiment, the collaboration has tested 359040 NUV-HD-cryo SiPMs at 77 K. The DarkSide-20k collaboration has determined the average breakdown voltage at 77 K to be $V_{\text{bd}} = 27.19 \pm 0.05 \text{ V}$ [20], which is consistent with the expected value reported in [12]. The single device we are testing originates from an engineering run used for pre-production assembly, and is of the same type as the DarkSide-20k SiPMs. Therefore, our measurement of $V_{\text{bd}} = 25.70 \pm 0.04 \text{ V}$ at $T_{\text{CSNT}} = 9.4 \pm 0.2 \text{ mK}$ is consistent with the expected behaviour of V_{bd} to decrease with lower temperatures for these devices.

3.4 Dark count rate

SiPMs can exhibit both uncorrelated and correlated noise contributions. The dark count rate (DCR) is the main source of uncorrelated noise. Each dark count is initiated by a thermally-generated charge carrier, which triggers an avalanche in the active region through the same process as a photon-generated charge carrier; thus DCR events are indistinguishable from single photon signals. The DCR of the SiPM is expected to increase with ΔV [12], since a higher ΔV increases the probability of an avalanche being triggered. The DCR is also expected to increase as a function of temperature. At higher temperatures, the DCR is dominated by field-enhanced thermal generation of charge carriers.

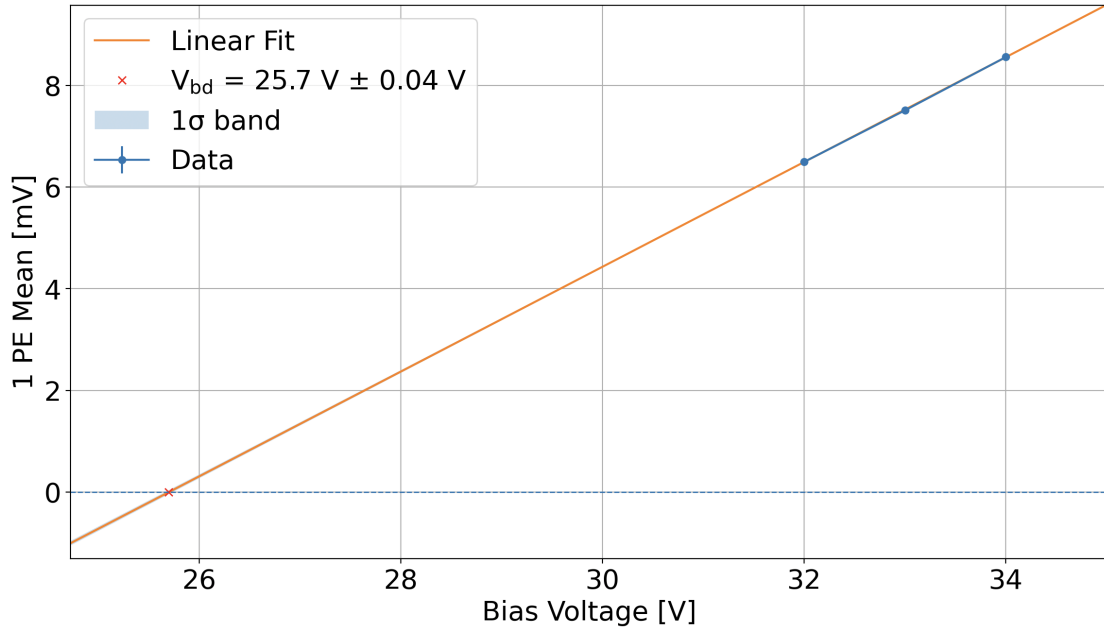


Figure 8. Mean SPE amplitude $\mu_{\text{SPE,amp}}$ as a function of V_{bias} , using the maximum amplitude obtained from the 8 ns MA filtered waveform. Linear fit and uncertainty band are also shown. The breakdown voltage extracted from extrapolating $\mu_{\text{SPE,amp}}$ down to zero is $V_{\text{bd}} = 25.70 \pm 0.04$ V.

However, below a certain temperature, the reduction of the DCR reaches a plateau, where the DCR becomes dominated by tunnelling effects that are weakly temperature dependent.

We characterise the DCR of the NUV-HD-cryo SiPM using the unshadowing method, based on using the time distribution between two consecutive pulses described in [21]. Given the incredibly low DCR of this technology, we need to consider the time differences of pulses across separate waveforms in addition to time differences within waveforms. Assuming a DCR of 5 mHz/mm² at 77 K [12], this corresponds to a DCR of 0.5 Hz for a single SiPM. If we were to acquire 600,000 randomly-triggered waveforms, we would expect only 30 dark counts. Thus, to only consider time differences within a waveform would require extremely long waveforms which is not feasible from a data acquisition perspective. We make the assumption that we lose a negligible number of pulses between waveforms as a result of acquiring the data using the Tektronix FastFrame Memory scheme [17] described earlier in section 2.2, which has a maximum dead time of 2.5 μ s per waveform segment.

To select primary pulses for this analysis, we make two requirements. First, we require $\mu_{\text{SPE,amp}} - 3\sigma_{\text{SPE,amp}} \leq A_{p_i} \leq \mu_{\text{SPE,amp}} + 3\sigma_{\text{SPE,amp}}$, where A_{p_i} is the amplitude of the primary pulse, and $\mu_{\text{SPE,amp}}$ and $\sigma_{\text{SPE,amp}}$ are the SPE amplitude mean and standard deviation respectively. Secondly, we require that the time difference between the primary pulse and the previous pulse is ≥ 1 ms. This cut is made to reduce the probability of the secondary pulse not being associated with the primary pulse to be negligible, whilst only removing a very small fraction ($\sim 0.05\%^2$) of DCR events.

Figure 9 (left) shows the measured time distribution of consecutive pulses whilst the SiPM is operated at $V_{\text{bias}} = 32$ V ($\Delta V = 6.3$ V). The dip in the dt distribution at close to the waveform

²Based on the time between subsequent events following an exponential distribution for a random Poisson process, and assuming a rate of 5 mHz/mm² \equiv 0.5 Hz.

acquisition window, $100 \mu\text{s}$, can be attributed to a loss in the pulse finder efficiency for pulses that occur near the end or beginning of waveform that may be partially cut-off, as well as from the dead time between waveforms itself, showing that the loss of pulses between waveforms is not necessarily negligible. We also observe that if we apply a stricter (larger) dt cut between the primary pulse and the preceding pulse, the effect of the dip is slightly enhanced.

Figure 9 (right) shows the unshadowed time distribution following the procedure in [21]. The unshadowed distribution is the average number of secondary pulses occurring between the bin limits of t_i and t_{i+1} defined in figure 9 (left), normalised per unit time to determine the pulse rate. The pulse rate distribution contains two contributions: the correlated delayed pulse rate, and the DCR. At large times, the rate of correlated delayed pulses is expected to disappear, leaving only the DCR. To extract the DCR, we compute the weighted mean of the asymptotic rate, between $1 \text{ s} < dt < 8 \text{ s}$. We calculate the DCR to be $4.046 \pm 0.217 \text{ mHz/mm}^2$ at $V_{\text{bias}} = 32 \text{ V}$ ($\Delta V = 6.3 \text{ V}$), the value of which is indicated by the horizontal dashed line in figure 9 to guide the eye. Figure 10 shows the measured DCR as a function of V_{bias} (ΔV). As expected, the DCR increases as a function of ΔV , increasing by 25% for $6.3 \text{ V} < \Delta V < 8.3 \text{ V}$.

We compute a cross-check of our DCR measurements from the unshadowed method by calculating $N_{\text{primaries}}/\text{lifetime}$ for each voltage, where $N_{\text{primaries}}$ is the number of selected primary pulses, and lifetime is defined as the total real acquisition time minus dead time. Assuming a maximum dead time of $2.5 \mu\text{s}$ per waveform segment, this corresponds to $< 0.0002\%$ of the total acquisition time per dataset, and thus is negligible. With this method, we calculate DCR values of 4.025 mHz/mm^2 , 4.516 mHz/mm^2 , and 4.594 mHz/mm^2 for $V_{\text{bias}} = 32 \text{ V}$, 33 V , and 34 V respectively. Based on the DCR values calculated from the unshadowed method shown in figure 10, we determine that these two independent DCR are consistent to within 0.1σ , 0.6σ , and 1.4σ for these three voltages respectively.

We note that the DCR values we calculate here are also consistent with the measured DCR value of $< 5 \text{ mHz/mm}^2$ at 77 K from [12], which could suggest that for temperatures below 77 K , the DCR is approximately constant. However, this is based on our measurement of a single device at $T_{\text{CSNT}} = 9.4 \pm 0.2 \text{ mK}$; measurements of several devices at more temperature points between 77 K and our base temperature would be required to confirm this.

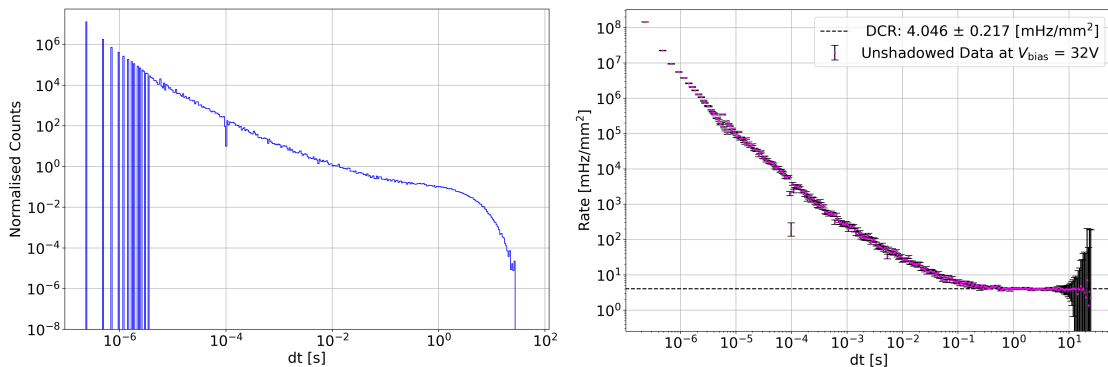


Figure 9. Left: measured time distribution (dt) of consecutive pulses at $V_{\text{bias}} = 32 \text{ V}$ ($\Delta V = 6.3 \text{ V}$). Right: unshadowed time distribution, based on dt distribution in left-hand plot. Horizontal dashed line indicates calculated DCR of $4.046 \pm 0.217 \text{ mHz/mm}^2$.

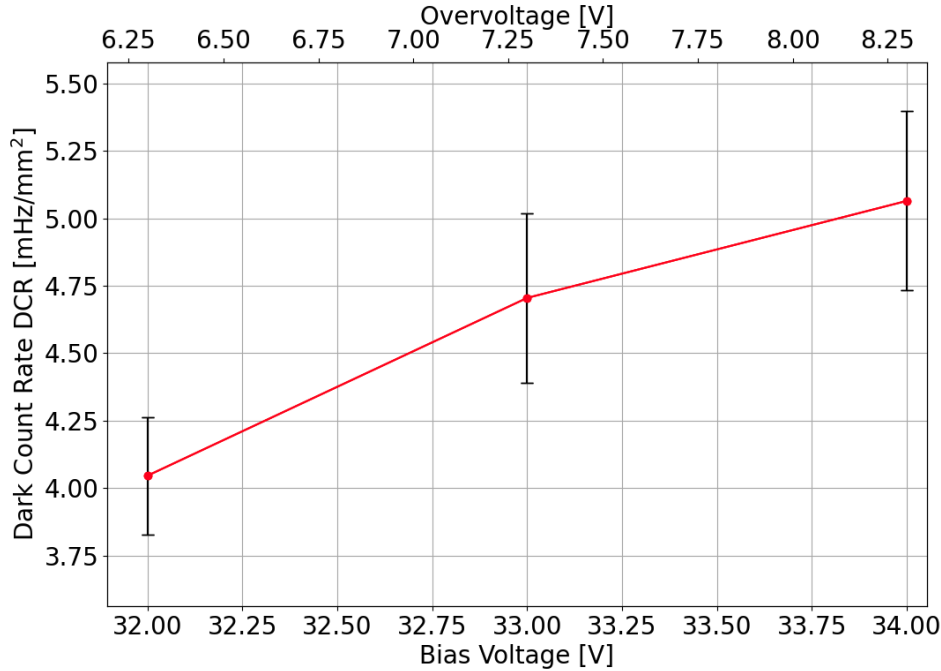


Figure 10. Measured DCR as a function of V_{bias} (ΔV). The DCR increases by 25% between $6.3 \text{ V} < \Delta V < 8.3 \text{ V}$. Red line is interpolated between the data points.

3.5 Direct crosstalk

Optical crosstalk is a source of correlated noise in a SiPM that occurs due to the emission of a small number of photons ($\sim 10^{-5}$ photons/electron) during an avalanche [22], which can trigger avalanches in the surrounding microcells. Optical crosstalk can be split into two categories:

1. Direct crosstalk (DiCT) occurs when a photon directly generates charge carriers in the active region of a neighbouring microcell, resulting in an almost simultaneous avalanche with respect to the primary avalanche.
2. Delayed crosstalk (DeCT) occurs when a photon generates charge carriers outside of the active region of a neighbouring microcell. The charge carriers can diffuse to the active region of the microcell, triggering a delayed avalanche.

DiCT manifests as a superposition with the 1 PE (single photon) pulse. For example, if a primary avalanche triggers two secondary crosstalk avalanches, the resulting output pulse will be a 3 PE pulse. DiCT can therefore be a limiting factor for applications requiring accurate photon-counting, since it artificially increases the number of detected photons and degrades the signal-to-noise ratio. The DiCT probability is expected to increase with ΔV , since the probability of an avalanche increases with ΔV . In addition, the number of emitted photons is linearly proportional to ΔV , which also contributes to an increase in the DiCT probability. There appears to be a weak dependence on the DiCT probability with temperature for the NUV-HD-cryo SiPM technology, as reported in [12].

To characterise the DiCT of the SiPM, we measure the pulse count rate as a function of threshold within a prompt coincidence window. As the threshold is increased, the count rate decreases in a step-like fashion, leading to a characteristic “staircase” where each step corresponds to excluding

events with increasing numbers of fired microcells (1 PE, 2 PE, etc). The rate measured at the top step, $R_{0.5PE}$, is due to all detectable pulses, whilst the rate measured at the second step, $R_{1.5PE}$, is due to DiCT pulses. The probability of DiCT, P_{DiCT} , can then be defined as:

$$P_{DiCT} = \frac{R_{1.5PE}}{R_{0.5PE}}, \quad (3.2)$$

Figure 11 (left) shows the characteristic staircase obtained at $V_{bias} = 32$ V ($\Delta V = 6.3$ V). By taking the ratio of the blue and red dashed lines, we determine $P_{DiCT} = 15.8\%$.

For this analysis, we only consider pulses which triggered the waveform, since we know the pulse trigger time. We first plot the distribution of trigger pulse amplitude times to identify the mean time; we only accept pulses for the DiCT analysis with amplitudes that reconstruct within 20 ns of the mean trigger pulse time. This requirement is made to reduce the bias on the DiCT measurement from DeCT pulses occurring very shortly after the primary pulse, in which case the MA filter and pulse finder will merge the pulses. We performed a test in which we: obtained a single 1 PE waveform at $V_{bias} = 32$ V; duplicated the waveform about the pulse and superimposed it on the original pulse (at $t_p = 0$); shifted it in time to study the change in the pulse shape. This is illustrated in figure 11 (right), which shows that for this particular waveform, two pulses separated by ≤ 80 ns will merge and reconstruct with an amplitude surpassing the 1.5 PE threshold.

Figure 12 shows P_{DiCT} as a function of ΔV . As expected, P_{DiCT} increases as a function of ΔV , increasing by 49% across 6.3 V $< \Delta V < 8.3$ V. At 77 K, the DiCT probability is reported to be 13% at $\Delta V = 5$ V for the NUV-HD-cryo technology [12], compared to our measurement of 15.8% at $\Delta V = 6.3$ V. Our measurement is approximately 20% higher, however we are also operating at ΔV about 1 V higher; it is plausible that at $\Delta V = 5$ V, our measurement would be consistent with the value reported in [12]. Thus, based on the single device we have tested, there is no clear temperature dependence on P_{DiCT} for the NUV-HD-cryo technology.

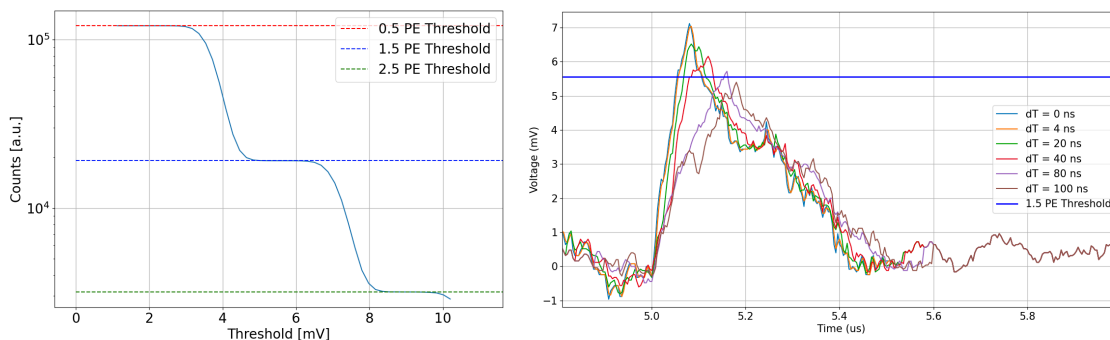


Figure 11. Left: pulse count rate as a function of threshold for $V_{bias} = 32$ V ($\Delta V = 6.3$ V). As the threshold is increased, the count rate decreases in a step-like fashion, leading to a characteristic “staircase”. Right: study to explore the effect on pulse reconstruction for delayed avalanches closely following a primary, obtained by selecting a 1 PE waveform $V_{bias} = 32$ V ($\Delta V = 6.3$ V), duplicating and superimposing the waveform onto the original, and shifting it in time to see the time difference at which the reconstructed amplitude drops below the DiCT threshold of 1.5 PE.

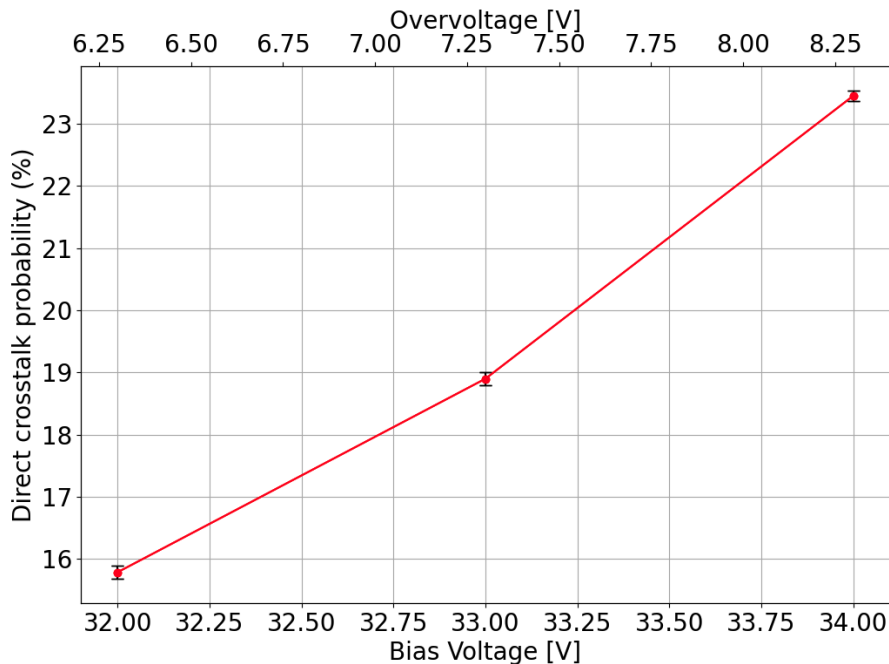


Figure 12. P_{DiCT} as a function of V_{bias} (ΔV). P_{DiCT} increases by 49% between $6.3 \text{ V} < \Delta V < 8.3 \text{ V}$. Solid red line is interpolated between the data points. Statistical error bars are shown, however are not necessarily visible as the errors are small.

3.6 Correlated delayed avalanches

Correlated delayed avalanches (CDAs) are a type of correlated noise in SiPMs, comprised of DeCT and afterpulses (APs). DeCT events occur within very short timescales, typically $\mathcal{O}(10 \text{ ns})$ after the primary avalanche, whereas AP events tend to occur on longer timescales, $\mathcal{O}(\text{ns}-\mu\text{s})$.

Afterpulsing is a process in which charge carriers generated in an avalanche can become trapped within defects of the silicon structure. Trapped charge carriers are then subsequently released, which can trigger avalanches in the same microcell after the primary avalanche has occurred. A single primary avalanche can result in multiple APs as a result of numerous trapped charge carriers being released at different timescales, or due to an AP triggering a second AP. If the timescale of the AP is short compared to the characteristic recharge time of the SiPM, the gain of the AP-induced avalanche is smaller than the primary avalanche, resulting in a pulse amplitude less than the 1 PE amplitude. If the timescale of the AP is long compared to the characteristic recharge time of the SiPM, APs are indistinguishable from 1 PE signals induced either thermally (from DCR) or optically (from a photon).

Similarly to DiCT, APs artificially increase the number of detected photons. APs can also distort integration measurements for high photon fluxes, effectively serving as an additional gain factor for the measured signal. The AP probability is expected to increase with ΔV [12], since a higher ΔV results in a larger initial avalanche charge. This, combined with an increase in the probability of an avalanche, results in a higher AP probability.

We determine the mean number of CDAs, \bar{N}_{CDA} , within three different time windows following a primary avalanche: $1 \mu\text{s}$, $5 \mu\text{s}$, and $10 \mu\text{s}$. The mean number of CDAs is calculated on a waveform-by-waveform basis. For each waveform, we count how many subsequent pulses are found within

a 1 μs , 5 μs , and 10 μs window after the primary pulse (i.e., the first pulse in the waveform that caused the trigger to occur). We apply no requirement on the amplitude of pulses found $< 1 \mu\text{s}$ with respect to the primary pulse, since the gain is decreased for APs occurring within the SiPM recharge time. This can be seen in figure 13 (left), which shows the 2D distribution of secondary pulse amplitudes versus dt for $V_{\text{bias}} = 32 \text{ V}$ ($\Delta V = 6.3 \text{ V}$).

However, we do apply a ToT requirement on pulses detected $< 1 \mu\text{s}$ from the primary pulse to reduce contamination from noise fluctuations occurring on pulse tails. Figure 13 (right) shows the ToT distribution for pulses occurring at $dt \leq 500 \text{ ns}$, $500 \text{ ns} < dt \leq 1 \mu\text{s}$, and $dt > 1 \mu\text{s}$ with respect to the primary pulse for $V_{\text{bias}} = 32 \text{ V}$ ($\Delta V = 6.3 \text{ V}$). There is an increase in identified pulses for dt values less than and on the order of the average length of a single photon pulse ($\sim 500 \text{ ns}$), a population which decreases with larger dt . We decide to exclude pulses from the CDA analysis with ToT $< 80 \text{ ns}$ for $dt \leq 1 \mu\text{s}$; this aligns with the second peak visible in the blue distribution of figure 13, above which we believe pulses are dominated by real CDAs rather than noise from tail fluctuations.

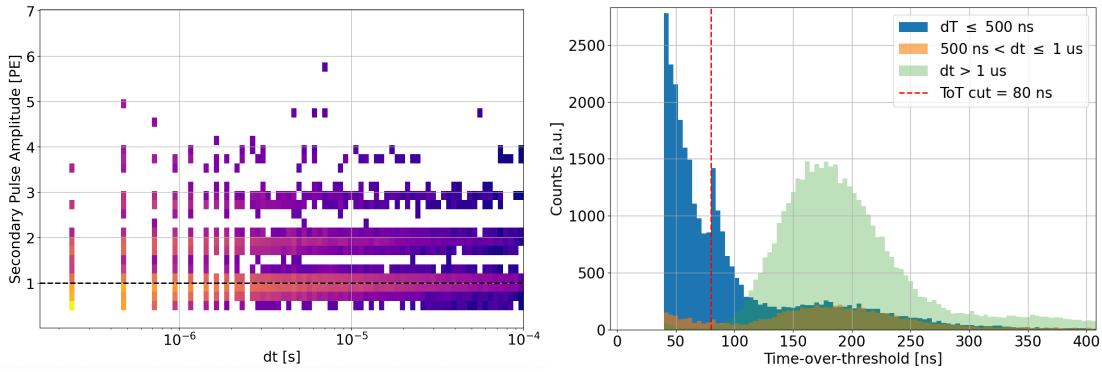


Figure 13. Left: 2D distribution of secondary pulse amplitudes versus dt for $V_{\text{bias}} = 32 \text{ V}$ ($\Delta V = 6.3 \text{ V}$), showing recharge-limited APs occurring within $dt < 1 \mu\text{s}$ of a primary pulse. Right: ToT distribution for pulses occurring at $dt \leq 500 \text{ ns}$ (blue), $500 \text{ ns} < dt \leq 1 \mu\text{s}$ (orange), and $dt > 1 \mu\text{s}$ (green) with respect to a primary pulse for $V_{\text{bias}} = 32 \text{ V}$ ($\Delta V = 6.3 \text{ V}$). Red dashed line corresponds to the ToT value above which we accept pulses (with $dt \leq 1 \mu\text{s}$) for the CDA analysis.

For pulses detected $\geq 1 \mu\text{s}$ from the primary pulse, we apply a requirement on the pulse amplitude to be $\geq \mu_{\text{SPE,amp}} - 3\sigma_{\text{SPE,amp}}$, to reduce leakage from noise. We make one additional requirement, that the time difference between the primary pulse of a waveform p_i and the last pulse of the preceding waveform p_{i-1} is $\geq 1 \text{ s}$. This cut is made to ensure that the primary pulse originates from DCR or from a photon-induced PE, and is not a CDA of a previous pulse. The choice of 1 s is informed from the unshaded time distribution in figure 9, which shows the asymptotic rate from pure DCR at time differences $\geq 1 \text{ s}$.

Figure 14 shows the CDA probability, P_{CDA} , as a function of ΔV within a 1 μs (red), 5 μs (blue), and 10 μs (green) window following a primary avalanche³. The CDA probability is measured to be $P_{\text{CDA}} = 15.7\% \pm 0.2\%$ (1 μs window), $34.0\% \pm 0.3\%$ (5 μs window), and $42.3\% \pm 0.3\%$ (10 μs window) at $\Delta V = 6.3 \text{ V}$. P_{CDA} increases by 59% (1 μs window), 80% (5 μs window), and 79% (10 μs window) between $\Delta V = 6.3 \text{ V}$ and $\Delta V = 8.3 \text{ V}$. However, we do note that as our pulse finder is unable

³The CDA probability is calculated using the mean number of correlated delayed avalanches with the following equation:

$$P_{\text{CDA}} = 1 - e^{-\bar{N}_{\text{CDA}}}$$

to resolve pulses that occur (on average) less than approximately 200 ns (Figure 3), there could be a systematic underestimation of P_{CDA} in our results. P_{CDA} increases as a function of ΔV as expected, driven by the increase in AP probability for higher ΔV . We also observe an increase in P_{CDA} for longer time windows following a primary avalanche for a given ΔV value.

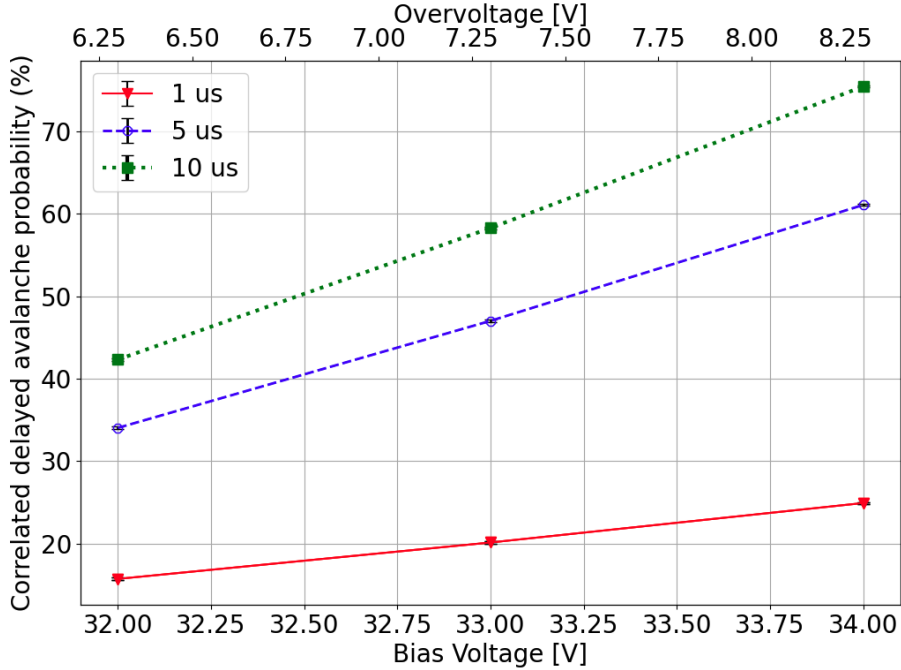


Figure 14. Correlated delayed avalanche probability (P_{CDA}) as a function of ΔV considering three time windows following a primary avalanche: 1 μs (red), 5 μs (blue), and 10 μs (green). P_{CDA} increases by 59% (1 μs window), 80% (5 μs window), and 79% (10 μs window) between $\Delta V = 6.3$ V and $\Delta V = 8.3$ V. Solid red, dashed blue, and dotted green lines are interpolated between the data points. Statistical error bars are shown, however are not necessarily visible as the errors are small.

3.6.1 Enhanced afterpulsing at low temperatures

At 77 K, the AP probability is reported to be 12% at $\Delta V = 5$ V [12]. We determine a CDA probability of 42% at $\Delta V = 6.3$ V assuming a 10 μs integration window, as is typically considered for computing AP probabilities in SiPMs [23, 24]. We assume that the number of DeCT events contributing to our P_{CDA} measurement is negligible; we make this assumption based on the timescale of DeCT events, which occur at $dt < 100$ ns, which we do not have the ability to resolve from the primary avalanche in our data. Therefore, we assume that our P_{CDA} measurement is dominated by APs.

We observe a substantial increase in the AP probability at $T_{\text{CSNT}} = 9.4 \pm 0.2$ mK compared to the measured value at 77 K. This behaviour has previously been reported at temperatures below 40 K [5, 6]. It has also been observed that below 40 K, the average time difference between a primary avalanche and AP(s) increases. We also see evidence for an increase in the delay times of APs with respect to the primary avalanche. The dt distribution shown in figure 9 shows long tails that extend significantly beyond the typical AP timescale of 10 μs . This is indicative of delayed APs that continue to occur even at much longer timescales.

The combination of a higher AP probability and an increase in the average time delay of APs can result in long-lasting AP “trains”, in which APs trigger further APs, causing a self-sustaining cascade of APs that can last for a long time at ultra-low temperatures, up to approximately 1 ms [5]. This effect is driven by the temperature-dependent trapping and release of charge carriers. During an avalanche, a fraction of carriers can be captured by deep-level traps in the silicon bulk and later released with a characteristic time. As the temperature decreases, the thermal emission rate from these traps is strongly suppressed, leading to much longer de-trapping times. When released, these carriers can retrigger avalanches, producing APs. At sufficiently low temperatures, a single primary avalanche can populate multiple traps with a range of lifetimes, resulting in a sequence of delayed releases and thus a correlated train of APs rather than a single event. This mechanism accounts for both the increased AP probability and the emergence of long AP trains at ultra-low temperature operation.

Figure 15 shows an example waveform trace acquired at $V_{\text{bias}} = 34 \text{ V}$ ($\Delta V = 8.3 \text{ V}$), consistent with a self-sustaining AP train. The presence of such AP trains clearly limits the photon counting ability of the SiPM, particularly at higher ΔV values. Whilst delayed APs will also distort integration measurements, figure 14 shows that the level of distortion caused by delayed APs can be mitigated by using shorter integration windows and operating at a lower ΔV .

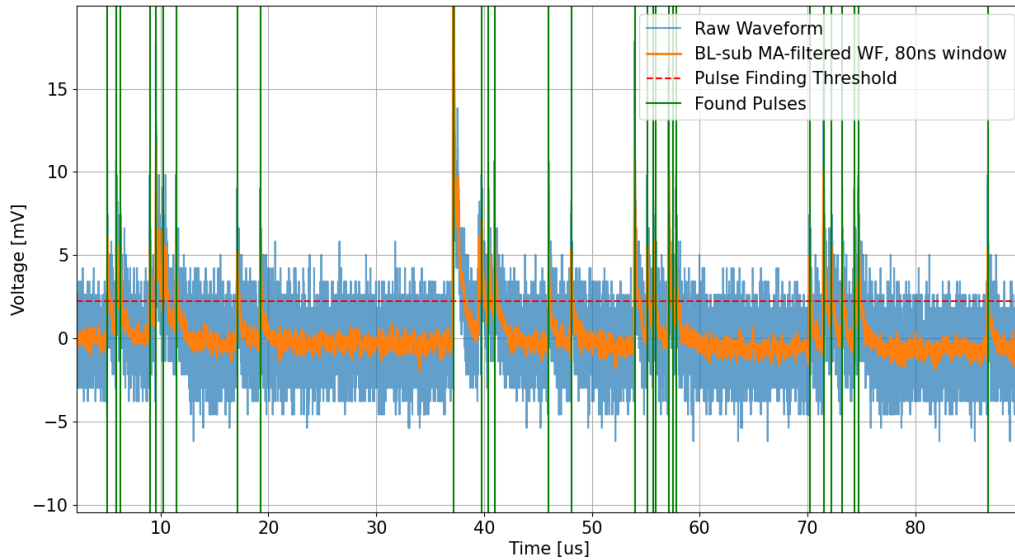


Figure 15. Example waveform trace acquired at $V_{\text{bias}} = 34 \text{ V}$ ($\Delta V = 8.3 \text{ V}$), illustrating long-lasting APs consistent with a self-sustaining AP train. Trace shows raw waveform (blue), baseline-subtracted moving-average filtered waveform (orange), pulse finding threshold (dashed red), and identified pulse times (green).

4 Towards a cryogenic cosmic-ray muon veto system

The main motivation for testing the NUV-HD-cryo SiPM technology in an ultra-low temperature environment is to determine whether it is feasible for SiPMs to be deployed as photon sensors for the cryogenic cosmic-ray muon veto system currently being designed for the QUEST-DMC experiment. As outlined in section 1, the concept design makes use of a scintillator volume surrounding the bolometer box holding the superfluid ^3He target, coupled to a SiPM. High energy cosmic-ray muons are minimum ionising particles which deposit approximately 2 MeV/g/cm^2 of energy in materials

they traverse. If a cosmic-ray muon passes through QUEST-DMC, it will also deposit a fraction of its energy into the scintillator surrounding the experiment, producing a flux of scintillation photons. These can be used as a veto signal, to remove any fiducial volume events from the dark matter search that occur in coincidence with a scintillator signal consistent with a cosmic-ray muon. If the SiPM is directly coupled to the scintillator (i.e., without using fibres to optically couple the physically separated scintillator and SiPM), it will need to be capable of operating at the same temperature as the scintillator, which is thermally coupled to a thermal radiation shield at a temperature of $\mathcal{O}(10 \text{ s mK})$.

The requirements of the cryogenic cosmic-ray muon veto system are set by both its primary function of tagging muon-induced events, and the constraints of operation near the main experimental volume at mK temperatures. The system must provide a sufficiently large response to muon energy deposits to allow a practical threshold to be defined in the presence of background interactions from environmental gammas, targeting a veto efficiency of $\geq 90\%$ while keeping accidental vetoes — and hence experimental dead time — minimal. A detailed characterisation of the SiPM noise is essential, as it sets the minimum achievable trigger threshold. In particular, correlated noise (DiCT, DeCT, APs) can increase the apparent charge of background signals and raise the accidental trigger rate, limiting how low the threshold can be set. In addition, the SiPM must operate stably at mK temperatures, meeting constraints on power dissipation and thermal load of the dilution fridge, and remaining compatible with the low-noise, low-background environment of the experiment.

In addition to the main dataset used for the characterisation analysis, we also acquired a second, shorter dataset where a small volume of scintillator was coupled to the SiPM. We acquired 2500 waveforms per voltage under the same trigger conditions as before, with a shorter $10 \mu\text{s}$ acquisition window ($1 \mu\text{s}$ pre-trigger, $9 \mu\text{s}$ post-trigger). The plastic scintillator used for this study was an early prototype for the SoLid experiment [25], and is reported to have a scintillation efficiency of approximately 10000 photons/MeV e^- at room temperature. This was a convenient test sample for the purpose of these initial proof-of-concept measurements.

Three small blocks of the scintillator, each with dimensions of 1 cm (W) \times 3 cm (L) \times 0.5 cm (D), are attached to the inside of a spare copper box lid using epoxy adhesive. The scintillator blocks are each embedded with three wavelength shifting fibres [26], which shift blue light to green with absorption at 420 nm and emission peaking at $\sim 494 \text{ nm}$ [25], and guide the photons produced in the scintillator towards the face of the SiPM. Figure 16 shows a picture of the scintillator blocks attached to the inside of the copper box lid. A cross-sectional diagram showing the integration of the scintillator inside the copper box when the lid is attached is shown in figure 16 (right). The SiPM is not physically in contact with the scintillator blocks when the box is closed; there is an extremely small gap of 1–2 mm between the scintillator blocks and the SiPM face when the box is closed. The SiPM sits in the central area of the three scintillator blocks, for maximum coverage and light collection efficiency.

Figure 17 (left) shows an example high-light event observed in this dataset, acquired at $V_{\text{bias}} = 32 \text{ V}$ ($\Delta V = 6.3 \text{ V}$). In this event, the number of photons incident on the SiPM is so large that it surpasses the dynamic range of the oscilloscope, illustrated by the clipping of the signal at 140 mV. For high photon fluxes such as this, we also observe a lengthening of the signal tail, compared to the SiPM recharge time ($< 1 \mu\text{s}$). This is caused by the superposition of pulses from real photons and correlated noise, which contribute to an overall increase in the duration of the signal.

Figure 17 (right) compares the pulse area distribution, computed as the $10 \mu\text{s}$ integral about the event trigger, for data acquired with the scintillator (red) versus data acquired without the scintillator

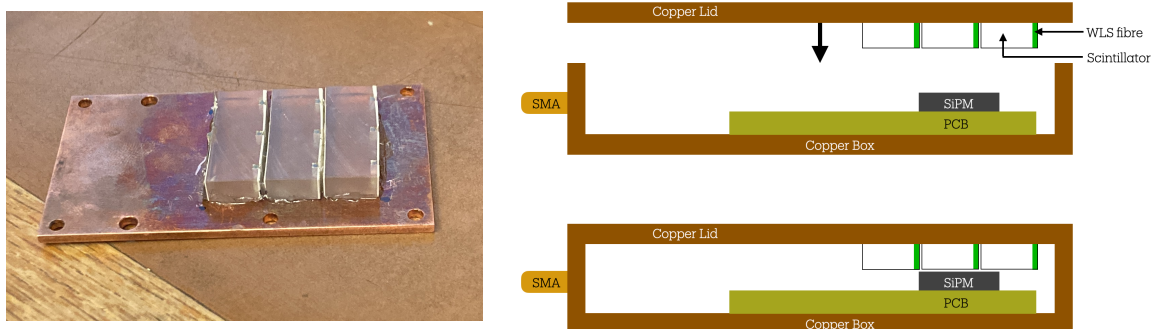


Figure 16. Left: image showing the three scintillator blocks attached to the inside of the spare copper box lid. Each block is embedded with three optical fibres, which guide photons generated in the scintillator towards the SiPM surface. Right: cross-sectional diagram showing integration of scintillator blocks inside copper box when lid is attached and box is closed.

(blue). Without scintillator, we only expect to observe signals associated with the SiPM’s intrinsic noise, resulting in a narrow distribution peaked at the single photon integral. However, once scintillator is introduced, we observe a much broader distribution and a high charge tail, corresponding to energy depositions in the scintillator from particle interactions. This distribution is likely a combination of environmental gamma-rays, which dominate the lower part of the charge distribution, and high-energy cosmic-ray muons populating the high charge tail, as observed in the cryogenic cosmic-ray muon veto system developed by the NUCLEUS experiment [11]. However, this has not been fully characterised here. This data demonstrates that the SiPM is clearly detecting light from the scintillator and optical fibres, which is easily distinguishable above the SiPM noise sources.

As discussed in section 3.6.1, the SiPM exhibits an increase in noise from long-lasting APs when operated at ultra-low temperatures. Whilst this additional noise source clearly limits the photon counting capability of the SiPM for a low photon flux, this is less of a concern for high photon fluxes generated from cosmic-ray muons. An increase in long-lasting APs will manifest as an increase to the integrated charge of a signal, but only in response to a primary energy deposition in the scintillator. As such, this should not limit the ability of the SiPM to tag cosmic-ray muons through their initial energy deposition. This is a first proof-of-concept of using scintillator coupled to a NUV-HD-cryo SiPM in a mK environment as a means of identifying high-light events consistent with through-going charged particles, such as from cosmic-ray muons.

5 Conclusions

We have shown that a FBK NUV-HD-cryo SiPM is capable of operating successfully inside a cryogen-free dilution refrigerator down to a base temperature of 9.4 ± 0.2 mK. No measurable increase in the dilution refrigerator base temperature was observed during SiPM operation, indicating that the associated power dissipation is compatible with the available thermal budget. This confirms that the device satisfies the cryogenic cosmic-ray muon veto system requirement for operation directly at the mK stage, coupled to the scintillator, without compromising the fridge performance. We measure the breakdown voltage and gain of the SiPM, and characterise several noise properties including

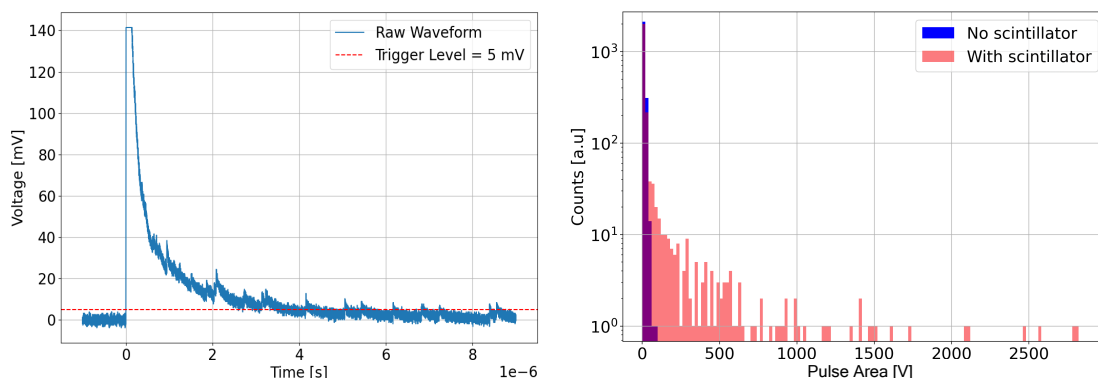


Figure 17. Left: example waveform trace taken at $V_{\text{bias}} = 32 \text{ V}$ ($\Delta V = 6.3 \text{ V}$), illustrating a high-light event consistent with through-going charged particles, potentially from a cosmic-ray muon. Right: pulse area distribution (integrated over a $10 \mu\text{s}$ window) about the event trigger at $V_{\text{bias}} = 32 \text{ V}$ ($\Delta V = 6.3 \text{ V}$) when SiPM is not coupled to scintillator (blue) versus when SiPM is coupled to scintillator (red).

the dark count rate, direct crosstalk probability, and correlated delayed avalanche probability, which we predominately attribute to afterpulsing.

We also present early measurements of deploying the SiPM coupled to scintillator with optical fibre assembly inside the dilution fridge and operated at $9.4 \pm 0.2 \text{ mK}$, as proof-of-concept for a cryogenic cosmic-ray muon veto system to be operated internal to a dilution refrigerator. In this run, we observe a significantly broader charge spectrum, consistent with high-light events originating from energy depositions from through-going charged particles, likely from environmental gamma-rays and high-energy cosmic-ray muons, which are not present in the data acquired without the scintillator.

The SiPM used in this work originates from an engineering run used for pre-production assembly for the DarkSide-20k experiment and was developed to meet the stringent requirements of single photon detection in a near background-free environment. In contrast, the scintillator-based cryogenic cosmic-ray muon veto system we propose for the QUEST-DMC experiment will operate in a regime of much larger signals, since muons traversing the scintillator will deposit significant energy and produce many photons. The requirements on the noise levels are therefore less stringent than the expected values outlined in [12], provided that muon signals remain sufficiently larger than background contributions from environmental gamma interactions and intrinsic SiPM noise. The system must be able to achieve at least a 90% muon veto efficiency, whilst minimising the experimental dead time due to accidental vetoes.

The dark count rate and direct crosstalk probability we measure in this work appear to be roughly consistent with the expected performance at 77 K for similar ΔV values [12], which would satisfy the noise requirements for the cosmic-ray muon veto system. These results indicate a weak temperature dependence of these properties between 77 K and 9.4 mK, however this would need to be confirmed with additional measurements at more temperature points across multiple devices. We do however observe a significant increase in the afterpulsing probability, measuring 42% at $\Delta V = 6.3 \text{ V}$ assuming a $10 \mu\text{s}$ integration window, compared to an expected afterpulsing probability of 12% at $\Delta V = 5 \text{ V}$ at 77 K [12]. In addition, we observe evidence of long-lasting afterpulse trains which can last for up to approximately 1 ms, a phenomena which has previously been observed in SiPMs at temperatures below 40 K [5, 6].

While this additional noise does not prevent operation of the veto system, it introduces event-by-event fluctuations in the reconstructed charge and can produce high-charge tails. This increases the probability that background events from environmental gamma interactions exceed the veto threshold, leading to a higher rate of accidental vetoes and hence increased dead time, unless a higher threshold is applied. However, raising the threshold may compromise the ability of the system to achieve the target veto efficiency of 90%. To minimise this excess noise, it will be important to ensure that the scintillation signal is sufficiently prompt, such that the primary light is collected within a short time window. This consideration will inform the next-stage of R&D of the cosmic-ray muon veto system, in particular in the choice of scintillator material, and in optimisation of the geometry and light collection (including the number of required SiPMs and the scintillator-SiPM coupling), to achieve a fast and well-defined muon signal. Following this, we will characterise the cryogenic cosmic-ray muon veto system as a function of temperature, from room temperature down to mK temperatures, in order to determine the minimum veto threshold required to achieve the target performance for the QUEST-DMC experiment.

Acknowledgments

This work was funded by UKRI EPSRC and STFC (Grants ST/T006773/1, ST/Y004434/1, EP/P024203/1, EP/W015730/1 and EP/W028417/1), as well as the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement no 824109 (European Microkelvin Platform). S.A. acknowledges financial support from the Jenny and Antti Wihuri Foundation. M.D.T. acknowledges financial support from the Royal Academy of Engineering (RF/201819/18/2). J.Sm. acknowledges support from the U.K. Research and Innovation Future Leader Fellowship MR/Y018656/1. A.K. acknowledges support from the U.K. Research and Innovation Future Leader Fellowship MR/Y019032/1.

References

- [1] DEAP collaboration, *In-situ characterization of the Hamamatsu R5912-HQE photomultiplier tubes used in the DEAP-3600 experiment*, *Nucl. Instrum. Meth. A* **922** (2019) 373 [[arXiv:1705.10183](#)].
- [2] LZ collaboration, *The LUX-ZEPLIN (LZ) Experiment*, *Nucl. Instrum. Meth. A* **953** (2020) 163047 [[arXiv:1910.09124](#)].
- [3] S. España et al., *Performance Evaluation of SiPM Photosensors in the Presence of Magnetic Fields*, *AIP Conf. Proc.* **1231** (2010) 171.
- [4] M. Biroth, P. Achenbach, E. Downie and A. Thomas, *Silicon photomultiplier properties at cryogenic temperatures*, *Nucl. Instrum. Meth. A* **787** (2015) 68.
- [5] O. Hanski et al., *Performance of Silicon photomultipliers at low temperatures*, [arXiv:2501.03812](#).
- [6] J. Zhang et al., *Scintillation detectors with silicon photomultiplier readout in a dilution refrigerator at temperatures down to 0.2 K*, *2022 JINST* **17** P06024 [[arXiv:2203.15631](#)].
- [7] QUEST-DMC collaboration, *QUEST-DMC superfluid ^3He detector for sub-GeV dark matter*, *Eur. Phys. J. C* **84** (2024) 248 [[arXiv:2310.11304](#)].
- [8] QUEST-DMC collaboration, *Dark matter attenuation effects: sensitivity ceilings for spin-dependent and spin-independent interactions*, *JCAP* **04** (2025) 017 [[arXiv:2502.10251](#)].
- [9] QUEST-DMC collaboration, *Dark Matter EFT landscape probed by QUEST-DMC*, *JCAP* **10** (2025) 044 [[arXiv:2505.17995](#)].

- [10] S. Autti et al., *QUEST-DMC: Background Modelling and Resulting Heat Deposit for a Superfluid Helium-3 Bolometer*, *J. Low Temp. Phys.* **215** (2024) 465 [[arXiv:2402.00181](https://arxiv.org/abs/2402.00181)].
- [11] A. Erhart et al., *A plastic scintillation muon veto for sub-Kelvin temperatures*, *Eur. Phys. J. C* **84** (2024) 70 [[arXiv:2310.08457](https://arxiv.org/abs/2310.08457)].
- [12] A. Gola et al., *NUV-Sensitive Silicon Photomultiplier Technologies Developed at Fondazione Bruno Kessler*, *Sensors* **19** (2019) 308.
- [13] A. Shibahara et al., *Primary current-sensing noise thermometry in the millikelvin regime*, *Phil. Trans. A. Math. Phys. Eng. Sci.* **374** (2016) 20150054.
- [14] J. Nyéki et al., *High-Performance Cryogen-Free Platform for Microkelvin-Range Refrigeration*, *Phys. Rev. Appl.* **18** (2022) L041002 [[arXiv:2209.07903](https://arxiv.org/abs/2209.07903)].
- [15] M.C. Runyan and W.C. Jones, *Thermal Conductivity of Thermally-Isolating Polymeric and Composite Structural Support Materials Between 0.3 and 4 K*, *Cryogenics* **48** (2008) 448 [[arXiv:0806.1921](https://arxiv.org/abs/0806.1921)].
- [16] Tektronix, *DPO72304DX Oscilloscope Datasheet*, https://www.rftesolutions.com/pdf/oscilloscopes/DP072304DX--Data_Sheet.pdf.
- [17] Tektronix, *Using FastFrame™ Segmented Memory*, https://download.tek.com/document/55W_12112_2.pdf.
- [18] F. Acerbi et al., *NUV and VUV sensitive Silicon Photomultipliers technologies optimized for operation at cryogenic temperatures*, *Nucl. Instrum. Meth. A* **1046** (2023) 167683.
- [19] P. Achenbach, M. Biroth, W. Lauth and A. Thomas, *Gain stabilization and noise minimization for SiPMs at cryogenic temperatures*, *Nucl. Instrum. Meth. A* **912** (2018) 110.
- [20] DARKSIDE-20K collaboration, *Quality assurance and quality control of the 26 m² SiPM production for the DarkSide-20k dark matter experiment*, *Eur. Phys. J. C* **85** (2025) 534 [[arXiv:2412.18867](https://arxiv.org/abs/2412.18867)].
- [21] A. Butcher et al., *A method for characterizing after-pulsing and dark noise of PMTs and SiPMs*, *Nucl. Instrum. Meth. A* **875** (2017) 87 [[arXiv:1703.06204](https://arxiv.org/abs/1703.06204)].
- [22] R. Mirzoyan, R. Kosyra and H.-G. Moser, *Light emission in Si avalanches*, *Nucl. Instrum. Meth. A* **610** (2009) 98.
- [23] F. Acerbi et al., *NUV Silicon Photomultipliers With High Detection Efficiency and Reduced Delayed Correlated-Noise*, *IEEE Trans. Nucl. Sci.* **62** (2015) 1318.
- [24] F. Acerbi and S. Gundacker, *Understanding and simulating SiPMs*, *Nucl. Instrum. Meth. A* **926** (2019) 16.
- [25] SoLID collaboration, *Optimisation of the scintillation light collection and uniformity for the SoLid experiment*, *2018 JINST* **13** P09005 [[arXiv:1806.02461](https://arxiv.org/abs/1806.02461)].
- [26] Lunix Solutions, *Plastic scintillating fibers*, <https://luxiumsolutions.com/sites/default/files/2021-11/Fiber-Product-Sheet.pdf>.