TF2 report

Simone Scopello,[1] Paolo Musico,[1] Samuel Hill,[2] Zoe Balmforth,[2] Daria Santone,[2] Valerio Ippolito,[3] Marco Rescigno,[3] Giacomo Petrillo,[4] Simone Stracka,[4] Eugenio Paoloni,[4] Luigi Rignanese,[5] David-Michael Poehlmann,[6] Luca Pagani,[6] Emilija Pantic,[6] Claudio Savarese,[7] Shawn Westerdale,[7] Yi Wang,[8] Tom Thorpe,[8] Hanguo Wang,[8] Vicente Pesudo,[9] Victor Goicoechea Casanueva,[10] Alexander Kish,[10] Pierre-Andre Amaudruz,[11] Ashlea Amanda Kemp,[12] Mark Stringer,[12] Emmanuel Le Guirriec,[13] Julie Rode,[13] Davide Franco,[13] and Paolo Agnes[14]

[1] *INFN Genova*
[2] *RHUL*
[3] *INFN Roma1*
[4] *INFN Pisa*
[5] *INFN Bologna*
[6] *UC Davis*
[7] *Princeton U*
[8] *UCLA*
[9] *CIEMAT*
[10] *U Hawaii*
[11] *TRIUMF*
[12] *Queens U*
[13] *IN2P3*
[14] *U Houston*

(Dated: April 26, 2021)

Contents

I. INTRODUCTION

A. Reduction of the number of TIAs

In the baseline configuration for the Tile and FEB of a PDM, its 24 SiPM are organized in 4 quadrants, each consisting of 6 SiPM arranged in 3 parallel branches of series of 2 SiPMs (2s3p configuration). The output of each quadrant is connected to the input of a transimpedance amplifier (TIA). The output of 4 TIAs is then added.

Other configurations are possible, in which one transimpedance amplifier receives the input of more than 6 SiPM, e.g., 24 SiPM (corresponding to the full Tile) or 12 SiPM (half the Tile), by wiring the SiPM differently.

A reduction in the number of components of the frontend board (FEB) affects heat dissipation and radiopurity.

B. Readout channel grouping

The reduction in the number of channels aims primarily to reduce the cost and complexity of the system.

When summing the output of several PDM together in the same readout channel, noise increases while the amplitude for the single photoelectron signal remains unchanged. This can be compensated by operating the PDMs at a higher overvoltage, thus increasing the signal amplitude (in the following, the quoted bias voltages refer to a Tile in the baseline configuration, in which the bias voltage is applied to a series of two SiPM; the quoted overvoltage is evaluated across a single SiPM)

II. HARDWARE

A. Summary

1. Reduction in number of TIA

- the impact in heat dissipation can be sizeable (Table I and Table II);
- the impact on radiopurity is marginal: O(10%) (Sec. II B 6);
- a possible worsening in signal-to-noise ratio with respect to the 4-TIA configuration (Sec. II B 3) must be taken into account when evaluating the compatibility of this choice with the reduction in the number of channels (the 4-TIA configuration is assumed throughout Sections II C 2 and IV).

2. Reduction in number of readout channels

- when performed before transmission, significant reduction in complexity is expected;
- significant cost savings possible, limited by bandwidth requirements in DAQ (Table VI);
- retaining similar performances for 1PE identification in digitizers and 1PE time resolution offline requires rising the operating overvoltage (Fig. 6 and Fig. 7). These studies consider, as a noise source, only the PDM electronics itself, in the 4-TIA configuration (i.e., noise resulting from transmission must be considered separately, e.g., as an extra safety factor). In proto-0, operating overvoltage was $5.5 - 6.5\,\mathrm{V}$;
- operating at higher overvoltage is associated to an increase in dark count rate and correlated noise probability (Fig. 2 and Fig. 3). At high correlated noise probability the response of the SiPM become difficult to model in simulation.
- the baseline layout (1 PDM/ch) offers the best signal quality and flexibility in operation (i.e., the possibility to adjust the overvoltage); grouping 2 PDMs in a single channel offers a good compromise between cost, complexity, and signal quality, while maintaining some flexibility during operations; grouping 4 PDMs or more in a single channels reduces complexity the most, and offers some additional cost reduction, at the price of higher correlated noise rates, worse signal quality, and lower margins for adjustments during operations.

B. Photoelectronics and transmission

1. LFoundry sensors performances

Traces from LFoundry tiles during laser runs have been collected at LNGS at different overvoltages [1]. An example of the average single electron responses for an LFoundry Tile is compared to that of a FBK Tile in Fig. 1.

Measurements of the dark count rate as a function of the overvoltage are reported in Fig. 2, while the behavior of direct cross-talk and afterpulsing is reported in Fig. 3.

For the optical cross-talk, only preliminary observations from ReD and at LNGS, not fully documented, are available.

The photodetection efficiency (PDE) of the sensors has not been measured at LAr temperatures. The values used in the simulations are based on educated guesses based on measurements at room temperature.

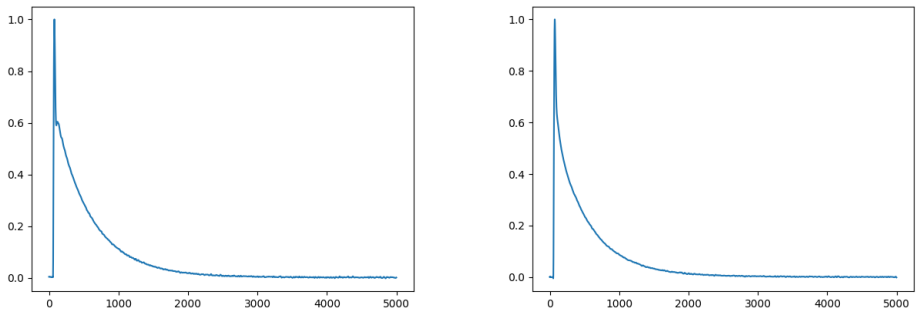

Figure 1: Comparison of the average single electron response for FBK (left) and LFoundry (right) sensors from LNGS laser runs, at 6 V overvoltage. Amplitude is in arbitrary units.

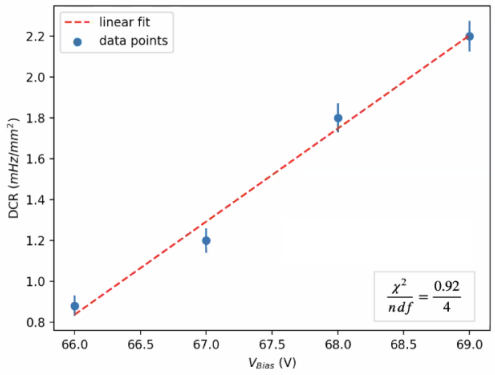

Figure 2: LNGS measurement of dark count rate as a function of bias voltage for a LFoundry tile. Technical Board meeting, 17 november 2020. From P. Kachru.

2. Operating overvoltage

Most of the measurements of the performances of the sensors (FBK or LFoundry) have been collected at overvoltages of 9 V or lower. Proto-0 sensors were operated at overvoltages of $5.5 - 6.5$ V.

It does not appear that an increase in sensor faults or ageing effects associated to operating SiPM's at a higher overvoltage has been observed and documented in the (lack of) existing literature.

The operational limit of the SiPMs can be set by the divergence of the correlated noise (correlated noise increases with the overvoltage). Using the geometric series approximation, when the probability P_{CN} of having a correlated noise event approaches one, the number of correlated noise events generated by a single, primary avalanche diverges. In many practical cases, SiPMs cannot be operated above $P_{CN} = 0.5$ [3, 4].

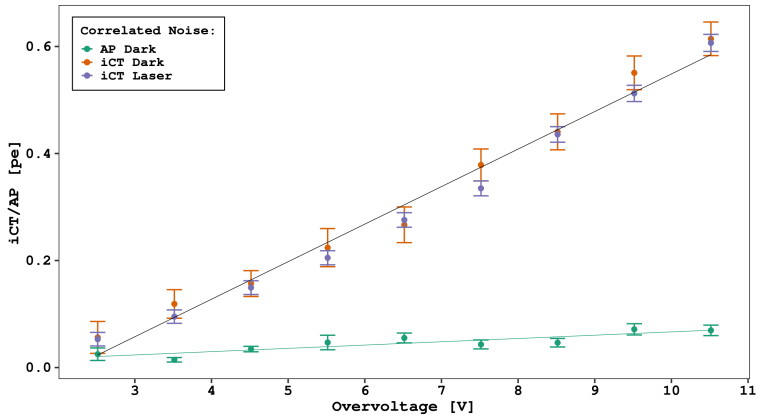

Figure 3: LNGS measurement of correlated noise as a function of overvoltage for a LFoundry tile. From A. Razeto.

3. FEB performances with a reduced number of TIA's

A reduction in the number of TIA's requires a redesign of the FEB and Tile circuitry. Also, the power supply choice differs between 4s and 2s configurations.

No measurement from a prototype FEB in the 1 TIA configuration was reported at TF2 meetings. It was therefore not possible to verify on data the absence of any change in signal-to-noise. Back-of-envelope calculations predict that the 1-TIA (4s6p), 2-TIA (2s6p or 4s3p), and 4-TIA (2s3p) configurations should have the same signal-to-noise ratio *at the PDM level*.

The signal amplitude at a given operating overvoltage in the 4s configuration (1-TIA design) is half of the amplitude in the 2s configurations (4-TIA design). This difference may be compensated by a gain in the adder circuit.

Since the channel grouping option already leads to a worsening in channel-to-noise ratio at a given overvoltage, the possibility that configurations with a reduced number of TIA's may result in a lower signal quality should be considered when evaluating the adoption of both solutions at the same time (i.e., reducing the number of TIAs while also grouping channels).

4. Transmission options

Recent reports on the status of transmission studies within the Photoelectronics group were presented here:

- optical transmission:
  - L. Rignanese - `https://agenda.infn.it/event/25940/contributions/131302/attachments/79551/103498/Opto-link_5min.pdf`
  - B. Bottino - `https://agenda.infn.it/event/26198/contributions/133083/attachments/80025/104314/VetoPE_15Mar2021.pdf`
- electrical transmission (flex cable):
  - P. Musico - `https://agenda.infn.it/event/25940/contributions/131301/attachments/80210/104682/CablesPaolo_210225.pdf`

A prototype for the optical link is available (transmitter and fiber), which shows good signal integrity. It was observed that the mechanical stability of the link is at this stage still very fragile. A prototype for the flex cable is not available, yet, but a design is expected by the end of March.

The compatibility of the two transmission options with the different cryostat designs has not been investigated. The transmission across the various interfaces (UAr-AAr, and LAr-RT) has not been discussed.

The above presentations include some preliminary cost estimates of 100 EUR/ch for the flex and $\sim$ 300 EUR/ch for the optical links. The reduction in the number of channels is expected to reduce the cost and complexity of both solutions (optical and copper).

5. Heat dissipation

Reducing the number of TIA amplifiers in the frontend board reduces the power dissipated by the photoelectronics. Table I reports the expected power dissipation in the FEB in different configurations. It should be noted that the FEB includes other active components in addition to the TIA's.

Reducing the number of components in the FEB also reduces the current drawn from the low-voltage power supply, and thus the power dissipated in the supply cables, as reported in Table II. The number of supply cables scales with the number of PDUs.

Finally, an optical driver (for the optical transmission option) or a differential driver (for the electrical transmission option) also contribute to heat dissipation. This contribution, reported in Table III, scales with the number of channels.

				Number of components per FEB			
FEB Component	Function	Power (mW)	I (mA) supply	4 TIA	2 TIA	1 TIA	1 TIA minimal
LMH 6629	TIA	58	15.5	4	2	1	1
OPA838	FEB adder	4.5	0.96	1	1	0	0
THS4521	FEB diff. driver	5.5	1.14	1	1	1	0
ADM7150	Linear regulator	24.1		1	1	1	0
P (1 FEB) (mW)				266	150	88	58
P (all FEBs) (kW)				2.20	1.24	0.73	0.48

Table I: Power dissipation (FEB only). Power dissipation in LV supply cables and optical or differential drivers are not included.

6. Radiopurity

Reducing the number of TIA amplifiers in the frontend board may improve the photoelectronics radiopurity.

The contribution to radioactivity of the individual components of the FEB and of Optolinks is reported here:

- V. Pesudo, `https://docs.google.com/presentation/d/15NB55f886_ug228wXxu44XiwvThhdAyti0bMxvI6kes/edit#slide=id.p`

It is expected that in the 1 TIA configuration the number of passive components in the FEB will be halved w.r.t. the 4 TIA configuration. As a result, the contribution of the FEB to the radioactivity

Configuration	4 TIA	2 TIA	1 TIA	1 TIA minimal
FEB supply current (mA)	64	33	17	16
PDU supply current (A)	1.60	0.83	0.42	0.39
Top cable resistance (Ω)			0.85	
Bottom cable resistance (Ω)			1.1	
Number of PDU			344	
Power top cables (kW)	0.38	0.10	0.03	0.02
Power bottom cables (kW)	0.48	0.13	0.04	0.03
Total (kW)	0.86	0.23	0.07	0.05

Table II: Power dissipated in LV supply cables, calculated according to typical supply current for each FEB active component: $P = 0.5 * I_{PDU}^2 * R_{cable} * N_{PDU}$. The estimate for the 4TIA case in doc-db:4310 assumes a supply current of 1.3 A instead of 1.6 A (a 20% difference).

Transmission option		Power (kW) per configuration			
Component	Power/ch (mW)	1 PDM/ch	2 PDM/ch	4 PDM/ch	5 PDM/ch
Optical driver (optical)	30	0.25	0.13	$0.07 - 0.09$	0.05
Differential driver (electrical)	22	0.18	0.10	$0.05 - 0.07$	0.04

Table III: Power dissipation due to optical or differential drivers (corresponding to the two transmission options, optical and electrical) as a funtion of the total number of channels.

budget will be most likely halved. The contribution from the optical module is expected to scale roughly linearly with the number of channels.

In the baseline configuration, the contribution of the FEBs to the radioactivity budget ("neutrons after cuts") represents roughly 40% of the Photoelectronics contribution, which is itself a fraction 31.6% of the total budget of the experiment (in "Plan A"). A similar picture can be derived for the calculation of γ activity (Table IV).

	n after cuts $[(200\,\mathrm{t\,yr})^{-1}]$	γ activity [Bq]
Baseline FEB TPC	2.25×10^{-2}	19.8
Baseline Optolink TPC	1.21×10^{-2}	11
Baseline Photoelectronics TPC	5.7×10^{-2}	76

Table IV: Contributions to radioactivity budget from FEB and Optolinks, compared to the total contribution from Photoelectronics, in the baseline configuration (4 TIA, 8280 channels).

7. Mechanical design

A reduction in the number of TIA's can be achieved without any modification of the mechanical design of the PDM, but requires a redesign of the Tile and FEB printed circuits.

Similarly, the grouping of the output of several PDM in the same readout channel could in principle be performed just before transmission, e.g., in the PCB hosting the optical driver, thus not requiring any significant modifications to the mechanical design of the PDU.

The specific implementation of the proposed modifications to the design of the sensors is beyond the scope of TF2 and is left to the Photoelectronics group.

C. DAQ

The DAQ discussion is limited to the TPC, due to the ongoing redesign of the VETO geometry.

1. Impact of reduction of number of TIA's

The reduction in the number of TIA's may have an impact on the signal-to-noise ratio, and thus on DAQ performances. However, a like-for-like comparison of a 1-TIA design prototype with the baseline 4-TIA design is not possible on the timescale of TF2. Therefore in the following it is assumed that 1-TIA (4s6p), 2-TIA (2s6p or 4s3p), and 4-TIA (2s3p) configurations behave according to back-of-envelope calculations, which predict that the signal-to-noise ratio *at the PDM level* does not depend on the number of TIA's. It should be noted, however, that the 4-TIA configuration was the result of several years of optimization efforts, and that some safety margin should be factored in when switching to a new configuration.

2. Impact of grouping several PDM's in the same readout channel

The signal-to-noise ratio (SNR) affects:

- the 1PE selection efficiency and the rate of accidental hits from noise in the boards after a filter suitable for implementation in the digitizer FPGA;
- the time resolution that can be achieved on single photoelectron pulses by means of a matched filter.

Grouping several PDM's in the same readout channel lowers the SNR. Since the signals from the PDM's contributing to the same readout channel are summed after at least an amplification stage in the FEB, it can be assumed that *at any given overvoltage* the signal-to-noise ratio with N PDM's in a single readout channel behaves as $SNR(N) \sim SNR(1)/\sqrt{N}$. We note that SNR improves linearly with overvoltage.

For 1PE identification in the digitizer boards we rely on a running average filter, $y(i) = \sum_{k=0}^{M-1} x(i-k)$. We choose $M = 128$ and assume the quantity $f(i) = y(i - 2M) - y(i)$ is compared to a threshold to identify 1PE pulses (here the difference accounts for baseline subtraction). This procedure has been characterized on waveforms from a single LFoundry tile collected at LNGS in laser runs at different overvoltages, as well as on eight FBK tiles (to get a more statistically meaningful sample), as shown in Fig. 4. We evaluate $f(\cdot)$ at a sample $i_0 = i_{\text{laser}} + M$, where i_{laser} is the sample synchronous with the laser pulse. The signal-to-noise ratio SNR is here defined as the mean of the $f(i_0)$ values obtained from waveforms containing a single photoelectron associated to the laser pulse, divided by the standard deviation of $f(i_0)$ values obtained in waveforms not containing photoelectrons associated to the laser pulse:

$$SNR = \frac{< f(i_0) >_{1PE}}{\sqrt{[< f(i_0)^2 > - < f(i_0) >^2]_{0PE}}}.$$

By requiring a 1PE identification efficiency better than 99%, a noise rate lower than 10 Hz (see Fig. 5), and accounting for a 5% variability (as in the FBK population) in the signal-to-noise ratio across different PDMs (while neglecting every other effect), we estimate the minimum operating overvoltage when

grouping N PDM's together to be

$$OV(N) = 4.5\,\mathrm{V} \times \sqrt{N}.$$

The rates calculated at the working point corresponding to 99% 1PE efficiency are shown in Fig. 6.

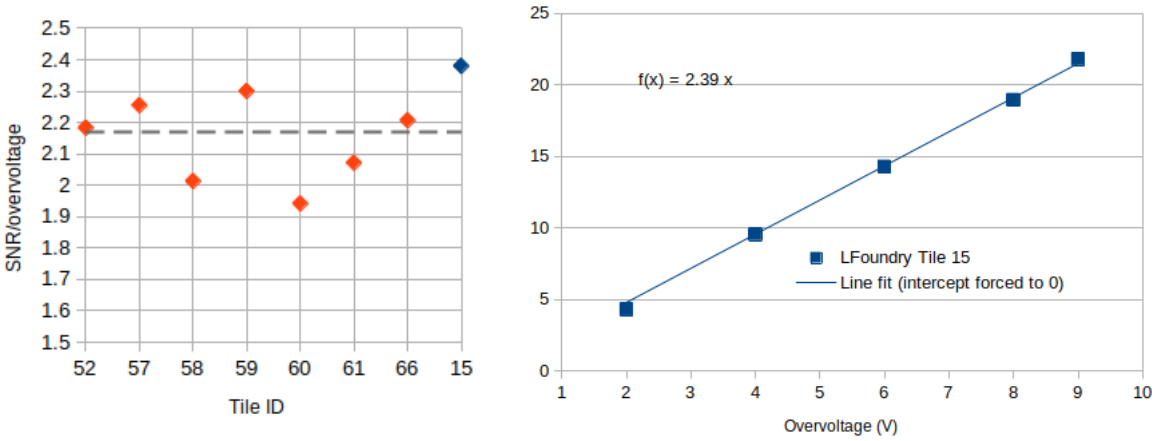

Figure 4: Left: signal-to-noise ratio per overvoltage after applying the 1PE selection algorithm described in Sec. II C 2, for LFoundry (blue) and FBK tiles (orange). The horizontal dashed line reports the mean. Right: signal-to-noise ratio as a function of overvoltage for LFoundry Tile 15, with a line fit overlaid (the intercept is forced to zero).
In both figures, SNR is defined as $SNR =< f(i_0) >_{1PE} / \sqrt{[< f(i_0)^2 > - < f(i_0) >^2]_{0PE}}$ (see text for the definition of $f(i_0)$).

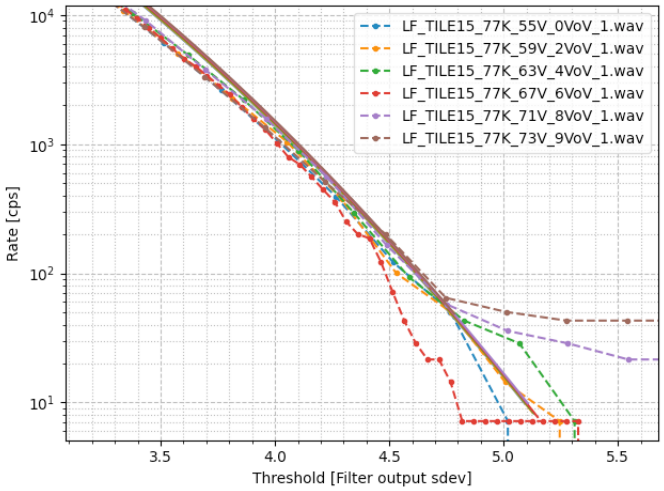

Figure 5: Rate due to noise as a function of the threshold (in units of the rms of the pedestal after filtering) obtained in the pretrigger region of laser runs with an LFoundry tile collected at LNGS. Dashed line: data; Solid line: theory. Departure from theory at higher thresholds for higher overvoltage may be due to dark count rate or light leaks in the measurement apparatus. From G. Petrillo.

To evaluate the effect of a lower SNR on time resolution, we study the same laser runs, collected by varying the operating overvoltage, and compare the reconstructed photon arrival time (by applying the

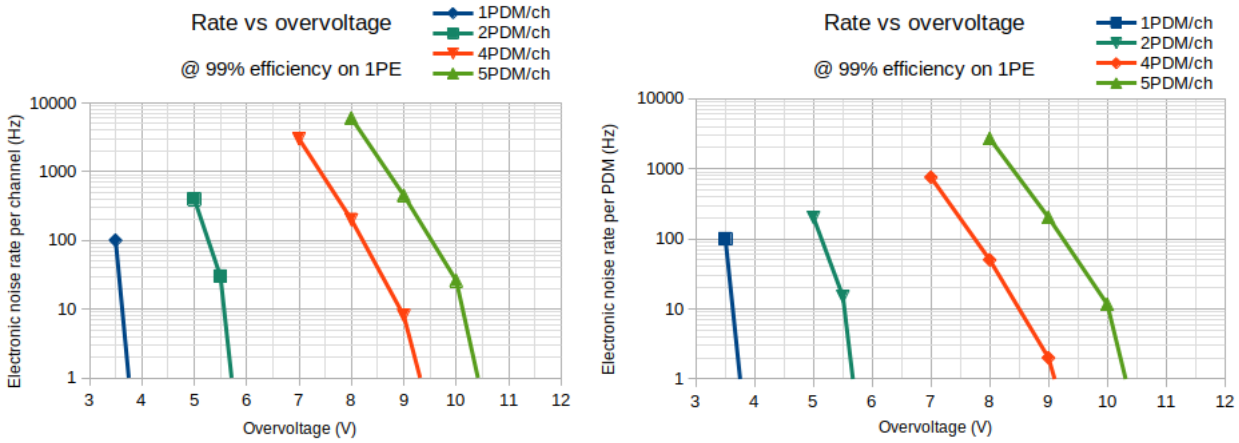

Figure 6: Calculated rate due to electronic noise (left: rate per channel; right: rate per PDM) at the working point associated to 99% efficiency on single photoelectron pulses, using the 1PE identification algorithm described in Section II C 2, when grouping several PDMs in the same readout channel. We assume $SNR = 2.0 \times V_{OV}(\mathrm{V})$ for at the PDM level. From S. Stracka.

matched filter on 1PE pulses, after downsampling to 125 MSPS) to the laser trigger time. The time resolution is calculated as the rms of the residuals, and the results are shown in Figure 7: requiring a time resolution on single photoelectrons equal to, or better than, 10 ns (in the offline analysis) results in a minimum operating overvoltage of $OV(N) = 4.0\,\mathrm{V} \times \sqrt{N}$ for Tile-15 (LFondry production). Including a 10% margin to account for a $\sigma_{SNR}/SNR = 5\%$ variability across different PDMs and neglecting all other effects, as done for the 1PE identification study, results in a lower bound on the minimum operating overvoltage, when grouping N PDM's together, of:

$$OV(N) = 4.4\,\mathrm{V} \times \sqrt{N}.$$

The above estimates do not include any safety factor due, e.g., to additional noise sources associated to transmission, or to different FEB configurations.

3. Bandwidth considerations

Grouping several PDM's together reduces the number of readout channels that must be connected to the digitizers, while increasing the occupancy for each readout channel (and thus the per-channel bandwidth utilization). With the TPC parameters currently assumed in the simulations, the occupancy (and thus the per-channel bandwidth utilization) is expected to scale linerly with the number of summed PDM in a single readout channel.

Both the overall number of channels and the per-channel bandwidth affect the number of digitizers that must be deployed for the TPC. In the baseline design, the output bandwidth for each digitizer is assumed to be 113 MB/s over a 1G link. CAEN is however developing a solution based on a 10G output link, which could allow a bandwidth up to $250 - 290$ MB/s. The digitizer boards feature rather large external DDR4 memory banks that may buffer a few seconds of data at 250 MB/s. Thus in the following we consider bandwidths as averaged over time (or, equivalently, over many events).

When operating a network in real life one should account for a loss in performance. We assume that during operations the network will sustain 80% of the maximum performance, i.e., 90 MB/s and 200 MB/s for the 1G and 10G solutions, respectively.

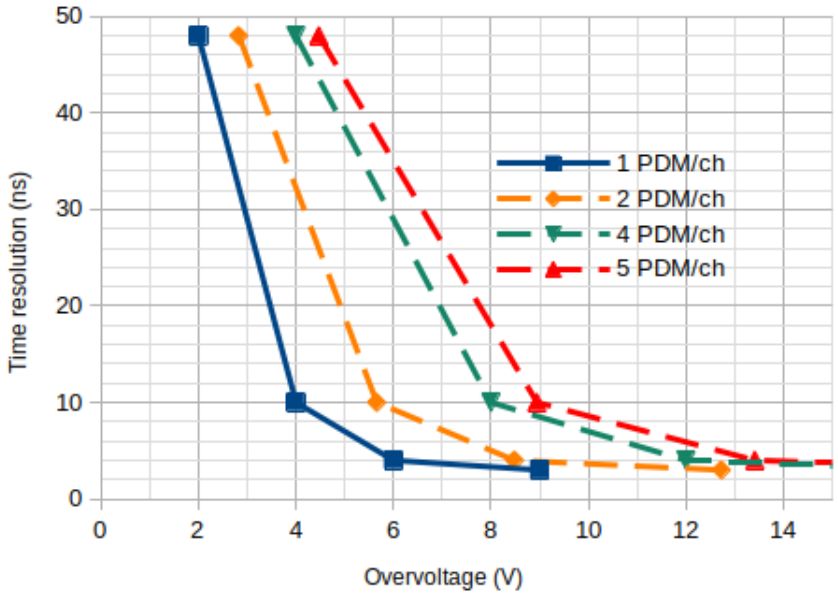

Figure 7: Time resolution with a matched filter on 1PE pulses as a function of overvoltage in Tile 15. Data for the 1 PDM/ch configuration come from a laser run on a LFoundry $5 \times 5\,\mathrm{cm}^2$ tile, after downsampling to 125 MSPS. The curves for $n = 2, 4,$ and 5 PDM/ch groupings are extrapolated according to the expected change ($\propto 1/sqrtn$) in signal-to-noise ratio. From S. Stracka.

PDM/ch	ch/PDU	BW/ch (MB/s)	#VX2740 (90 MB/s)	#VX2740 (200 MB/s)
Calibration runs				
1	25	1.5	136	136
2	13	2.8	136	68
4	9	4.1	136	64
4	7	5.3	136	64
5	5	7.5	136	64

Table V: Required number of digitizers (TPC only) in different configurations for grouping several PDM's in the same readout channel. With the sole exception of the 1 PDM/ch configuration, the limitation is the aggregated VX2740 output bandwidth, rather than the number of channels.

For science runs, we assume each PDM requires 1 MB/s of output bandwidth from the VX2740 (based on Fig. 8, which assumes a β/γ rate of 145 Hz).

Calibration campaigns are associated to additional bandwidth requirements. In the following it is assumed that a bandwidth usage of 1.5 MB/s/PDM will be sufficient for data-taking during calibration campaigns (thus including the β/γ background as well as any extra rate due to calibration sources). A more thorough study of the required data taking conditions during calibrations is needed to finalise this choice.

The number of digitizers extrapolated for every PDM grouping option is reported in V. It should be noted that the 10G option for the output link is still under development by CAEN and cannot be taken as granted. Should the 10G option not be feasible, the mitigation strategies already envisioned against excess bandwidth demand can be applied, including lossy and lossless compression in the digitizer FPGA's, with tradeoffs in physics performances.

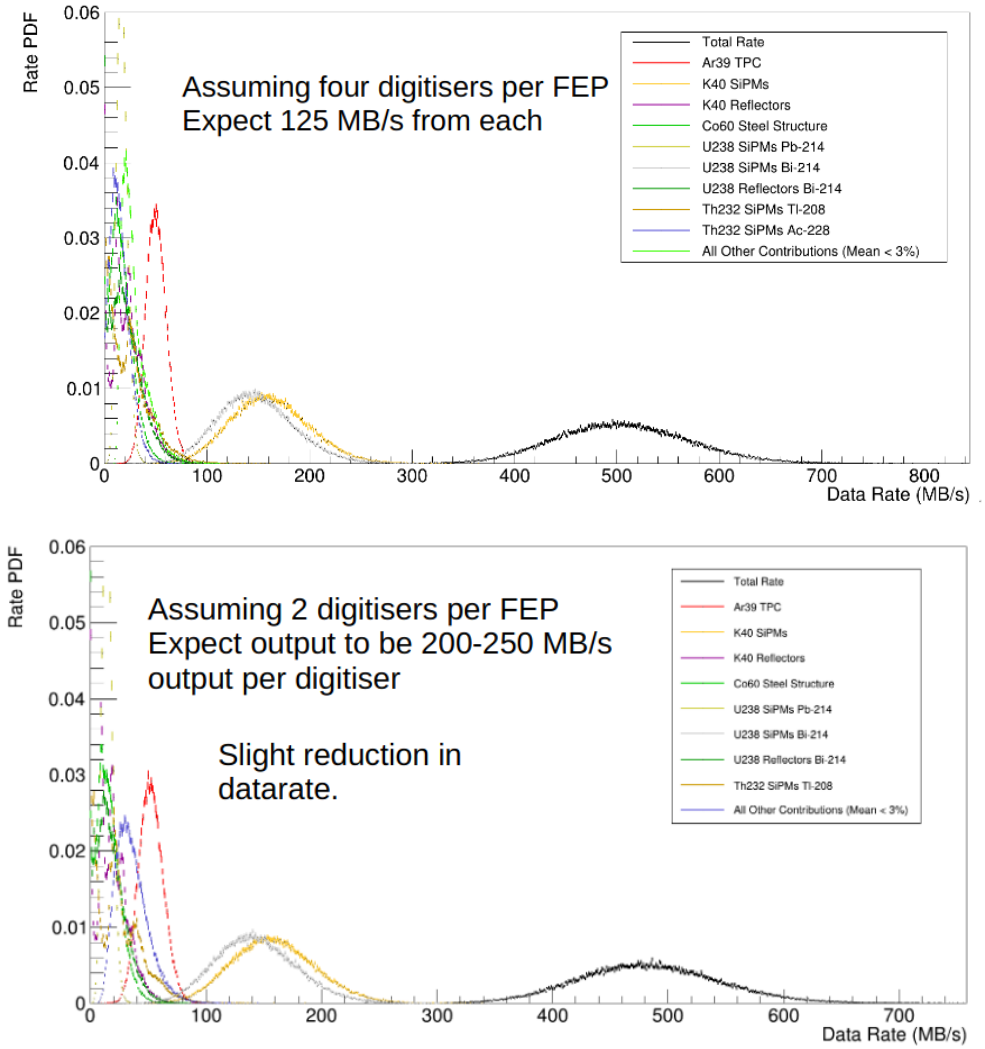

Figure 8: Required bandwidth at frontend server input for 2 PDM/channel (top) and 5 PDM/channel (bottom) configurations, due to β and γ backgrounds. Additional rate expected during calibration runs is not considered here. No other element in the detector geometry has been changed. Full description of the scenario at `https://agenda.infn.it/event/26025/contributions/131754/attachments/79727/103812/DAQCall_2ndMarch.pdf`. Correlated noise is neglected in these estimates, as is the variation of PDE with operating overvoltage. From M. Stringer.

4. Impact on the online farm

The output from several digitizers is transferred through an aggregation network to a frontend server. Each frontend server collects data from 1/16 of the TPC instrumented area, and studies show that the input bandwidth for each FE server is stable against variations in the number of channels grouped together. Therefore, the size and architecture of both the frontend (after the aggregation switch) and TSP farms are unaffected by the channel grouping.

It should be noted that, although a 290 `MB/s` throughput assumed in Sec. II C 3 falls short of the theoretical maximum for a 10G link, in the current network architecture the 10G NICs in the frontend processors set a limit on the aggregated data flow from all the digitizers feeding a single server.

Significantly increasing the throughput for the digitizer output would require a revision of the network architecture, shifting costs from the digitizers to the farm.

5. Operating at higher overvoltage

An increase in the PDM operating overvoltage is associated to an increase in DCR and correlated noise. The DCR values assumed in the estimates of bandwidth usage exceed the ones measured at LNGS for a tile operated at values of V_{bias} in range $[66 - 69]$ V and reported in Fig. 2.

Correlated noises are not included in the estimates in Fig. 8 and Table V. Using the input from Fig. 3, direct cross-talk and afterpulse are expected to increase the bandwidth usage by up to 10%. Preliminary observations, not fully documented, of optical cross-talk in ReD and at LNGS, combined with optical simulations in `https://agenda.infn.it/event/26322/contributions/133602/attachments/80195/104646/sandbox%20%283%29.pdf`, suggest optical cross-talk would increase the bandwidth usage by less than 10%. It is worth stressing that as the probability for correlated noise increases modelling its impact becomes more difficult.

Photo-detection efficiency is also expected to increase with overvoltage (increasing rapidly at low overvoltages, and then plateauing at higher overvoltages). No measurement of PDE at cryogenic temperatures is available. For the estimates in Fig. 8 a value of 40% at a wavelength of 400 nm was assumed. The bandwidth usage is expected to scale linearly with PDE.

D. Costs of digitizers and transmission

A reduction in the number of readout channels has an impact on costs. Using the information collected in TF2 we are in a position to estimate how the proposed configuration affects the total cost of cables and digitizers (for the TPC only). This is reported in Table VI. All other costs related to photoelectronics and DAQ are *not* included.

Table VI does not include any contingency or spares. This cost estimate should therefore be considered extremely preliminary. We remind the reader that in the bandwidth estimates used to calculate the number of required digitizers the effects of correlated noise, as well as the variation of PDE with operating overvoltage, are neglected. Both these effects are expected to increase the required bandwidth. Correlated noise probability and PDE are also expected to increase with overvoltage, and are therefore more relevant for the configurations with 4 and 5 PDM/ch.

			Cost (MEUR)	
PDM/ch	#ch	#VX2740	Optical	Electrical
Only 1G link available				
1	8280	136	4.05	2.40
2	4312	136	2.86	2.00
4	2984	136	2.46	1.87
4	2112	136	2.20	1.78
5	1656	136	2.06	1.73
10G link available				
1	8280	136	4.05	2.40
2	4312	68	2.08	1.21
4	2984	64	1.63	1.04
4	2112	64	1.37	0.95
5	1656	64	1.23	0.90

Table VI: Breakdown of costs for digitizers and transmission for configurations with different number of channels, for the TPC only. Both scenarios presented in the text and in Table V are considered: one in which only the 1G output link is available for the VX2740 boards, and one in which also the 10G option is available. For each digitizer we assume a cost of $11,520$ EUR. For the optical transmission we assume a cost of 300 EUR/ch. For the electrical transmission we assume a cost of 100 EUR/ch. VAT is not included. Spares are not included.

III. SIMULATIONS

A. Physics cases

The main physics goal of the DS20k experiment is the high-mass WIMP search. The WIMP signal is a single sited nuclear recoil (NR), in the fiducial volume (FV, the innermost 20 t of UAr, 30 cm away from the TPC lateral walls), in the [30, 250] keV$_{NR}$ energy range. We require that the detector enables to successfully reconstruct these evens with high efficiency, good 3D vertex reconstruction and no saturation (nor in S1 or S2).

Backgrounds are mainly coming from surface α's, internal β's and γ-rays from the detector materials, and span over a wider energy range (few MeV). γ-ray cascades from neutron captures on Gd in the vicinity of the TPC may deposit up to 9 MeV in the active TPC volume. DarkSide-50 and DEAP-3600 demonstrated how a good understanding of the backgrounds over the whole energy range is beneficial to constrain the backgrounds at low energy. However, the detector performance requirements at high energy are looser compared to the signal region: good energy resolution over the full energy range, good 3D vertex reconstruction for single sited events, and high-efficiency to correctly identify multi-sited (several S2s) events.

A thorough discussion of the absolute performance figures for the quantities mentioned above for both signal and backgrounds goes beyond the scope of this document.

Avoiding saturation for both S1 and S2 in the WIMP signal region is the only "hard" requirement that **must** be satisfied by any of the considered proposed configurations. To provide a more complete picture, we identified 3 areas where the SiPM performance and channel readout grouping will have a measurable effect on the DS20k physics reach. They are described below:

- **Saturation.** S1 saturation is not expected to affect NR events in the WIMP search region, due to the high number of readout channels ($>10^3$), the time characteristics and the good uniformity across the TPC of the S1 scintillation signals. However, MeV events will occasionally induce saturation on a fraction of the readout channels. We will estimate the effect of saturation on the energy resolution and linearity up to 2.6 MeV.

  S2 signals are more likely to induce saturation on channels very close to the S2 location. We will measure the probability for saturation to occur for NRs at the upper bound of the WIMP search region; and for single sited ERs of 700 keV, about the endpoint of the 39Ar β-decay spectrum.

- **Fiducialization (XY reconstruction).** The fiducialization along the vertical axis (Z) will use the drift time and it is expected to reach a resolution at $\sim$mm level. In the horizontal plane, the (X,Y) coordinates of events are provided by the light pattern of S2 PEs on the top array of photo-detectors. A plethora of algorithms is available, spanning from basic charge barycenter to machine learning based approaches. We will quantify the effect of correlated noise levels and channel grouping on the XY position reconstruction resolution as a function of the radial position of the event, over a wide energy range, covering the predicted S2 range for NRs in the WIMP search ROI.

- **S1 Pulse Shape Discrimination.** The SiPM response parameters (especially PE time resolution and the correlated noise levels) and the channel grouping may have an impact on the statistics of the PSD, as previously demonstrated [2]. We will assess variations in the NR acceptance induced by the proposed modifications to the baseline.

B. Simulation tools

1. G4DS

G4DS is the GEANT4 DarkSide simulation tool. It includes a detailed simulation of the geometry of the detectors of the DarkSide program, including the latest design of the DS20k TPC; an optical model to describe the TPC response, tuned on the real data from DS-50; a calibration of the LAr response to ER and NR based on data from DS-50, ARIS, SCENE and ReD. A summary of the calibrations is provided in the Appendix.

Full optical simulations of the DarkSide-20k detector predict an average detection probability of 19.5% for 128 nm scintillation light, resulting in a S1 light yield of 10 (9) PE/keV at null field (200 V/cm) at 40 keVee.

The S2 gain g_2, namely the multiplication factor in the gaseous phase, has been artificially scaled to roughly match the value reported by DarkSide-50, corresponding to $g_2 \sim 30$ PE/e-.

The TPC geometry implementation matches the latest technical drawings and conceptual design: the TPC is an octagonal prism, 3.5 m tall and 3.55 m wide (distance from two parallel walls). The reflector on the lateral sides (98% reflectivity for 400 nm incident photons) is coated with TPB. The TPB model is propagated from DS50, and assumes 100% VUV light conversion efficiency. The anode and cathode windows are also coated with TPB. We neglect however the clevios electrode coating, whose optical properties (rindex and transparency) are not known. The acrylic anode and cathode windows are 5 cm thick, and the optical properties are derived from DEAP-3600.

The baseline distance between the SiPM and the acrylic windows is 1 cm, to minimise the amount of passive LAr in front of the light detectors. The rule of thumb for optimal XY position reconstruction is that the distance between the gas pocket and the photo-detectors should match the linear distance of the photodetectors. This condition is automatically satisfied when PDM are readout individually. We thus repeated certain simulations after having increased the SiPM to gas pocket distance by additional 5 cm (10 cm total). A more refined optimisation of this distance parameter and the impact on the overall background budget go beyond the scope of this document.

Full optical simulations provide the inputs for the studies presented in this section. The PDM mechanics and design geometry have no impact on these results.

2. pyreco

(D Franco, L Pagani, M. Poehlmann, J Rode, S Hill, M Le Guirriec, P Agnes, C Savarese, A Razeto)

pyreco is the collection of high-level analysis tools developed to fully emulate the detector electronics response, and the online and offline DAQ and analyses chains. Starting from initial optical simulations (from G4DS or toy generators), it simulate realistic waveforms (using DSElec, the model for the SiPM response developed at LNGS); it then emulate the electronics chain and the DAQ operations; it implements hit-finding and event reconstruction, building on the software tools developed, in particular, for the analysis of ReD and proto-0 data. A schematics of the workflow is given in Figure 9.

C. Optical simulation samples

(Z Balmforth)

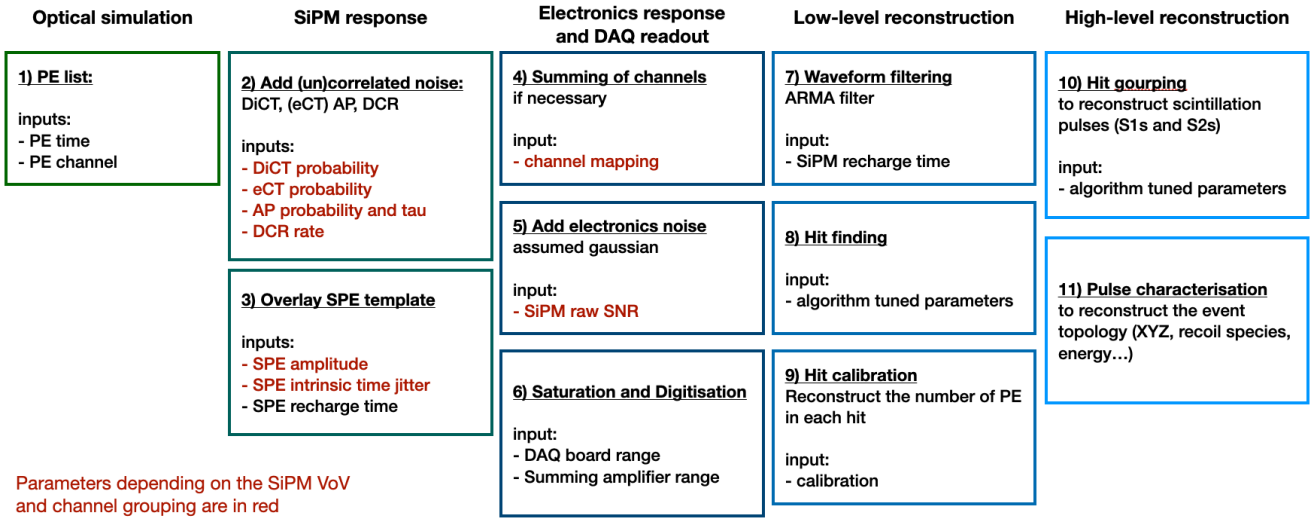

Figure 9: Schematics of the `pyreco` workflow.

Where not otherwise stated, the input simulation used for the results presented in the following Sections are generated with `G4DS` and reconstructed with `pyreco`. Several samples are available for the analysis, as listed below. Simulations involving S2 have been prepared in 2 geometrical configurations, namely with the baseline 1 cm distance between the SiPM planes and the anode window, and with 5 cm distance.

Sample list:

A: S1 for 300 keVNR NRs (1E3 PE) distributed in the full ACTIVE volume.

B: S1 for 700 keVee ERs (4.7E3 PE), uniform distribution in the FV

C: S1 2.6 MeV gammas from the PDMs

D: S1 9 MeV gamma cascade from Gd-captures (needs implementation of Gd in TPC walls). *Not ready yet, as of March 24th*

E(1): S2 (∼3.5E3 PE) from 300 keVNR NR, distributed in the full ACTIVE volume, assuming a larger S2 gain (g2 50 PE/e-, S2/S1 3). 1 cm SiPM offset

E(1): S2 (∼3.5E3 PE) from 300 keVNR NR, distributed in the full ACTIVE volume, assuming a larger S2 gain (g2 50 PE/e-, S2/S1 3). 5 cm SiPM offset

F(1): S1+S2 2.6 MeV gammas from the PDMs. (1cm offset)

F(2): S1+S2 2.6 MeV gammas from the PDMs. (5cm offset)

G(1): S2 [from 100 to 1E4] PE in the full ACTIVE volume. 1 cm SiPM offset

G(2): S2 [from 100 to 1E4] PE in the full ACTIVE volume. 5 cm SiPM offset

D. Geometry and Grouping Configurations

Five configurations have been identified to test the Physics performance. Those are chosen to determine the effect of grouping channels by 2 and by 4, w.r.t. to the baseline configuration (each channel read-out individually). The grouped options are tested in two geometrical arrangements at the top of the detector, as described above: 1 cm and 5 cm away from the anode window.

A 5x grouping configuration is not considered explicitly, the 2x2 configuration being a reasonable proxy for it. The tested geometrical configurations are listed below:

Parameter	1x	2x1	2x2	comments
OverVoltage	6 VoV	7 VoV	9 VoV	To determine correlated noise and DCR
Raw SNR (per channel)	Used 5 in `pyreco` *			To keep same hit finder efficiency
hit finding efficiency	>90%	>90%	>90%	from `pyreco`
SPE time resolution	4 ns	7 ns	7 ns	
SPE amplitude	6 mV	7 mV	9 mV	
DiCT probability	25%	30%	45%	Inspired from Fig. 3 **
AP probability	5%	6%	10%	Inspired from Fig. 3 **
eCT probability (emission)	20%	25%	40%	anecdotal
DCR / PDM	2.5 Hz	5 Hz	25 Hz	
PDE (max)	40%	42%	45%	

Table VII: SiPM input parameters used as input for the simulations. * In the actual `pyreco` simulations we assumed the same SNR, to avoid underestimating the hit detection efficiency and over-estimating the rate of "fake hits", due to a yet-to-be-tuned algorithm. ** Figure 3 actually reports the mean number of detected PE's per primary PE. Using these values as emission probabilities overestimates the noise levels, as they are simulated recursively.

(i) **The reference configuration:** 5x5 cm^2 PDMs, 1 cm distance between SiPM planes and TPC top window.

(ii) **2x1 grouping**, no offset. 10x5 cm^2 readout channels and 1 cm SiPM-TPC distance.

(iii) **2x1 grouping**, with offset. 10x5 cm^2 readout channels and 5 cm SiPM-TPC distance.

(iv) **2x2 grouping**, no offset. 10x10 cm^2 readout channels and 1 cm SiPM-TPC distance.

(v) **2x2 grouping**, with offset. 10x10 cm^2 readout channels and 5 cm SiPM-TPC distance.

E. Low-level SiPM input parameters

(C. Savarese)

The SiPM operation in configurations (ii) and (iii) is not expected to differ substantially from configuration (i), namely the SiPMs are operated at about 6 VoV. Configurations (iv) and (v) above requires to increase the SiPM VoV to $\sim$ 9 VoV, with the effect of increasing PDE, DCR and correlated noise levels.

The SiPM input parameters are thus adjusted accordingly, based on the results of existing measurements (e.g. the correlated noise levels are obtained from Figure 3) or on the results of dedicated simulations (e.g. to assess the hit finding efficiency and the resolution on the hit time).

These parameters are listed in Table VII. The Raw SNR entry in the table corresponds to the value used in input to the simulation: an artificially high value is employed to avoid underestimating the hit detection efficiency (or, conversely, over-estimating the rate of "fake hits"), due to a yet-to-be-tuned algorithm. The low-level parameters that depend on SNR are nevertheless included in the simulation, after estimating them from real, test-bench data (e.g., Fig. 7).

IV. IMPACT ON PHYSICS PERFORMANCE

A. Summary

1. Saturation

- Saturation occurs earlier for channels grouping several PDMs together, all other conditions being equal (Sec. IV B);
- Increasing overvoltage (e.g., when grouping 4 PDM per channel) decreases the dynamic range (when expressed in photoelectrons)
- Moving the sensors away (5cm instead of 1cm) partially compensates the increased occupancy of 4PDM/ch w.r.t. 1PDM/ch;
- Saturation is not expected to occur for WIMP events in the ROI (Fig. 11). Also expected to be manageable offline for other classes of events.

2. XY resolution

- 2x2 tiles grouping worsens XY resolution compared to 1 tile configuration. The same pattern is observed using different algorithms (Fig. 18 and Table IX). According to the offline group, a slightly worse XY resolution should be manageable;
- 2x2 tile grouping does not worsens the leakage of surface events in the fiducial volume, as reported in Fig. 19 (under assumptions in the text).
- Shift in PDM distance to 5 cm (from 1 cm) worsens XY resolution (Table IX).

3. Pulse shape discrimination

- Similar PSD performances are expected between 1 tile and 2x2 tiles configurations, after adjusting the overvoltage to compensate the lower SNR (see Sec. IV D). This study assumed ideal signal reconstruction performances, which may be an overly optimistic assumption, especially for the 4 PDM/ch readout case.

B. Saturation

(L Pagani)

Saturation at the level of each read out channel (due to the summing stage of multiple tiles or at the level of the single tile) is tested on sample C and G. Sample C represents the highest energy ERs background expected in the experiment due intrinsic radioactivity of SiPMs. S1 signal due to 2.6 MeV γ-rays are considered originated from the bottom SiPM where higher light collection is expected. Sample G represents S2 in a range extending well beyond the upper bound of the WIMP search region (provided $g_2 \sim 25$ PE/e-).

Different configuration are considered where both the SiPMs arrays are offset by 1 or 5 cm together with different grouping options (namely baseline where 1 tile corresponds to a single read out (RO)

	i	iv	v	
Saturation in sample C (S1 at 2.6 MV)	0.0%	22%	6%	only very close to windows
Energy resolution from S1 at 2.6 MeV	1.5%	1.7%	1.7%	
Saturation in sample G (S2 up to 12000 PE)	none	0.1%	none	only above the WIMP ROI

Table VIII: Comparison of the performance in terms of saturation.

channel or 2x2 where 4 tiles (2 by 2) are grouped together to became a RO channel). The levels of correlated noise and DCR are taken from Table VII. An increased level of external cross-talk (20%) is used for the results shown here, for both the configurations.

For these various options, we assessed the following quantities:

- the fraction of events which saturate,
- the linearity of the reconstructed signal, and
- for sample C, the resolution of the full absorption peak at 2.6MeV (all the γ-ray energy is deposited in the TPC)

An event is considered saturated if in any of the active RO channel the waveform amplitude exceeds the threshold determined by the electrical characteristics of the summing stage amplifier. The effect of saturation is studied both in the full and in the active volume, where the latter is defined by excluding roughly 40 cm from top and bottom of the TPC.

Saturated events can be spotted in plots of Figures 11 and 10 where the maximum amplitude (V_{max}) of the reconstructed waveform (for convenience rescaled by the amplitude of the single PE (mV_to_PE)) among the active RO channel is displayed for each event. A vertical line indicate the dynamic range for the corresponding amplitude of the single PE.

Figures 11 and 10 also show the linearity of the energy response, where the number of simulated PEs (mc_npe) is compared to the reconstructed (assessed using the full simulation and reconstruction chains). A linear fit of the energy response at low energy (<6000 PE) is projected at high energy. As in DS-50 this effect can be compensated (e.g. using the slow component of the S1 pulse).

To assess the energy resolution at 2.6 MeV a gaussian fit is performed on the full absorption peak. The resolution is then calculated as the ratio between the standard deviation and the mean of the gaussian.

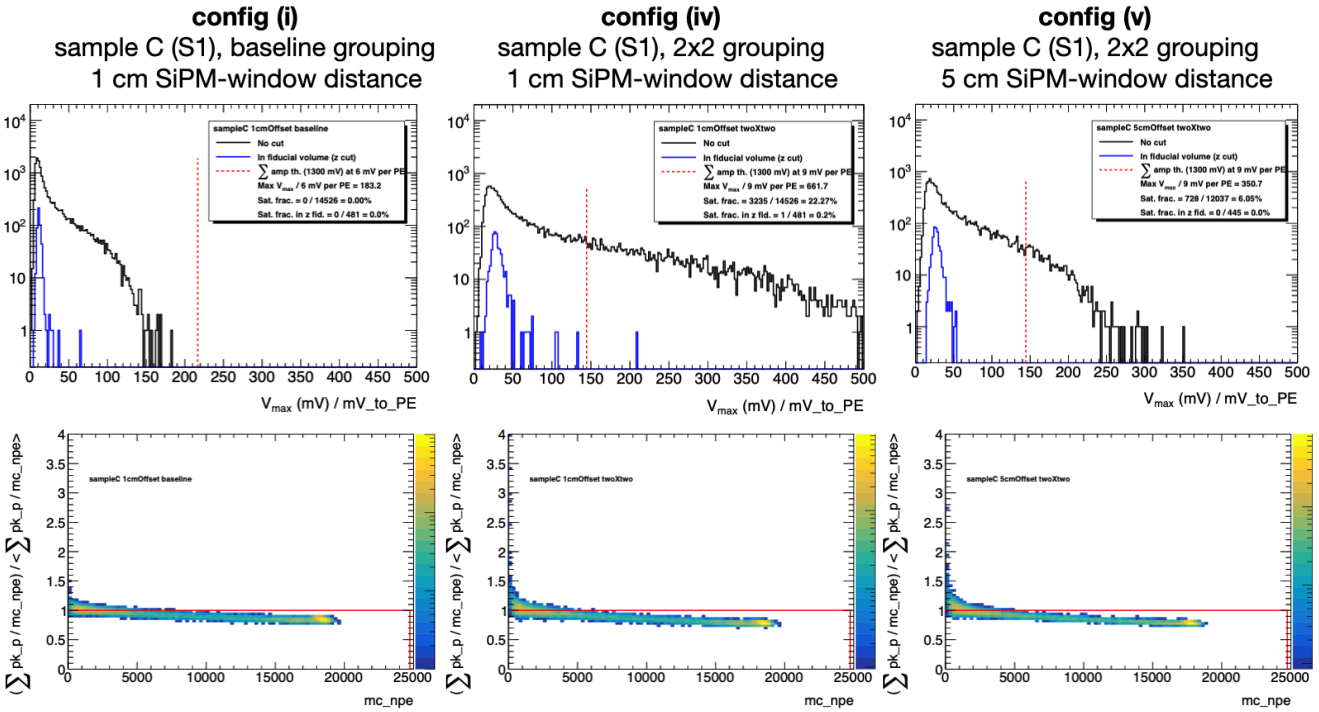

Figure 10: Sample C - S1 of 2.6MeV originated from the bottom SiPM array. Top: maximum amplitude (V_{max}) of the reconstructed waveform (rescaled by the amplitude of the single PE (mV_to_PE)) among the active RO channel spectra for baseline design at 1 cm offset compared with the remapped scenario (2x2) at 1 and 5 cm offsets. The effect on the fraction of saturated events when a z fiducial cut is applied is also shown (blue histogram). Bottom: scatter plot of the number of PE simulated vs. reconstructed (assessed using prominence). A linear fit (red) of the energy response at low energy (<6000PE) is projected at high energy.

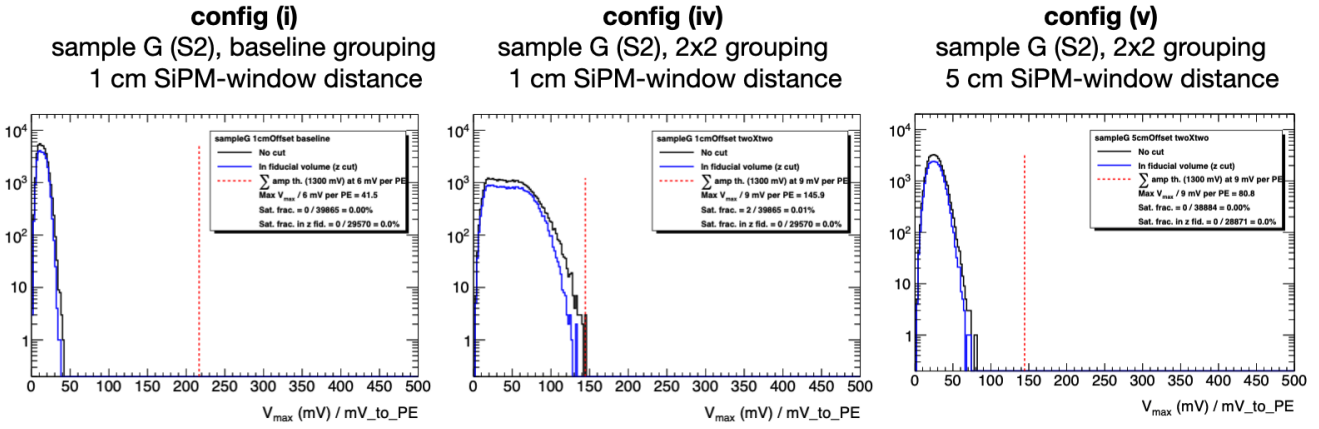

Figure 11: Same as Figure 10 but for Sample G - S2 corresponding to the upper bound for NRs in the WIMP search region.

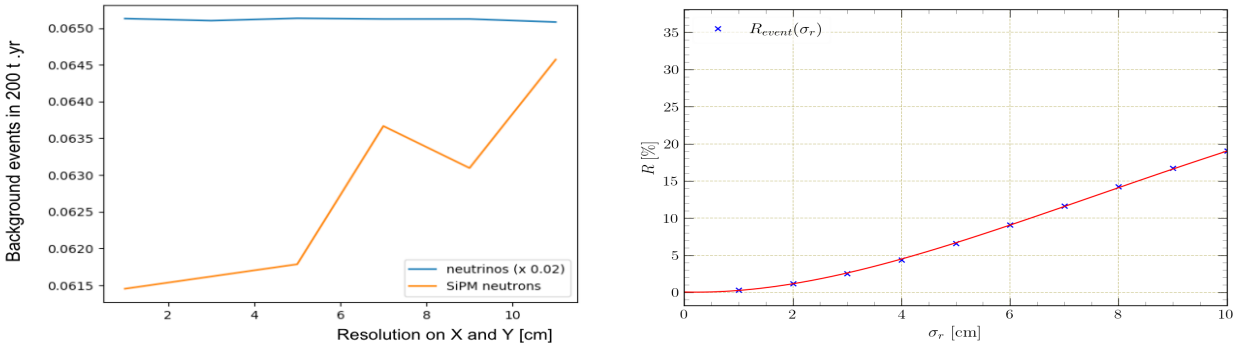

Figure 12: Absolute NR background rate in the FV in 200 t.yr (left) and relative increase of the gamma-ray penetration (right) as a function of XY resolution σ_r. On the left-hand-side plot, the neutrino event rate is scaled by 0.02 for easier comparison. Fluctuations are due to the toyMC statistics. On the right-hand-side plot, R is the relative rate increase, defined as the ratio between the number of events reconstructed in FV with the given resolution and the number of events in FV with perfect resolution.

C. XY position reconstruction

(M Poehlmann, D Franco, V Goicoechea Casanueva, S Copello)

1. Impact of XY position reconstruction resolution on background rates

Figure 12 provides the rate of events reconstructed within the fiducial volume (FV) as a function of the resolution on the XY position reconstruction. The FV is defined removing 30 cm from the lateral walls of the TPC and 70 cm from anode/cathode. The calculation is performed with a dedicated toyMC, for the followong background sources:

- neutrinos. The spatial distribution is uniform in both X,Y and Z.
- neutrons from the PDMS. The spatial distribution of WIMP-like events from this source is taken from dedicated G4DS simulations.
- gamma-rays from the detector materials. The spatial distribution of these events is also taken from dedicated G4DS simulations.

The effect of resolutions up to 10 cm is negligible for uniformly distributes sources. The spatial distribution of neutron induced NR and the gamma-ray ER backgrounds are exponentially falling as the distance from the TPC walls increases. For these sources, the increased rate in the FV is at the level of few per-cent for XY resolution of 10 cm.

2. CoG method

The baseline XY position reconstruction resolution is estimated using charge (hits) barycenter on the top array of photo-detectors for the samples of S2 pulses. The charge barycenter is calculated only in regions of plane where the signal density is high. Several algorithms can be used to identify high density regions: the simplest consists in drawing a MxN region of readout channels around the channel with the largest share of S2 signal; another option is to rely on specialised machine learning algorithms (such as DBSCAN [6] or KMEANS [7]). In the following, we will start from the results obtained with the first method (MxN, the most conservative approach); we will then test a slightly more sophisticated approach based on DBSCAN.

The charge barycenter algorithm is expected to bias the reconstructed event position towards the detector center for events close to the physical boundaries of the detector. The bias, which could be corrected in the real-life data analysis using data from external calibration or MC simulations, will get smaller for smaller MxN regions, at the cost of worsening the resolution. The most conservative approach is then to use the smallest MxN region around the signal maximum, without explicitly correcting for the reconstruction bias. The results obtained with the DBSCAN reconstruction are instead corrected for the bias.

The resolutions are calculated along both X and Y directions, and correspond to the RMS (or to the sigma after a gaussian fit) of the distribution of the distance between real and reconstructed coordinates (Xreco - XMC) and (Yreco - YMC). Results are presented for one direction only, since the resolution values are identical for X and Y. They are extracted from sample G in two energy ranges: 1000 PE $<$ S2 $<$ 3000 PE, corresponding to the lower bound of the WIMP search region; and 5000 PE $<$ S2 $<$ 13000 PE.

Results are presented here for configurations (i), (iv) and (v). The MxN approach uses both a 3x3 and 5x5 square around the channel with the maximum fraction of S2 for configuration (i), while for (iv) and (v) only a square of 3x3 is used, given the larger size of the readout channels. In the DBSCAN approach, the algorithm determines the number of channels used for the charge barycenter calculation.

Figures 13, 14 and 15 show the results obtained with the MxN approach, plotted as a function of the true radial coordinate of the event. Results are consistent across the energy range, and only shown here for the low energy range (1000 PE $<$ S2 $<$ 3000 PE).

Figures 16 and 17 show the results obtained with the DBSCAN approach for configurations (i) and (v), plotted as a function of the true radial coordinate of the event. Results are only shown here for the whole energy range (1000 PE $<$ S2 $<$ 13000 PE).

Table IX compares the XY resolutions extracted using the approaches described above. Table X brakes down the effect of correlated noise and event reconstruction on the resolutions. It is obtained by applying the DBSCAN algorithm to a sample of MC-truth PE, MC-truth PE with the addition of correlated noise source, and to fully reconstructed data. Fully reconstructed data are subject to larger fluctuations, due to the non linearity of the hit-finding algorithm applied to high-occupancy waveforms.

It is interesting to note how an increased shift in PDM distance (1 cm to 5 cm) worsens the XY resolution. The reason for this is attributed to the geometry of the detector and to the PE spatial distribution for S2 signals. A few supporting plots are shown in the Appendix.

More sophisticated algorithms have been tested on simulated samples.

3. Machine learning approaches

Machine learning has gained popularity in the last 10-15 years thanks to the rapid increase in the availability (and decrease in cost) of computational resources and of data. While for most methods the

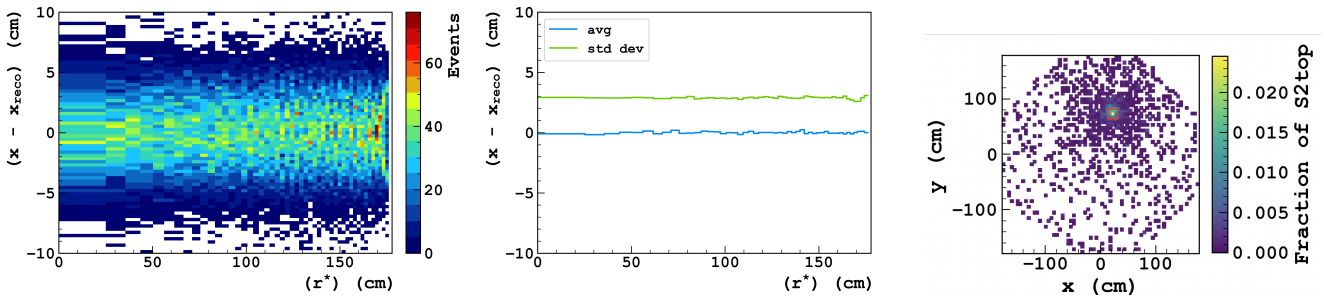

Figure 13: Distribution of the distance between true and reconstructed position after bias correction (left) and average bias and resolution (right) for S2 events, 1000 PE $<$ S2 $<$ 3000 PE, for events uniformly distributed in the TPC, as a function of the true event position (r*, in cm, is the distance from the TPC center along the line which intersects with the nearest wall). The MxN algorithm computes the charge barycenter using a 3x3 square of channels centered around the channel with the largest S2 share. The SiPM to window distance is 1 cm. This is configuration (i). From M Poehlmann.

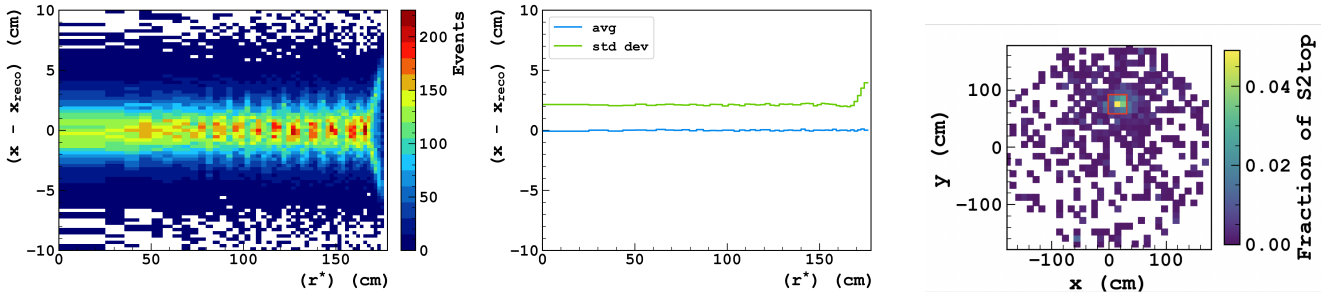

Figure 14: Same as Figure 13. Here the readout used here is configuration (iv), 10x10 cm^2 readout channels. The SiPM to window distance is 1 cm. From M Poehlmann.

training phase is computing-intensive, evaluating the output of, e.g., a trained classifier can be relatively cheap.

Several mature software packages provide a suite of powerful tools for classification and regression. Some of these tools are particularly suited to tasks of interest to DarkSide-20k. An example is convolutional neural networks (CNN), which find widespread applications to many computer vision tasks (a relevant example from HEP is the clustering, reconstruction, and classification of calorimeter images in many, if not most, collider experiments, in some cases starting from raw detector hits). An example

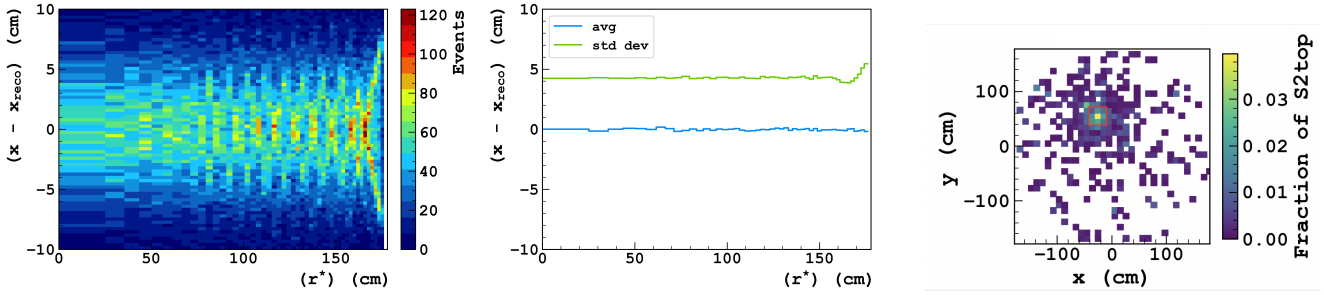

Figure 15: Same as Figure 13. Here the readout used here is configuration (v), 10x10 cm^2 readout channels. The SiPM to window distance is 5 cm. From M Poehlmann.

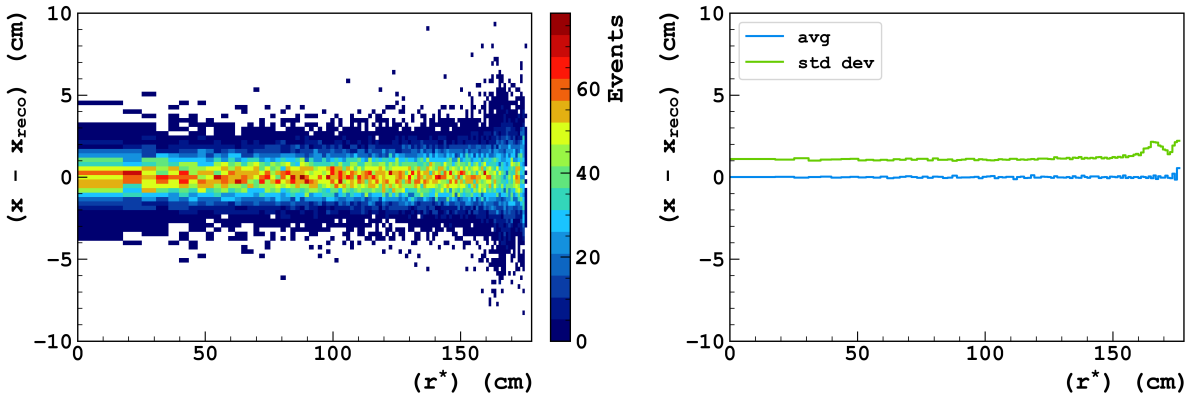

Figure 16: Distribution of the distance between true and reconstructed position after bias correction (left) and average bias and resolution (right) for S2 events, 1000 PE < S2 < 10000 PE, for events uniformly distributed in the TPC, as a function of the true event position (r*, in cm, is the distance from the TPC center along the line which intersects with the nearest wall). The DBSCAN algorithm is used to identify the PDMs included in the charge barycenter calculation. The SiPM to window distance is 1 cm. From M Poehlmann.

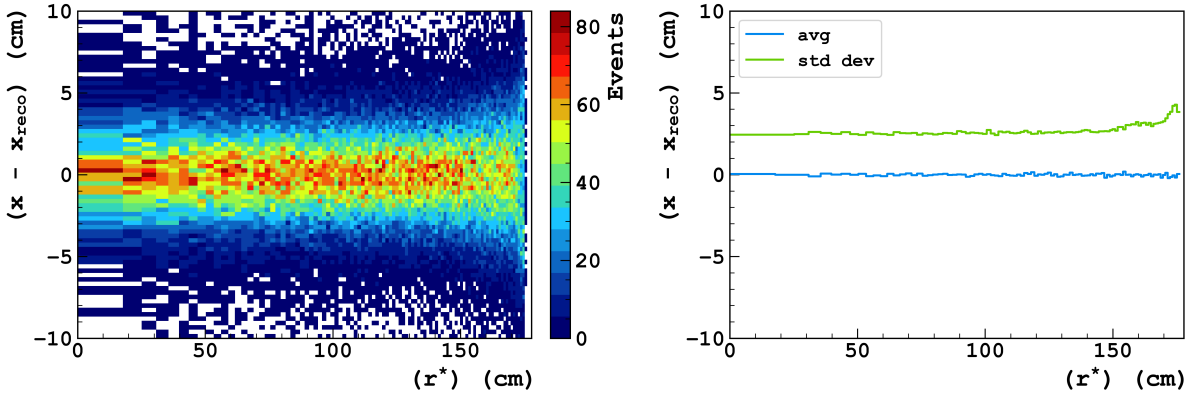

Figure 17: Same as Figure 13. Here the readout used here is configuration (v), 10x10 cm^2 readout channels. The SiPM to window distance is 5 cm. From M Poehlmann.

of the application of CNN to the reconstruction of the S2 position is shown in Fig. 18 (as for the CoG algorithm, in this exercise we are assuming that the detector response is perfectly characterized and that systematic effects due to data/MC discrepancies are negligible).

Depending on the chosen algorithm, a machine learning approach may deal with missing or partial data (e.g., due to faulty or saturating sensors), have the ability to satisfactorily approximate analytic methods, or cope efficiently with multivariate samples that may be difficult to describe with analytical models (e.g., because of non-linear correlations). This may often result in gains in terms of prototyping speed.

Disadvantages may include a lower degree of interpretability, the need to put in place reliable procedures for the validation of the trained models and the evaluation of systematic uncertainties. For the more complex models, this requires that sizeable collections of Monte Carlo, calibration and suitable control-sample data be available.

Grouping	PDM offset	XY Resolution: NPE = [1000, 3000]		
		DBCAN	3x3 Max Chan	5x5 Max Chan
Individual PDMs	1 cm	1.081 ± 0.005 cm	2.849 ± 0.008 cm	1.765 ± 0.006 cm
2x2 PDMs	1 cm	2.179 ± 0.004 cm	2.057 ± 0.004 cm	1.829 ± 0.003 cm
	5 cm	2.467 ± 0.004 cm	3.779 ± 0.006 cm	3.038 ± 0.005 cm

Grouping	PDM offset	XY Resolution: NPE = [5000, 13000]		
		DBCAN	3x3 Max Chan	5x5 Max Chan
Individual PDMs	1 cm	0.592 ± 0.002 cm	2.531 ± 0.005 cm	1.634 ± 0.004 cm
2x2 PDMs	1 cm	1.316 ± 0.002 cm	1.873 ± 0.003 cm	1.639 ± 0.002 cm
	5 cm	1.381 ± 0.002 cm	3.521 ± 0.004 cm	2.696 ± 0.003 cm

Table IX: XY position reconstruction resolutions, averaged across the TPC Fiducial Volume, for the tested approaches. For the 2x2 grouped PDMs case, we report the results for a 3x3 square around the channel with maximum share of S2; for the individual PDMs case, both 3x3 and 5x5 squares are used. The latter allows for an area similar to the area used for the 2x2 PDMs grouping. Errors are statistical.

XY Resolution = [1000, 3000] NPE - DBSCAN			
Configuration	MC truth	MC truth + DiCT&AP	Full reconstruction
Individual PDMs 1 cm offset	0.667 ± 0.004 cm	0.761 ± 0.004 cm	1.081 ± 0.004 cm
2x2 PDM grouping 1 cm offset	0.818 ± 0.003 cm	0.994 ± 0.003 cm	2.179 ± 0.004 cm
2x2 PDM grouping 5 cm offset	1.158 ± 0.003 cm	1.387 ± 0.003 cm	2.467 ± 0.004 cm

Table X: XY position reconstruction resolutions, averaged across the TPC Fiducial Volume, for the same simulated sample and XY position recontruction algorithm (DBSCAN) applied to different reconstruction methods: no reconstruction (MC truth), addition od correlated noise sources (MC truth + DiCT&AP) and full `pyreco` reconstruction.

4. Surface events

We tested the performance of the DBSCAN approach for events close to the TPC boundary (using a sub-sample of sample G, obtained by selecting events within 2 cm from any TPC wall and selecting events with S2 $>$ 1000 PE). We calculated the fraction of events reconstructed at smaller radii, a proxy to determine whether one readout configuration may induce a significant increase in the surface bg events compared to another. The results are presented in Figure 19, where the fraction of surface events reconstructed at smaller radii is provided as a function of the reconstructed radius. The boundary of the TPC is at R* = 175 cm. The FV bounday is at R* = 145 cm.

We limited this test to a MC truth dataset (using the number of optical and correlated noise PE's) rather than hits reconstructed with the reconstruction software. In these conditions, similar leakage probabilities for the 1x1 and 2x2 are found.

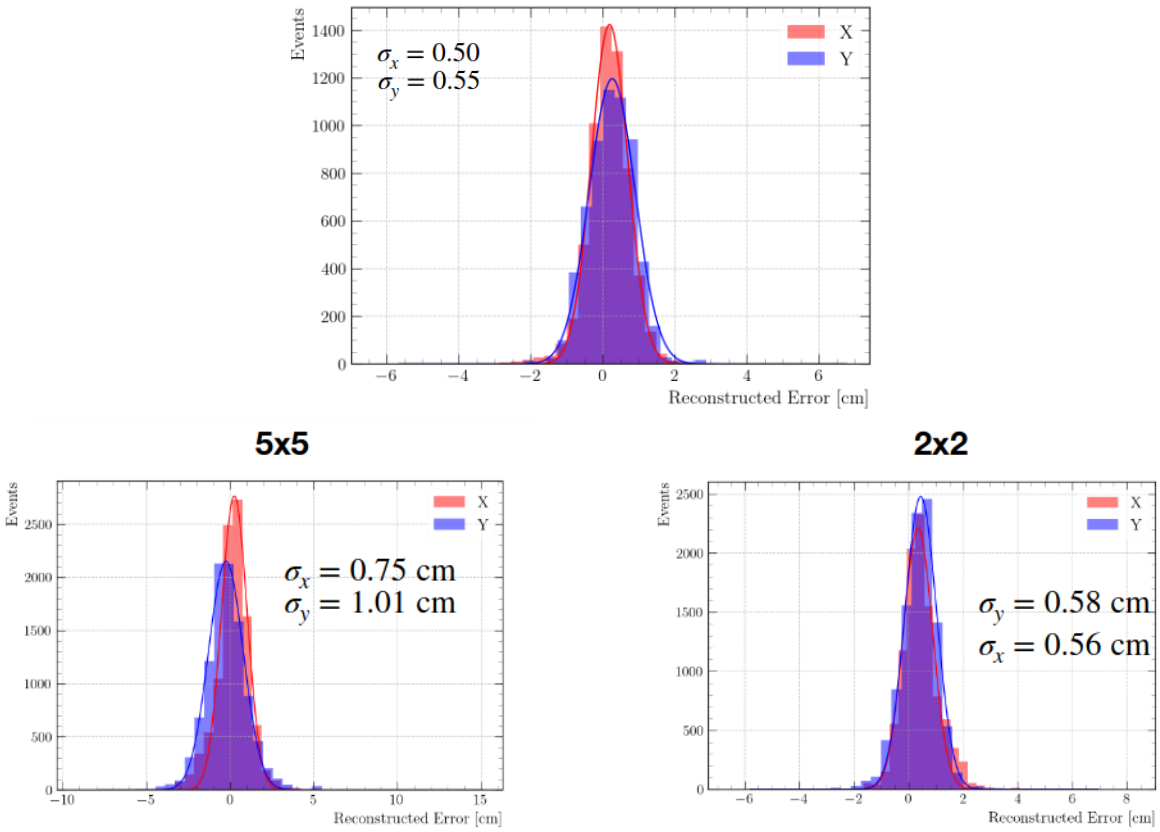

Figure 18: Position resolution for 2D CNN regression on a S2 signal with 20 000 photoelectrons in the top plane, for 1 PDM/ch (top), 4 PDM/ch (bottom right) and 25 PDM/ch (bottom left). The algorithm extracts the position of the cluster from the charge deposited on the 6×6 channels around the maximum. From V. Goicoechea Casanueva.

In general, the use of the algorithm on fully reconstructed data results in larger, non-Gaussian tails for events close to the TPC boundary. We interpret this as a feature induced by the event reconstruction procedure. The hit-finding algorithm is non linear in high-channel occupancy regimes, such as in channels above the location of an S2 production. This induces large fluctuations in how the S2 signal is distributed among the most illuminated channels. We argue this effect could be mitigated by an improved tuning of the hit-finding algorithm and/or by using a different approach (e.g. waveform integration in high-occupancy channels). Further work is needed to implement these improvements.

D. S1 PSD

(S Westerdale, C Savarese)

We tested the impact of the channel grouping and SiPM behaviour on the PSD performance using large statistics toyMC for NRs and ERs. The simulation consists of three steps:

- optical simulations;
- addition of DCR and exCT hits;
- overlay of the single channel SiPM response (DiCT and AP);

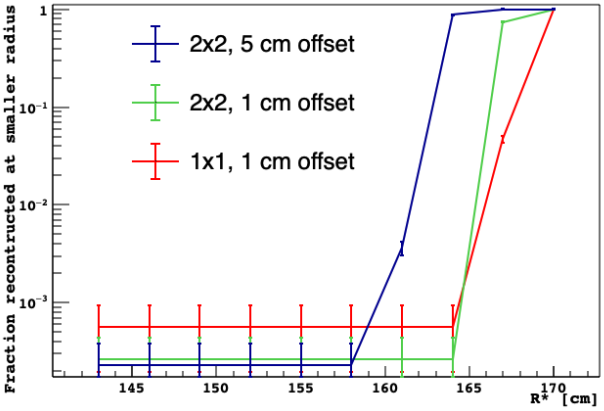

Figure 19: Leakage probability of surface events at smaller radii, using the DBSCAN approach on MCtruth PE's (neglecting the hit-finder efficiency). Where no events are reconstructed within a certain radius, the 90% CL upper limit is provided. The statistics of the 1x1 sample is smaller than the other two samples: this explains why right-hand-side plot the asymptotic limit of the red curve (< 155 cm) is above the other other two.

- extraction of the NR acceptance curve.

The first step consists in generating large statistics optical simulations, using the approach documented at this link: `https://docs.google.com/presentation/d/13bkng-tRf1ih5sHOodCgmQIchJwCKFB90IRp0WtiHy0`. This approach embeds the optical model implemented in G4DS, since it adopts G4DS-based light maps, to account for the photon travel time for events uniformly distributed in the Fiducial Volume. The NR and ER energy scales and the prompt-fraction calibrations are (approximately) taken from DS-50 and ARIS and include fluctuations due to quenching and electron-ion recombination. The PDE is adjusted, according to the operating VoV, as described in Table VII. The chosen energy range is at the PSD threhsold: 5-25 keVee ERs vs 20-100 keVNR NRs. 1E7 events for NRs and for ERs have been generated.

The second step is to populate the optical simulation output with additional PE's due to DiCT, DCR and exCT. DCR hits are uncorrelated in time and space (channel index) to the signal PEs; exCT are modelled according to `https://agenda.infn.it/event/26322/contributions/133602/attachments/80195/104646/sandbox%20%283%29.pdf`. The rate of DCR and exCT depends on the detector configuration, as described in table VII. Configurations (i) and configuration (v) were specifically targeted.

The third step is to determine the average SiPM response to a single PE. Simulations of realistic waveforms including single PE's and AP's were performed with `pyreco` to determine a p.d.f. for secondary PE's induced by a primary avalanche. In order to disentangle the effects of the hit finding algorithm and the reconstruction chain from the effects of correlated noise sources, we assumed the same SNR (and thus hit finding efficiency) for both the configuration. It is worth noting that a reduction of the SNR may imply not only a lower hit finding efficiency, but also higher rates of "fake hits". This effect, which may impact in particular the 2x2 configuration, have been neglected in this study. The single PE time resolution is however adjusted, according to Table VII, in order to account for the lower SNR of the 2x2 configuration.

The rate of DiCT and AP are taken from Table VII. The output of the optical simulations, including DCR and exCT, is then reconstructed by convolving each PE with the single PE response. The average photo-electron yields are shown for both NR's and ER's in Figure 20.

The optical photon yields are corrected for the expected SiPM PDE. We stress here that the PDE measurements (absolute value and dependence on operating VoV) are performed at room temperature.

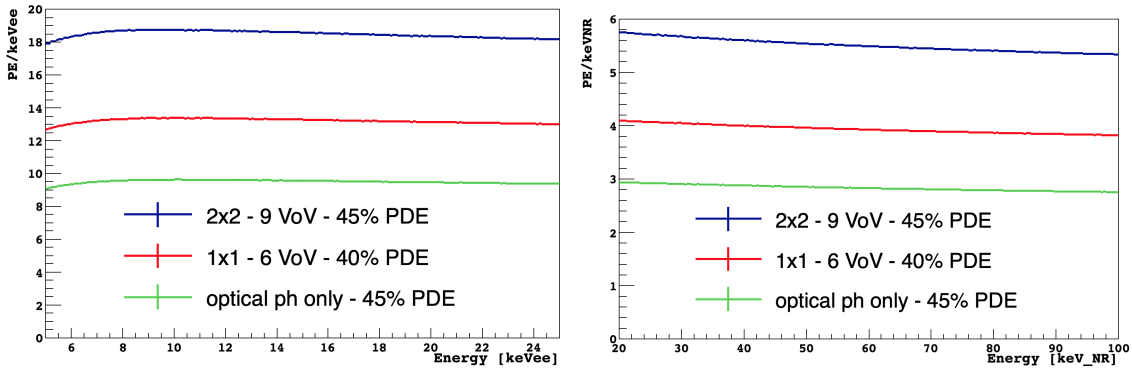

Figure 20: Extracted photo-electron yield for ER's (left) and NR's (right) from the toyMC. The blue and red curves correspond to the samples used in this study (2x2 and 1x1 respectively). The green curve, for reference, shows the optical photon yield only, obtained by neglecting all sources of noise. From P Agnes.

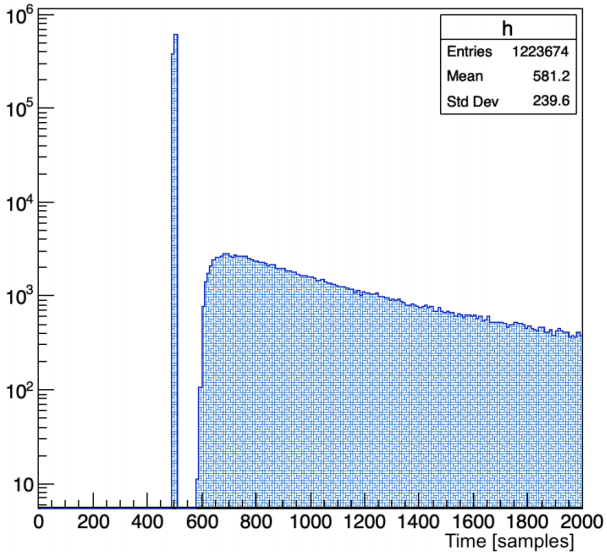

Figure 21: Time distribution for afterpulses. The narrow peak around 500 samples represents the arrival times of primary PEs. From C Savarese.

No measurement in LAr is yet available.

The forth step is to extract NR acceptance curves as a function of the NR recoil energy. 2D distributions of f200 (reconstructed prompt fraction) vs S1 (reconstructed number of S1 hits) are generated for both NRs and ERs. Both primary optical photons, induced DiCT, exCT, AP and DCR are accounted for in the the S1 signal. The 2D distributions are shown, for configurations (i) and (v), in Figure 22. Vertical slices of the 2D distributions, corresponding to a narrow S1 range, are used to determine: the NR acceptance as a function of f200 for each value of S1; and the corresponding expected ER background. The NR acceptance curve is calculated by requiring at most 0.1 ER background events in the full exposure (200 t.yr). The ER population in the fiducial volume and in the considered energy region is dominated by ^{39}Ar decays. The expected statistics is $\sim 2 \times 10^8$ events in the target exposure, and would require huge simulated statistics ($\sim 10^9$ in each narrow S1 interval). To overcome the statistics limitation, we used an analytical fit to extrapolate tails at the levels of 10^{-9} for the ER distributions. A discussion on

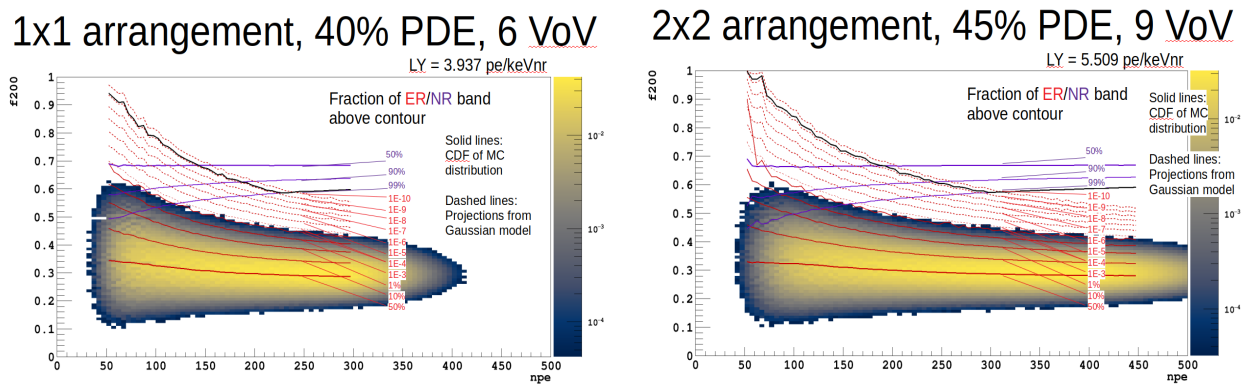

Figure 22: The output of the toy MC for the baseline 1x1 (left) and 2x2 (right) cases. The leakage curves are calculated with analytical extrapolations. From S Westerdale.

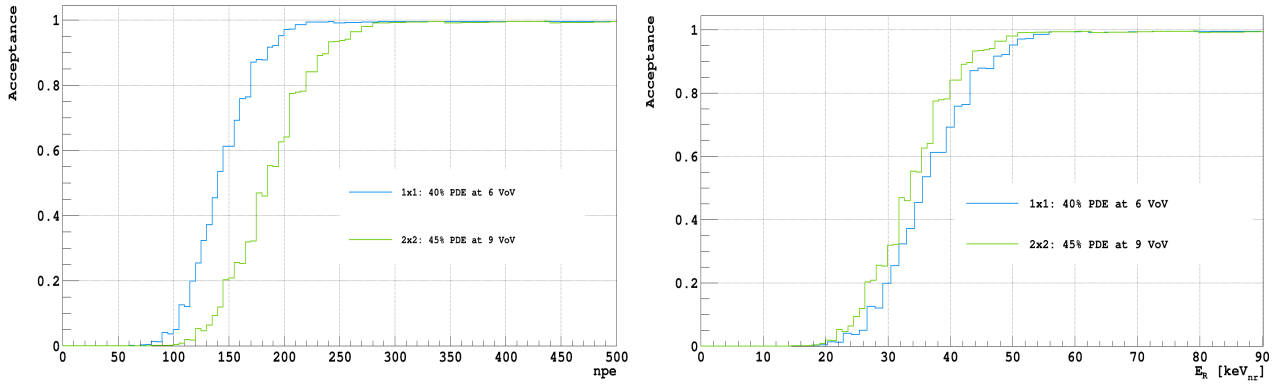

Figure 23: Acceptance curves as a function of S1 (left) and as a funcion of recoil energy (right). From S Westerdale.

the model itself goes beyond the scope of this work, but we heavily built on the past experience from both DEAP-3600 and DarkSide-50. The model used for the acceptance calculation is the "tail" model from DS-50 [5].

The NR acceptance is extracted as a function of the S1 size (number of detected PEs), shown in Figure 23 (left). This includes the extra PEs due to correlated and uncorrelated noise sources, which are different for the two considered configurations. The result can be expressed as function of true NR energy by scaling the measured S1 for the measured PE yield, as shown in Figure 23 (right). The acceptance is slightly better in the 2x2 case compared to the 1x1. This could be explained by the higher PDE, as suggested by the results of prior studies [2]. Further investigations are ongoing to quantify the extent of this effect.

The outcome of this study indicates that increased levels of correlated noise sources due to higher operation VoV will not induce a significant acceptance loss, for the WIMP search, compared to the baseline solution. This is true under the assumption that the SNR it kept high enough that the hit-finding efficiency is comparable. Thus, this results must be taken with some caution.

V. CONCLUSIONS

Two proposed (not mutually exclusive) options for the optimization of the detector have been studied: the first one involves the reduction in the number of amplifiers in each front-end-board; the second one consists in the reduction in the number of readout channels.

Both solutions aim to simplify installations (e.g., by reducing detector complexity) and/or operations (e.g., by reducing power consumption) at the price of lower signal quality compared to the baseline configuration. The impact of the signal quality on DAQ and physics performances has been studied with real data and simulations. The conditions chosen for these studies should be considered mildly optimistic: it was not possible to account for all sources of noise at this stage (e.g., transmission across the interfaces has been so far neglected, choice of different configurations of the amplifiers, ...). The loss in signal-to-noise ratio can be compensated by increasing the operating overvoltage, as long as dark count rate and correlated noise probabilities are low. Increasing dark count rate and correlated noise probabilities limit the operating overvoltage range from above, while photon detection efficiency and signal-to-noise ratio limit the operating overvoltage range from below. The possibility to adjust the overvoltage is one of the main knobs to recover performance, so one should allow for sufficient safety margins.

Reducing the number of amplifiers may enable a significant reduction in the heat dissipated by the electronics, which would simplify the cryogenic aspects of detector operations. Based on the power consumption of the individual active elements of the FEB, the heat load of photoelectronics could be reduced from $\sim 3\,\mathrm{kW}$ (4-TIA) to $1.5\,\mathrm{kW}$ (2-TIA) or $0.6 - 0.8\,\mathrm{kW}$ (1-TIA, depending on the adopted configuration) (Tables I-II). Early measurements by LNGS with a 1-TIA prototype suggest the signal-to-noise ratio would worsen by roughly 20% (compared to the 4-TIA configuration).

The main drive for the reduction in the number of channels is the simplification in the data transmission and installation (e.g., routing of cables), along with the potential cost savings. This reduction may occur at several stages, with different impact on the readout complexity and cost. In particular, this can happen at the PDU level before transmission (over cables or optical fibers), or just before digitization. A significant reduction in complexity can be achieved (only) by reducing the number of channels before transmission, whereas a significant cost reduction (summarized in Tab. VI for various configurations) is possible in both cases. We have assumed signal-to-noise worsens by a factor $\sqrt{n}$ when summing the output of n PDMs in a single readout channel, compared to the configuration in which each PDM is readout individually.

The impact on high-level physics performance for the reduction in the number of channels has been studied. In particular, we have analyzed the impact on saturation, X-Y resolution, and PSD performances. Saturation is not expected to occur for WIMP events in the ROI in any of the tested configurations, and it is expected to be manageable with offline analysis for other classes of events. A 2x2 tiles grouping is expected to worsen the XY resolution compared to 1 tile configuration. However, a slightly worse XY resolution is not expected to be a show-stopper. It should also be noted that the applied XY reconstruction algorithms are at an early stage of development. Finally, similar PSD performances are expected between 1 tile and 2x2 tiles configurations, after adjusting the overvoltage to recover a similar SNR. This study assumed ideal signal reconstruction performances, which may be an overly optimistic assumption, especially for the 4 PDM/ch readout case. For PSD studies, an offline time resolution $\sigma_t < 10\,\mathrm{ns}$, to be achieved by a matched filter, has been assumed for all configurations [2]. The minimum operating overvoltages resulting from the requirement on σ_t is summarized in Fig.-7.

For DAQ operations, one of the key elements is the ability to identify 1 PE pulses with low background rate ($< 10\,\mathrm{Hz}$) with simple algorithms (so that it could fit in the resources available in the digitizers FPGA). The minimum operating overvoltages resulting from this requirement can be extracted from Fig. 6.

To summarize: among the options considered in this report the baseline layout (1 PDM/ch) offers the best signal quality and flexibility in operation (i.e., the possibility to adjust the overvoltage); grouping 2 PDMs in a single channel offers a good compromise between cost, complexity, and signal quality, while maintaining some flexibility during operations; grouping 4 PDMs or more in a single channels reduces complexity the most, and offers some additional cost reduction, at the price of higher correlated noise rates, worse signal quality, and lower margins for adjustments during operations.

[1] A. Razeto, L. Consiglio, and G. Korga, Traces for LFoundry Tile 15, collected in February 2021. Retrieved from URL `http://ds50tb.lngs.infn.it:2180/SiPM/Tiles/LFOUNDRY/pre-production-test/TILE_15/LF_TILE15_77K_*_*_*.wav` on February 20, 2021.
[2] C. Savarese, DS collaboration meeting March 2020. `https://agenda.infn.it/event/21595/contributions/111721/attachments/71155/89175/SiPM_on_Science.pdf`
[3] Gola A., et al., Sensors **19**, 308 (2019). `https://doi.org/10.3390/s19020308`
[4] Gola A., et al, Phys. Med. Biol. **59**, 3615 (2014).
[5] `https://www.dropbox.com/s/m2bgafujdv8hr7x/psd.pdf?dl=0`
[6] `https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html`
[7] `https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html`

Appendix A: Appendix

a. S2/S1 and energy scale in DS-50

The calibration of the energy scales and the detector observables is based on past work, mainly from DS-50 and ARIS. We report below some reference plots and a summary table comparing some of the reconstructed quantities for a selection of events.

Energy	$\mathcal{L}_{eff}$	S1 in DS-50	S2 in DS-50
10 keVee ER	-	75	2200
60 keVee ER	-	420	22000
500 keVee ER	-	3000	200000
2600 keVee ER	-	$>$ 20000	-
30 keV NR	0.27	70	2000
300 keV NR	0.33	600	3500

Table XI: Typical values for NR and ER signal and backgrounds.

b. Optical Modelling of the DarkSide-20k TPC

The optical simulations have been performed with G4DS. The sequence of materials implemented in the simulations (and thire refractive index at 420 nm) is reported in Figure 26 for the case of 1 cm and 5 cm SiPM-window distance.

Full optical simulations of single electrons extracted in the gas at the TPC center can be used to record the true position of optical photons hitting the silicon. The cumulative distribution of photons

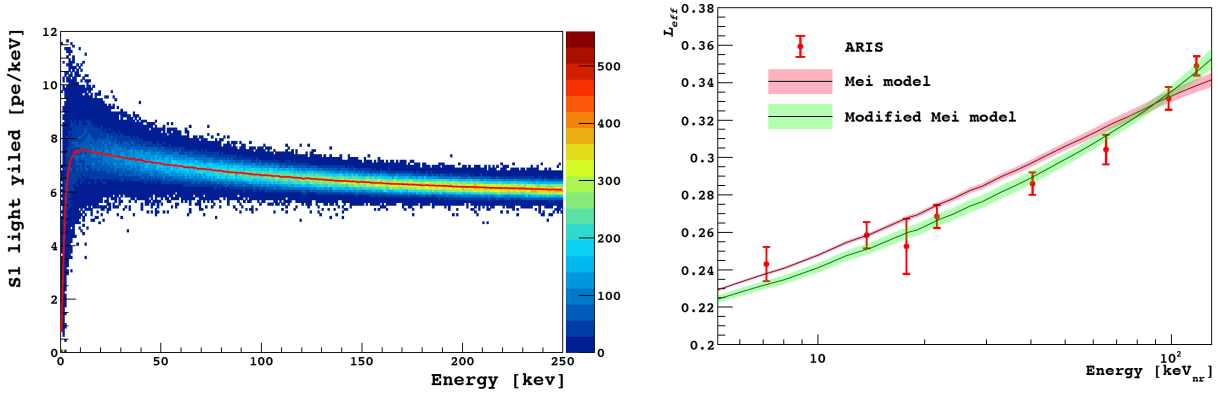

Figure 24: S1 LY for ER in DS-50 and relative scintillation efficiency ($\mathcal{L}_{eff}$) for NRs from DS-50 and ARIS.

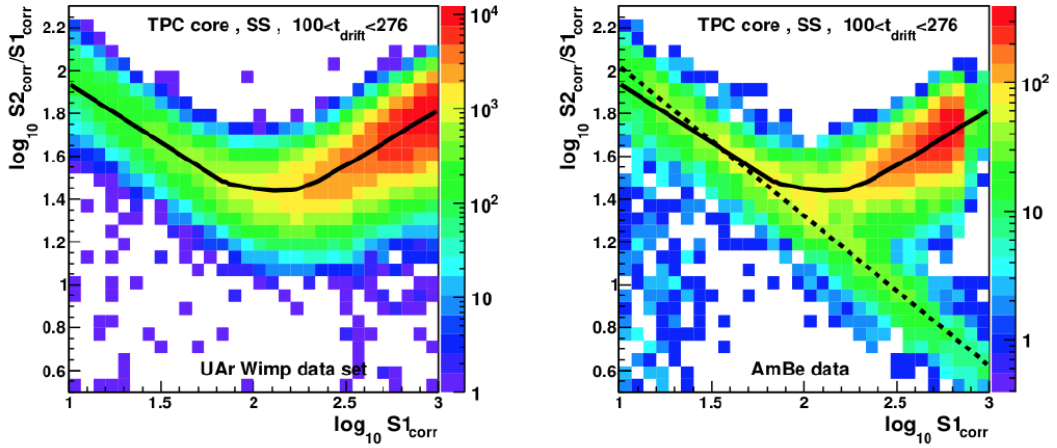

Figure 25: S2/S1 for ER and NR as measured in DS-50. $g_2 = 23.1 \pm 1$. PE/e-. From M. Ave Pernas.

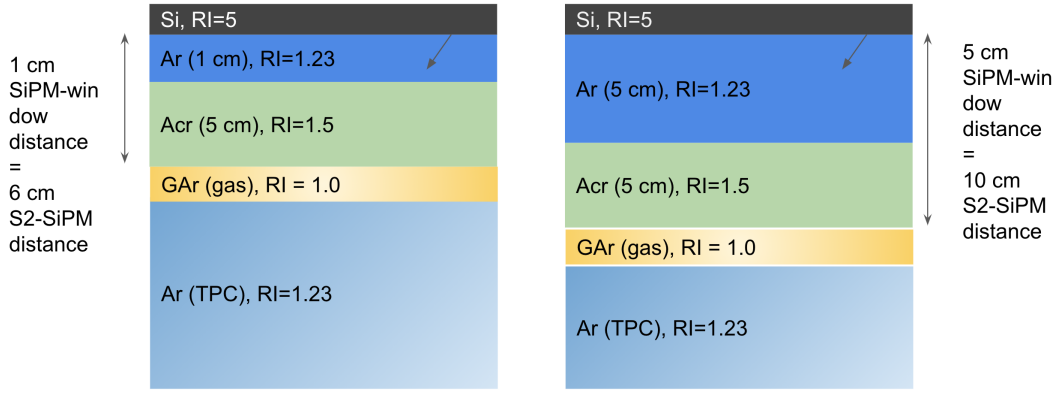

Figure 26: Sequence of materials and refractive indexes in G4DS for S2 simulations.

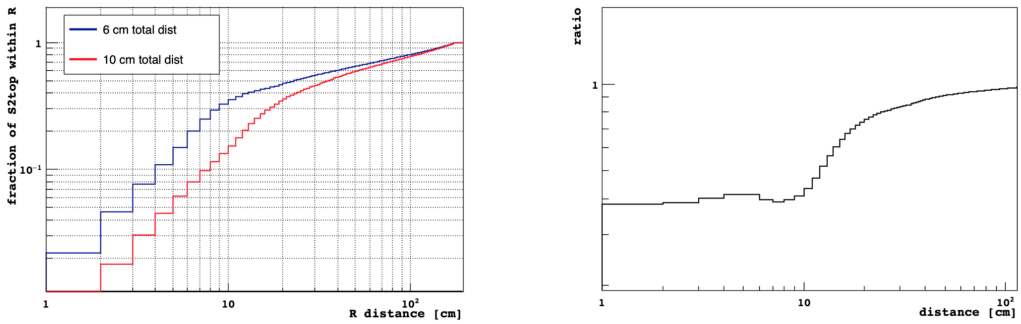

Figure 27: The cumulative distribution of photons collected as a function of the radial coordinate, for events at the TPC center (left). The right-hand-side plot shows the ratio between the red and blue curve of the plot on the left.

collected on the top array of detectors is shown in Figure 27. A 5 cm radius circle, centered around the XY coordinate of the interaction, contains the 12% (5%) of the S2 signal collected by PDMs on the top array. Due to total internal reflection at the exist of the acrylic, the amount of photons collected per unit area scales with the solid angle only for R $<$ 10 cm.

c. Revision History

The first full draft was circulated for comments within the TF2 WG on the 10/04/2021.

- **11/04/2021**: selection of XY resolution plots refined and minor text fixes to capture DF comments. **21/04/2021**: yet to be finalised.
- **13/04/2021**: revision of the text in the Conclusions Section to capture TT comments.
- **14/04/2021**: update of saturation plot selection, following to DF comments. Update of the analysis, since the previous version included too much correlated noise, from LP.
- **15/04/2021**: update of the right-hand-side plot of Fig. 12 on gamma penetration from VGC.
- **15/04/2021**: revision of the text in Section IV D to capture CS comments.
- **18/04/2021**: revision of the plots in Section IV D to capture MR comments. Found a bug in the S1 simulation, yielding S1 integral to be reduced by 1/3. Added updated LY plots.
- **21/04/2021**: update of the NR acceptance plots in Fig. 23 and Fig. 22 to reflect the higher LY simulations.
- **22/04/2021**: some text revision to respond to MR's comments.