nature portfolio

Peer Review File

DarkSide-20k sensitivity to light dark matter particles

Corresponding Author: Dr Fabrice Hubaut

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript projects sensitivities for light dark matter searches with the DarkSide-20k experiment. Since the projections are world-leading for a range of benchmark models, they will be highly relevant for the field for several years. I expect to recommend publication if the authors can address some questions and comments.

{} Please add csv files or something similar to the supplemental material to distribute the (mass, cross-section) datapoints for the projected sensitivities. It also helps to add them to the arXiv preprint, if you update this.

{} What exactly are the "expected limits" shown in all the plots? Are these mean or median background-only limits? And could you also indicate the band of expected limits in at least one of the figures -- traditionally the band between the +- 1 sigma quantiles of background-only limits is shaded green.

{} l. 255: DS-20k will have a much larger drift time than DS-50, so electron clouds experience more diffusion and S2s will be wider. Thus, assuming the same resolving power in drift/depth as DS-50 seems unrealistic. The resolution may also be a nontrivial function of depth.

{} l. 282: "no fiducialization is performed in the electron drift direction". In the large LXe detectors I am familiar with, S2-only analyses can effectively fiducialize in the drift direction -- at least for S2s larger than a couple electrons -- by a cut on the S2 width, exploiting diffusion of the electron cloud during drift. Is such a cut assumed here, or not?

If not, why do you not expect a large background from radioactive elements stuck on the cathode, or external backgrounds near the top of the detector, as we see in LXe detectors? If yes, it would be good to elaborate on how the signal acceptance and background rejection of such a cut is modeled (as the S2 width is a nontrivial function of S2 as well, for small S2s).

{} l. 267-9: Why would cuts against accidental coincidences -- if I understand them correctly as events with unrelated S1 + S2 signals -- reduce the effective livetime of an S2-only analysis like this by half? Perhaps this refers instead to cuts of livetime following a large S2, to reduce the few-electron background?

{} l. 317: Is there a reference for this factor 100?

{} l. 322 - 324: it is not clear to me what "a linear uncertainty on such effects" means for the model. Perhaps a reference to another DS-20k paper suffices here, otherwise, a plot of the assumed spectral shape and its uncertainty (all the way to 0 eV) would be helpful.

{} l. 343: Is there a reference for this factor 2.5?

{} l. 350: As detailed in the suggestion below, I would here immediately say something about how this background is much more uncertain than the others, and that you will consider two different scenarios (ultimate / conservative). I would also put the sentence that the model is the one from [18] in front, before the short summary of the model.

{} Table II / l. 408-9: It would help to clarify these are not the actual uncertainties you currently have on the parameters, but the uncertainties you _expect_ to have at the start of the analysis. The actual uncertainty on some parameters (like the 39Ar activity) would be much larger and difficult to quantify; you account for this by choosing conservative assumptions (and you report on some variations of the nominal conditions later in the paper).

{} l. 435-438: As mentioned above, a figure or reference to such that details the 39Ar spectral shape model would be appreciated to help contextualize this remark.

{} l. 441: It may be worth clarifying that the 90% C.L. CLs procedure you chose has a similar median sensitivity as the two-sided Feldman-Cousins approach recommended in [31] and used by several of the results you compare with. (A 90% one-sided search without CLs would have a better median sensitivity, and would not be a completely fair comparison.)

{} l. 453-455: Here it may be worth clarifying that the impact of the quenching fluctuations, and the uncertainty on the availability of the SE model (encoded in the ultimate vs conservative projection), is even larger than this.

{} Figure 3: Are the positive signal contours just shown for historical purposes? If I understand correctly (see e.g. https://arxiv.org/pdf/1506.03924 and elsewhere), the CoGent excess is now well-understood as being due to an underestimated background. For both CoGent and CDMS, I believe later limits from the same collaborations rule out a WIMP interpretation of their earlier results.

{} Figure 4:
* I would not use nearly-invisible-pink for PandaX-4T (also not in in figure 3), since it is leading in the bottom two figures on the left (and figure 3). Perhaps use the bright magenta that is currently used for PandaX-II, which does not seem very relevant to show anymore.
* The lower right panel reserves a color for NuSTAR in the legend, but I did not see it in the plot. Perhaps this information is better communicated with some text inset in the plot. It is also a bit strange to feature a benchmark model that has been so heavily excluded -- perhaps a spin-dependent dark matter model would be more interesting? Or is there a reason why these limits remain relevant after the NuSTAR result?
* Please encourage the authors of [16] to publish an erratum that you can then cite here.
* I found myself writing captions with the different subfigures with a pen, e.g. "Migdal + NR" or "Dark photons", after scanning the y-axes and captions. You could consider saving other readers the trouble. Using a larger font for the axes and labels might also help.

[Below I grouped some related suggestions with a general comment.]

The manuscript shows sensitivities under several different scenarios, corresponding to conservative and optimistic scenarios for the main unknowns:
* how LAr responds to low-energy nuclear recoils (quenching fluctuations or not);
* whether the spurious electron background appears as expected and can be accurately modeled;
* whether the Migdal effect takes place as expected.

Although it would be more convenient to have a single projection that is "realistic but on the conservative side", I agree with the author's choice here -- the uncertainties are large and quite impactful, and opinions differ on how to best treat them. However, given this, I would give the reader more information to make up their minds on which model they would like to use. Specifically:

{A} It would help to reference (planned) efforts to resolve the uncertainties. Are there plans (by Darkside-20k or others) to measure the NR yield fluctuations in LAr? Similarly, while the spurious electron background will likely have to be characterized in-situ, references to efforts to characterize this background might bolster the case that the "ultimate" projection is attainable.

{B} I would recommend being clearer about the limitations of the SE background model. The authors assume the empirical model from reference [18], which has been fitted to Darkside-50 data. I doubt this model has predictive power for the 1000x larger detector that DarkSide-20k expects to be. For example, reference [18] assumes the single electron rate scales linearly with event rate, which would make sense in some range around the DS-50 observations. For a 80/1.5 ~ 50x increase in event rate, I would think it depends on exactly how delayed single electrons are distributed in time following their progenitor, on which [18] is inconclusive. The model in [18] was also derived for the particular event selections of DS-50, which may have to be substantially changed in DS-20k.

{C} For 39Ar, the authors assume the same activity as in Darkside-50. This is conservative. However, they note that "significant improvement in sensitivity is expected if the UAr extraction plant would further reduce the contamination of 39 Ar". It may be worth briefly commenting (or referencing) plans to such improvements -- even though the authors do not here assume they will come to pass. This helps the reader judge how likely it is that the sensitivities will be significantly better than those quoted here.

{D} For Migdal, you could consider adding some statement (ideally with references) that experimental efforts to measure this effect are under way and/or that they are inconclusive at the moment. Again, this explains to readers why you show two plots rather than just one.

Reviewer #2

(Remarks to the Author)

This paper describes a light dark-matter sensitivity study for the next-generation detector of the DarkSide liquid-argon direct dark-matter search program, DarkSide-20k. It estimates achievable limits for the expected experimental parameters using a simulation model and background estimates informed by results from the smaller DarkSide-50 experiment.
This is a fairly bread-and-butter type of paper, not containing any new experimental physics result. Nevertheless, such a documented study of projected sensitivity is a useful addition to the literature, in the context of a community of numerous competing experiments with different approaches.

The DarkSide program is by now quite mature. I do not spot obvious errors or inaccuracies in the provided information, although most details are embedded in references.

I recommend publication.

I have the following minor suggestions:

Does "Clevios" need a "TM" or the like?

Fig 1: I suggest toning down (or turning into grey) the colors for items not referenced in the caption (for example, more than one thing is pink).

It is worth citing the recent PANDA-X and XENON measurements of the neutrino "fog" in the relevant section.
Line 360: "backgrounds to the DM search" ("background" depends on what you are looking for...)

Table II: most of the uncertainty sizes are conveyed in the pre-fit plot and presumably details are in the reference. Can you indicate the approximate size of the "atomic exchange and screening" contribution? Presumably that is combined with shape uncertainties related to the uncertain Q-value in the pre-fit background spectra (it is of interest given the statement about eventual limiting 39-Ar beta shape uncertainties).

Line 435:
-> "As 39-Ar is the dominant background, its spectral shape will"

Reviewer #3

(Remarks to the Author)
The paper “DarkSide-20k sensitivity to light dark matter particles” assesses the sensitivity of the future DarkSide-20k to models predicting light dark matter particles below 10 GeV. The authors claim at least one order of magnitude improvement in dark-matter interaction cross-sections for WIMPs with respect DarkSide-50, which is the predecessor of DarkSide-20k. The analysis focuses on exclusion limits for spin-independent WIMP nucleon recoil with quenching fluctuations, although other signals and no quenching are also considered.

As mentioned in the article, DarkSide-20k is designed to perform a nearly instrumental background-free search for WIMP masses above 20 GeV. This paper represents an effort to understand the capabilities of the future DarkSide-20k in the region below 10 GeV. It is important to note that a proposal by scientists largely overlapping with the DarkSide-20k collaboration has recently published (DarkSide-LowMass PRD 107, 12006 (2023)) addressing specifically the search for dark matter lighter than 10 GeV, with exactly the same physics scope as this paper, and also based on DarkSide-50 results. The reference is included in the paper.

This paper does not contain any physics results with new data but prospects based on successful results obtained with a previous smaller detector (DarkSide-50) extrapolated to a larger detector with several hardware differences. The methodology, analysis techniques, and detector and background models are essentially the same as in previous publications. The developments in the detector design/construction are not covered in full detail in this paper.

This may be acceptable for this particular Focus collection, but it seems in contradiction with the general publication criteria of Communications Physics. Therefore, I would rely on the editor’s decision on the suitability of this study for publication in Communications Physics.

My impression is that this paper contains few novel ingredients compared to analyses and studies already published and included in the references. In particular, the paper largely relies on previous published DarkSide-50 work. Many assumptions are based on this detector with an active volume a factor of 1000 smaller and a different readout and veto. The paper lacks a detailed assessment of the validity of the assumptions for DarkSide-20k.

Regarding the interest of the paper to the community, no new methods or ideas are put forward in the paper, compared to previous publications. Nonetheless, the conclusions can be of interest to the direct detection dark matter community since they present new expected limits by DarkSide-20k for light dark matter particles.

On the other hand, I consider that the paper is very well written, concise and clear. In case of publication, I think additional information should be added to the supplemental material (or to the paper) to better understand the details of the selection, analysis techniques and background model. In particular, it would be good to summarize the main assumptions that differ from DarkSide-50, the robustness of the assumptions for DarkSide-20k, and their impact on the presented prospects. There is a paragraph on this in the manuscript (lines 462-475) but I consider it rather insufficient.

I also think it would be interesting to point to ideas for further improvements in the detector, new measurements or analysis techniques that could straighten the conclusions or even improve the results.

Given the level of detail provided in the manuscript, I am not sure about the ability of a researcher to reproduce the results, as suggested by Communication Physics.

Other technical comments/questions:

- S1 undetectability: Fiducialization along the drift is not possible because of the undetected S1 signals. In order to lower the threshold, most of the selected events only have a S2 signal detected. Could you please comment on the S1 detection threshold? What is the expected scintillation light yield in DarkSide-20k? How are S1 signals produced in the region close the Photo Detector Modules distinguished from S2 signals?

- Neutron tagging and suppression: DarkSide-20k proposes an inner veto made of new material (PMMA doped with Gd) for neutron detection. What is the efficiency of this veto? How is this included in the analysis? This should be mentioned in the paper.

- Line 233: The S2 yield is about 25 photo-electrons per ionization electron extracted in the gas pocket. Where does this estimation come from? How is the performance of the Photo Detector Modules included in the study?

- Could you please comment on the assumed DarkSide-20k trigger efficiency at low energy and how this has been taken into account in the analysis? Some information should be included in the paper.

- Lines 261-262: It is mentioned that surface alpha-backgrounds can be efficiently suppressed. Could you please quantify the expected surface activities in DarkSide-20k and the contribution to the background rate?

- Line 285: “an exposure of 17.4 ton.year for one year of data” -> The exposure already has units. Adding “for one year of data” is misleading.

- Line 304: please add a more detailed explanation and/or a reference on how the 23% single electron response resolution is computed and to what extent this calculation is reliable for DarkSide-20k.

- Line 317: Is the reduction of 85Kr by a factor of 100 compared to DS-50 already demonstrated?

- It looks like the spurious single electrons, whose origin is still unknown, are the dominant backgrounds at the lowest energies. Is there any strategy in place to understand and possibly mitigate this background?

- Neutrons with energy in the MeV range can produce an energy deposit in the TPC that mimics a WIMP signal. In the paper, the nuclear recoil rate from radiogenic and cosmogenic neutrons is considered negligible compared to electron recoil backgrounds. The authors should quantify this statement in the paper.

- Improvement of figures: some colors used in Figures 2, 3, 4 and supplementary material are difficult to distinguish. The authors should add the names of the legend close to the lines in the plot.

- Fit strategy: what is the reason to extend the fit up to 170 Ne- and not to restrict it to the Ne- region where the signal is expected? Is the information from this higher energy region helping somehow the fit?

I would recommend this manuscript for approval once my previous comments and questions are addressed.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thank you for addressing all my comments. I recommend this manuscript for publication.

One remaining suggestion:

>> What exactly are the "expected limits" shown in all the plots?
> They are the usual 90% C.L. expected median limits, obtained on background-only pseudo-data.

I would recommend clarifying this in the manuscript, e.g. by changing 'expected' -> 'expected median' in an appropriate place. Without this, some readers may misunderstand "expected limit" as an average.

Reviewer #3

(Remarks to the Author)
Overall, the authors have addressed the questions I raised in my previous report.

Regarding my concerns about the poor novelty and originality of the paper, the editors clarified that a collection article has different criteria than a regular article in Nature Communications Physics, and therefore more relaxed standards may apply in a Focus collection.

For all these reasons, I support the publication of this collection article.

Rebuttal letter for manuscript COMMSPHYS-24-0785

Date: 15.10.2024

Dear reviewers and editor,

We thank you for your useful comments regarding our manuscript titled "DarkSide-20k sensitivity to light dark matter particles". We appreciate your inputs that definitely help to improve our manuscript. Please find below the list of changes made to the Figures and our answers to your comments. The changes made to the text are highlighted in the revised manuscript.

We look forward to hearing from you regarding our revised manuscript. We would be glad to respond to any further questions and comments that you may have.

Best regards,

the DarkSide-20k collaboration

List of changes to figures

Figure 2:

- The CEvNS model was not including the latest radiative corrections. This has been corrected, with a fully negligible impact on all results.
- Increased size of x- and y-axis labels.

Figure 3:

- Removed positive signal contours from CoGent and CDMS, as they have both been excluded later by the same collaborations.
- Changed color for PandaX-4T 2023 (from light pink to bright magenta).
- Changed x-axis tick from 1.2 to 1 GeV.
- Increased size of x- and y-axis labels.

Figure 4a:

- Changed color for PandaX-4T 2023 (from light pink to bright magenta).
- Increased size of x- and y-axis labels.

Figure 4b:

- Changed color for PandaX-4T 2023 (from light pink to bright magenta).
- Removed the limit from PandaX-II, which is not relevant to show anymore.
- Increased size of x- and y-axis labels.

Figure 4c:

- It was pointed to us that the Freeze-in benchmark has been recently corrected, see 2312.14152. The corresponding line and the associated reference have been updated accordingly.
- Changed color for PandaX-4T 2023 (from light pink to bright magenta).
- Removed the limit from PandaX-II, which is not relevant to show anymore.
- Increased size of x- and y-axis labels.

Figure 4d:

- Increased size of x- and y-axis labels.

Figure 4e:

- Increased size of x- and y-axis labels.

Figure 4f:

- The DS-50 limit, which was pointed to be wrong in Ref [16], has been properly corrected by the authors, who submitted an erratum to PRL. The limit has been updated accordingly.
- ‘NuSTAR’ has been removed from the legend on the right, and included directly in the plot with a down arrow to show that the limits are below the y-axis scale.
- Increased size of x- and y-axis labels.

Figure Supplementary 1:

- Removed positive signal contours from CoGent and CDMS, as they have both been excluded later by the same collaborations.
- Changed color for PandaX-4T 2023 (from light pink to bright magenta).
- Changed x-axis tick from 1.2 to 1 GeV.
- Increased size of x- and y-axis labels.

Figure Supplementary 2:

- Changed color for PandaX-4T 2023 (from light pink to bright magenta).
- Increased size of x- and y-axis labels.

Figure Supplementary 3:

- Added the +- 1 sigma and +- 2 sigma bands (brazilian bands).
- Removed positive signal contours from CoGent and CDMS, as they have both been excluded later by the same collaborations.
- Changed color for PandaX-4T 2023 (from light pink to bright magenta).
- Changed x-axis tick from 1.2 to 1 GeV.
- Increased size of x- and y-axis labels.

Figure Supplementary 4:

- Changed x-axis tick from 1.2 to 1 GeV.
- Increased size of x- and y-axis labels.

Reviewer #1 (Remarks to the Author):

This manuscript projects sensitivities for light dark matter searches with the DarkSide-20k experiment. Since the projections are world-leading for a range of benchmark models, they will be highly relevant for the field for several years. I expect to recommend publication if the authors can address some questions and comments.

{} Please add csv files or something similar to the supplemental material to distribute the (mass, cross-section) datapoints for the projected sensitivities. It also helps to add them to the arXiv preprint, if you update this.

→ We have added a data availability statement. The expected limits reported in Figs. 3-4 and Supplementary Figs. 1-3 can be found on the Zenodo repository: https://zenodo.org/records/13911875.

{} What exactly are the "expected limits" shown in all the plots? Are these mean or median background-only limits? And could you also indicate the band of expected limits in at least one of the figures -- traditionally the band between the +- 1 sigma quantiles of background-only limits is shaded green.

→ They are the usual 90% C.L. expected median limits, obtained on background-only pseudo-data. For illustration, we have added the +/- 1 sigma and +/- 2 sigma bands (brazilian bands) in supplementary Figure 3 (adding its description to the caption). For the other Figures, for which we show limits for both conservative and ultimate fit scenarios, the plot would be too crowded with overlapping bands.

{} l. 255: DS-20k will have a much larger drift time than DS-50, so electron clouds experience more diffusion and S2s will be wider. Thus, assuming the same resolving power in drift/depth as DS-50 seems unrealistic. The resolution may also be a nontrivial function of depth.

→ Only ~1.5% (resp. ~9%) of the background events in the Ne range between 4 and 20 (resp. 170) [see Figure 2] are due to external backgrounds (mainly from PDMs) in the bottom half of the TPC. Only these events could induce multiple S2 pulses (Ar39 background only induces single scatter events), which could be merged by diffusion effect. Among those events, we checked with simulations that less than 35% of them deposit energy in several S2 pulses separated by more than 2 mm along the drift direction. In conclusion, even without any resolving power in depth, the increase of the overall background would be around 0.5% (resp. 3%) in the Ne range below 20 (resp. 170). The impact on the limits is fully negligible. This has been added to the text at line 264.

{} l. 282: "no fiducialization is performed in the electron drift direction". In the large LXe detectors I am familiar with, S2-only analyses can effectively fiducialize in the drift direction -- at least for S2s larger than a couple electrons -- by a cut on the S2 width, exploiting diffusion of the electron cloud during drift. Is such a cut assumed here, or not?
If not, why do you not expect a large background from radioactive elements stuck on the cathode, or external backgrounds near the top of the detector, as we see in LXe detectors? If yes, it would be good to elaborate on how the signal acceptance and background rejection of such a cut is modeled (as the S2 width is a nontrivial function of S2 as well, for small S2s).

→ We do not perform any fiducialization along the drift direction, as stated in the paper. This leads to conservative limits. As for Xenon TPCs, this could be done by analyzing the S2 width in order to further reject backgrounds from top and bottom. However, external backgrounds from X- and gamma-rays are sub-dominant with respect to the internal 39Ar background (see Figure 2). They mainly come from the optical planes (mainly SiPMs, labeled ‘PDM’), from TPC walls as well as anode and cathode (labeled ‘TPC’). Note that the strategy of no fiducialization along the drift direction was similar in the DS-50 data analyses, while the contribution from the classical photo-multipliers was higher than the one foreseen from the SiPMs.

{} l. 267-9: Why would cuts against accidental coincidences -- if I understand them correctly as events with unrelated S1 + S2 signals -- reduce the effective livetime of an S2-only analysis like this by half? Perhaps this refers instead to cuts of livetime following a large S2, to reduce the few-electron background?

We acknowledge that the original statement was unclear. To clarify, we neglect S1 pulses in this analysis because, at very low energies, a significant part of them are undetectable, and we prefer to avoid introducing an S1 efficiency curve. The cut mentioned in the text (±3.7 ms isolation) is designed to reject multiple scatter events. Since we cannot distinguish between correlated pulses and random pile-ups that occur within the maximum drift time window, we reject any S2 pulses within this time frame, regardless of their size. This leads to a deadtime of 49%. We have revised the sentence to: “*The accidental coincidences within the 3.7 ms isolation window between the S2 pulses induced by this background and by the signal result in an effective livetime of 51%.”*

{} l. 317: Is there a reference for this factor 100?

→ This (conservative) factor is an educated guess by the experts based on the UAr extraction plant design and simulations. This plant will have multiple distillation column system, allowing to get rid of 85Kr. There is unfortunately no reference for it. Actual measurements will be available next year.

{} l. 322 - 324: it is not clear to me what "a linear uncertainty on such effects" means for the model. Perhaps a reference to another DS-20k paper suffices here, otherwise, a plot of the assumed spectral shape and its uncertainty (all the way to 0 eV) would be helpful.

→ We rephrased to “*(...) with a 200 eV threshold [24, 25]. Below this value, a linearly increasing uncertainty on the corresponding corrections is assumed, reaching 25% at 0 eV [14]*”. This makes clearer how the uncertainty was assigned. Details can be seen in the inset of Figure 4 of PRD 107 (2023) 063001, where the shaded area corresponds to this uncertainty. We have added this reference to the sentence.

{} l. 343: Is there a reference for this factor 2.5?

→ The factor 2.5 is the outcome of our study, integrating the external X- and gamma-ray backgrounds from Figure 2 (labeled ‘PDM’, ‘TPC’ and ‘Vessel’), and comparing to the DS-50 analysis (see Figure 7 of PRD 107 (2023) 063001, curves labeled ‘PMTs’ and ‘Cryostat’).

{} l. 350: As detailed in the suggestion below, I would here immediately say something about how this background is much more uncertain than the others, and that you will consider two different scenarios (ultimate / conservative). I would also put the sentence that the model is the one from [18] in front, before the short summary of the model.

→ The two scenarios refer to the fit strategy, described in section “Results and discussion”, and not to the spurious electron background modeling. So we prefer not to introduce them there. The SE model is from a dedicated study performed with the DS50 data, updating [18], that will give rise to a dedicated paper – currently at work within the collaboration. Only the scaling from DS-50 to DS-20k is the same as in [18].

According to your suggestion, we clearly state that we assume the same ionization spectrum for these events in DS-20k and in DS-50 and a sentence has been added to the paper at line 366 to clarify this: "*This background will need to be thoroughly characterized once real DS-20k data is available.*"

{} Table II / l. 408-9: It would help to clarify these are not the actual uncertainties you currently have on the parameters, but the uncertainties you _expect_ to have at the start of the analysis. The actual uncertainty on some parameters (like the 39Ar activity) would be much larger and difficult to quantify; you account for this by choosing conservative assumptions (and you report on some variations of the nominal conditions later in the paper).

→ This has been clarified at line 422: "*Systematic uncertainties that are believed to be reachable at the time of data taking …*"

{} l. 435-438: As mentioned above, a figure or reference to such that details the 39Ar spectral shape model would be appreciated to help contextualize this remark.

→ As answered above, this can be seen in Figure 4 of PRD 107 (2023) 063001, where the inset details the uncertainties at low energy. This reference has been added to the modified sentence at line 330, where the model is described.

{} l. 441: It may be worth clarifying that the 90% C.L. CLs procedure you chose has a similar median sensitivity as the two-sided Feldman-Cousins approach recommended in [31] and used by several of the results you compare with. (A 90% one-sided search without CLs would have a better median sensitivity, and would not be a completely fair comparison.)

→ As mentioned in the text, we set the 90% CL exclusion limits following the guidelines outlined in https://iopscience.iop.org/article/10.1088/0954-3899/28/10/313, which is the same methodology employed by the DarkSide-50 collaboration in their 2023 publication. It is important for us to remain consistent with the exclusion limits established by DS50 in order to maintain coherence across results, since this is our main point of comparison across the paper.

{} l. 453-455: Here it may be worth clarifying that the impact of the quenching fluctuations, and the uncertainty on the availability of the SE model (encoded in the ultimate vs conservative projection), is even larger than this.

→ For SE, a sentence has been added to the paper at line 366, as already mentioned. We also added a statement at line 439: "*given the unknowns on the SE background before data taking (...)*".

For quenching fluctuations, we moved the sentence referring to the NQ results (originally in "Detector response model" section) here, where it indeed fits better.

Finally, we have added the following sentence at the end of the paper: "*Further improvements in the sensitivity of DarkSide-20k could be achieved through dedicated measurements aimed at quantifying the fluctuations in quenching for nuclear recoils or identifying the processes underlying the generation of spurious electrons, thereby enabling their effective suppression.*"

{} Figure 3: Are the positive signal contours just shown for historical purposes? If I understand correctly (see e.g. https://arxiv.org/pdf/1506.03924 and elsewhere), the CoGent excess is now well-understood as being due to an underestimated background. For both CoGent and CDMS, I believe later limits from the same collaborations rule out a WIMP interpretation of their earlier results.

→ Indeed, the positive signal contours from CoGent and CDMS have both been excluded later by the same collaborations, so we have removed them from Figure 3 (and consistently in supplementary material Figures).

{} Figure 4:
* I would not use nearly-invisible-pink for PandaX-4T (also not in in figure 3), since it is leading in the bottom two figures on the left (and figure 3). Perhaps use the bright magenta that is currently used for PandaX-II, which does not seem very relevant to show anymore.

→ We changed the color for PandaX-4T from light pink to bright magenta, and removed the limit from PandaX-II. This has been done consistently for Figure 3 and for Figures in supplementary material.

* The lower right panel reserves a color for NuSTAR in the legend, but I did not see it in the plot. Perhaps this information is better communicated with some text inset in the plot. It is also a bit strange to feature a benchmark model that has been so heavily excluded -- perhaps a spin-dependent dark matter model would be more interesting? Or is there a reason why these limits remain relevant after the NuSTAR result?

→ 'NuSTAR' has been removed from the legend on the right, and included directly in the plot with a down arrow to show that the limits are below the y-axis scale. These are astrophysical constraints (indirect and model-dependent limits), so this is relevant to report on our potential to set complementary direct detection limits. We clarified this in the caption: "*The indirect detection limits set by the NuSTAR experiment [67], which looks for anomalous X-ray lines from radiative sterile neutrino DM decays, extends downwards to $|U_{e4}|^2=10^{-13}$ at 20 keV/c^2.*"

* Please encourage the authors of [16] to publish an erratum that you can then cite here.

→ This has been done, and we will add the reference as soon as we have it.

* I found myself writing captions with the different subfigures with a pen, e.g. "Migdal + NR" or "Dark photons", after scanning the y-axes and captions. You could consider saving other readers the trouble. Using a larger font for the axes and labels might also help.

→ We have increased the font size of the x- and y-axis labels (for all figures of the paper), improving the readability of the Figures, in particular Figure 4 which has 6 plots.

[Below I grouped some related suggestions with a general comment.]

The manuscript shows sensitivities under several different scenarios, corresponding to conservative and optimistic scenarios for the main unknowns:
* how LAr responds to low-energy nuclear recoils (quenching fluctuations or not);
* whether the spurious electron background appears as expected and can be accurately modeled;
* whether the Migdal effect takes place as expected.

Although it would be more convenient to have a single projection that is "realistic but on the conservative side", I agree with the author's choice here -- the uncertainties are large and quite impactful, and opinions differ on how to best treat them. However, given this, I would give the reader more information to make up their minds on which model they would like to use. Specifically:

{A} It would help to reference (planned) efforts to resolve the uncertainties. Are there plans (by Darkside-20k or others) to measure the NR yield fluctuations in LAr? Similarly, while the spurious electron background will likely have to be characterized in-situ, references to efforts to characterize this background might bolster the case that the "ultimate" projection is attainable.

→ There are several ongoing efforts involving part or the entire collaboration (e.g., ReD, BLEND, X-ArT) using neutron sources, guns or neutron beams below 100 keV, aimed at improving the NR quenching factor and fluctuations and at characterizing spurious electrons as a function of contaminants. ReD data analysis is underway, and results are expected in a few months while other projects have a timeline of 2-3 years. Therefore, they could directly impact the DS-20k models. However, there are currently no citable references for these efforts. We have added the following sentence at the end of the paper: "*Further improvements in the sensitivity of DarkSide-20k could be achieved through dedicated measurements aimed at quantifying the fluctuations in quenching for nuclear recoils or identifying the processes underlying the generation of spurious electrons, thereby enabling their effective suppression.*"

{B} I would recommend being clearer about the limitations of the SE background model. The authors assume the empirical model from reference [18], which has been fitted to Darkside-50 data. I doubt this model has predictive power for the 1000x larger detector that DarkSide-20k expects to be. For example, reference [18] assumes the single electron rate scales linearly with event rate, which would make sense in some range around the DS-50 observations. For a 80/1.5 ~ 50x increase in event rate, I would think it depends on exactly how delayed single electrons are distributed in time following their progenitor, on which [18] is inconclusive. The model in [18] was also derived for the particular event selections of DS-50, which may have to be substantially changed in DS-20k.

→ As mentioned above, the unknown on SE have been underlined at lines 366 and 439 according to your suggestions. We are currently concluding an analysis of DarkSide-50 data focused on spurious electrons (SEs), which has allowed us to update the model used in reference [18]. This analysis, which will be published in the coming months, demonstrates that the number of electrons per SE (also referred as SE multiplicity), follows a Poisson distribution with a constant mean, independent of the energy deposited by previous events or drift time (please keep this information confidential).

Under the same assumption of reference [18] - i.e. SEs result from impurities that capture and re-emit electrons during the drift process- the rate of SEs is extrapolated by scaling the DS-50 one as in Ref. [18].

We have therefore employed this updated model in our analysis.

{C} For 39Ar, the authors assume the same activity as in Darkside-50. This is conservative. However, they note that "significant improvement in sensitivity is expected if the UAr extraction plant would further reduce the contamination of 39 Ar". It may be worth briefly commenting (or referencing) plans to such improvements -- even though the authors do not here assume they will come to pass. This helps the reader judge how likely it is that the sensitivities will be significantly better than those quoted here.

→ Assuming the same 39Ar activity in DS-20k UAr than in DS-50 is our baseline, as it will be extracted from the same deep mine. As you correctly point out, this is conservative. From DS-50 UAr extraction, observation of residual oxygen might point to an accidental air infiltration in the plant, which could have inflated the residual 39Ar contamination. This will be quantified next year, when UAr extraction will start.

{D} For Migdal, you could consider adding some statement (ideally with references) that experimental efforts to measure this effect are under way and/or that they are inconclusive at the moment. Again, this explains to readers why you show two plots rather than just one.

→ We have added such a statement at line 395: "*As experimental efforts to confirm the existence of such effects in nuclear scattering are still underway [35–37], WIMP signals with and without them (...) are considered in this article.*". This points to 3 references of experimental efforts to measure the Migdal effect.

Reviewer #2 (Remarks to the Author):

This paper describes a light dark-matter sensitivity study for the next-generation detector of the DarkSide liquid-argon direct dark-matter search program, DarkSide-20k. It estimates achievable limits for the expected experimental parameters using a simulation model and background estimates informed by results from the smaller DarkSide-50 experiment.
This is a fairly bread-and-butter type of paper, not containing any new experimental physics result. Nevertheless, such a documented study of projected sensitivity is a useful addition to the literature, in the context of a community of numerous competing experiments with different approaches.

The DarkSide program is by now quite mature. I do not spot obvious errors or inaccuracies in the provided information, although most details are embedded in references.

I recommend publication.

I have the following minor suggestions:

Does "Clevios" need a "TM" or the like?

→ Good point, the ™ symbol has been added after "Clevios".

Fig 1: I suggest toning down (or turning into grey) the colors for items not referenced in the caption (for example, more than one thing is pink).

→ We rephrased the caption of Figure 1 to avoid any confusion: "*Cross-section of the DS-20k detector showing at its center the TPC, with its acrylic walls in green and electrodes in pink. The stainless steel vessel surrounding it is shown in gray, immersed in the ProtoDUNE-like cryostat.*"

It is worth citing the recent PANDA-X and XENON measurements of the neutrino "fog" in the relevant section.

→ This has been added at line 372, referencing both XENON-nT and PandaX-4T preprints.

Line 360: "backgrounds to the DM search" ("background" depends on what you are looking for...)

→ According to your suggestion, "to the DM search" has been added after "Other backgrounds".

Table II: most of the uncertainty sizes are conveyed in the pre-fit plot and presumably details are in the reference. Can you indicate the approximate size of the "atomic exchange and screening" contribution? Presumably that is combined with shape uncertainties related to the uncertain Q-value in the pre-fit background spectra (it is of interest given the statement about eventual limiting 39-Ar beta shape uncertainties).

→ The size of the atomic exchange and screening contribution is energy–dependent but the most significant impact is at low energy (<10 keV). For example it amounts to ~50% at 0.1 keV. The caption of Table II refers to PRD 107 (2023) 063001, where Figure 4 shows the 39-Ar ionization spectrum for DarkSide-50. Details can be seen in the inset, where the shaded area corresponds to the uncertainty from atomic exchange and screening effects. The description of the model has been rephrased at line 330, adding this reference: "*(...) with a 200 eV threshold [24, 25]. Below this value, a linearly increasing uncertainty on the corresponding corrections is assumed, reaching 25% at 0 eV [14].*"

Line 435:
-> "As 39-Ar is the dominant background, its spectral shape will"

→ This sentence has been modified according to your suggestion.

Reviewer #3 (Remarks to the Author):

The paper "DarkSide-20k sensitivity to light dark matter particles" assesses the sensitivity of the future DarkSide-20k to models predicting light dark matter particles below 10 GeV. The authors claim at least one order of magnitude improvement in dark-matter interaction cross-sections for WIMPs with respect DarkSide-50, which is the predecessor of DarkSide-20k. The analysis focuses on exclusion limits for spin-independent WIMP nucleon recoil with quenching fluctuations, although other signals and no quenching are also considered.

As mentioned in the article, DarkSide-20k is designed to perform a nearly instrumental background-free search for WIMP masses above 20 GeV. This paper represents an effort to understand the capabilities of the future DarkSide-20k in the region below 10 GeV. It is important to note that a proposal by scientists largely overlapping with the DarkSide-20k collaboration has recently published (DarkSide-LowMass PRD 107, 12006 (2023)) addressing specifically the search for dark matter lighter than 10 GeV, with exactly the same physics scope as this paper, and also based on DarkSide-50 results. The reference is included in the paper.

This paper does not contain any physics results with new data but prospects based on successful results obtained with a previous smaller detector (DarkSide-50) extrapolated to a larger detector with several hardware differences. The methodology, analysis techniques, and detector and background models are essentially the same as in previous publications. The developments in the detector design/construction are not covered in full detail in this paper.

This may be acceptable for this particular Focus collection, but it seems in contradiction with the general publication criteria of Communications Physics. Therefore, I would rely on the editor's decision on the suitability of this study for publication in Communications Physics.

My impression is that this paper contains few novel ingredients compared to analyses and studies already published and included in the references. In particular, the paper largely relies on previous published DarkSide-50 work. Many assumptions are based on this detector with an active volume a factor of 1000 smaller and a different readout and veto. The paper lacks a detailed assessment of the validity of the assumptions for DarkSide-20k.

Regarding the interest of the paper to the community, no new methods or ideas are put forward in the paper, compared to previous publications. Nonetheless, the conclusions can be of interest to the direct detection dark matter community since they present new expected limits by DarkSide-20k for light dark matter particles.

On the other hand, I consider that the paper is very well written, concise and clear. In case of publication, I think additional information should be added to the supplemental material (or to the paper) to better understand the details of the selection, analysis techniques and background model. In particular, it would be good to summarize the main assumptions that differ from DarkSide-50, the robustness of the assumptions for DarkSide-20k, and their impact on the presented prospects. There is a paragraph on this in the manuscript (lines 462-475) but I consider it rather insufficient.

→ DarkSide-20k analysis strategy is directly adapted from DarkSide-50 low mass search paper. We think that all the details of the selection are already given in the "selection" section. The same analysis technique is used by considering the "number of electrons" as the only discriminating variable. The background modeling of the two main components (spurious electron and the 39-Ar) is directly coming from DarkSide-50 and adapted to the DarkSide-20k case.

Concerning the detector assumptions, for your convenience here is a short summary table of the main relevant characteristics for the low-mass DM search as measured for DS-50 and assumed for DS-20k. Note that all these numbers are already in the paper draft

Detector	DS-50	DS-20k
Electron lifetime (ms)	16	16
Single electron resolution (%)	27	23
39Ar activity (mBq/kg)	0.73	0.73
85Kr activity (mBq/kg)	1.9	0.019

Concerning their impact on the presented prospects, we have added at line 484 when discussing the impact of the assumptions: "*The electron lifetime, single electron response and $x - y$ resolutions and have 39Ar activity have been varied by a factor 2 with respect to the nominal assumptions*."

I also think it would be interesting to point to ideas for further improvements in the detector, new measurements or analysis techniques that could straighten the conclusions or even improve the results.

→ We have added the following sentence at the end of the paper: "*Further improvements in the sensitivity of DarkSide-20k could be achieved through dedicated measurements aimed at quantifying the fluctuations in quenching for nuclear recoils or identifying the processes underlying the generation of spurious electrons, thereby enabling their effective suppression.*"

Given the level of detail provided in the manuscript, I am not sure about the ability of a researcher to reproduce the results, as suggested by Communication Physics.

Other technical comments/questions:

- S1 undetectability: Fiducialization along the drift is not possible because of the undetected S1 signals. In order to lower the threshold, most of the selected events only have a S2 signal detected. Could you please comment on the S1 detection threshold? What is the expected scintillation light yield in DarkSide-20k? How are S1 signals produced in the region close the Photo Detector Modules distinguished from S2 signals?

→ We neglect S1 pulses in this analysis because, at very low energies, a significant part of them are undetectable, and we prefer to avoid introducing an S1 efficiency curve. You can appreciate the S1 threshold as observed in DS-50 in Figure 2 of PRD 107 (2023) 063001. It lies at O(1 keV_er). The scintillation yield in DarkSide-20k is expected to be around 20%, i.e. similar to DS-50 which was measured to be 16%.

As for the unresolved S1+S2 pulses from the top of the TPC, they do not pose any issue in this analysis, as their rate is approximately two orders of magnitude lower than the trigger rate and, within the energy window of our analysis, over three orders of magnitude lower than the background of electron recoils (ERs). That said, in cases where the S1 and S2 pulses cannot be resolved, their pile-up would be interpreted as a single S2 pulse, with the contribution of photoelectrons from S1 being minimal compared to S2. This scenario would introduce only a negligible bias in the energy reconstruction.

- Neutron tagging and suppression: DarkSide-20k proposes an inner veto made of new material (PMMA doped with Gd) for neutron detection. What is the efficiency of this veto? How is this included in the analysis? This should be mentioned in the paper.

→ Neutron contribution is negligible, as their rate is approximately O(1-10) NRs/day from radiogenic and cosmogenic sources. This is significantly lower, more than 4 orders of magnitudes, compared to the expected rate of ERs. As a consequence, the efficiency of the neutron veto, which is not employed in the analysis, has no impact on the results.

- Line 233: The S2 yield is about 25 photo-electrons per ionization electron extracted in the gas pocket. Where does this estimation come from? How is the performance of the Photo Detector Modules included in the study?

→ This is based on simulations of the electric field and the length of the gas pocket. This g2 factor was included in the Monte Carlo simulation, taking into account full optical propagation, instrumental effects related to the SiPMs (photo-detection efficiency, cross-talks, after pulses, dark count rate), DAQ emulation, and reconstruction effects. The optical properties in the Monte Carlo, as well as the noise in the electronics, were introduced based on actual measurements.

- Could you please comment on the assumed DarkSide-20k trigger efficiency at low energy and how this has been taken into account in the analysis? Some information should be included in the paper.

→ DarkSide-20k operates as a triggerless system, with waveforms processed in real-time using matched filtering to optimize the S/N ratio. Integrated information, such as charge and timing from SiPMs, is extracted, and the pulse finder achieves an efficiency close to 100%. This has been added at line 248: "*DarkSide-20k adopts a triggerless data acquisition system ensuring to save all SiPM signals with close to 100% efficiency for single photo-electrons.*"

- Lines 261-262: It is mentioned that surface alpha-backgrounds can be efficiently suppressed. Could you please quantify the expected surface activities in DarkSide-20k and the contribution to the background rate?

→ Surface alpha events from the lateral walls in DarkSide-20k produce a large S1 signal but a small S2 signal, as most electrons are absorbed by the walls themselves due to distortions of the electric field. We do not yet have all the elements for a definitive estimate of this background. However these events are easily identified and rejected using the S2/S1 ratio, as successfully done in DS-50, where the residual alpha background was negligible. Additionally, any surviving electrons are reconstructed near the edges in the xy plane and further rejected through fiducialization, which is by far more effective in DS-20k due to the larger TPC size compared to DS-50. Please note that alphas from the cathode induce in average large S2 pulses, above the range of interest for this analysis, and therefore are neglected.

- Line 285: "an exposure of 17.4 ton.year for one year of data" -> The exposure already has units. Adding "for one year of data" is misleading.

→ To quantify the exposure that we target to reach in one year of data taking, we need to precise "for one year of data" (it would be different to reach 17.4 ton.year exposure e.g. after 10 years of data taking). In the paper, we give results for both 1 year and 10 years of data taking (i.e. exposures of 17.4 t.y and 174 t.y).

- Line 304: please add a more detailed explanation and/or a reference on how the 23% single electron response resolution is computed and to what extent this calculation is reliable for DarkSide-20k.

→ The resolution is obtained through a full optical Geant4 simulation of thermal electrons extracted in the gas pocket of DS-20k, followed by the simulation of the electronics (including waveform simulation and SiPM random and correlated noises), and by the emulation of the DAQ up to the raw data. It is important to stress that the optical properties in the Monte Carlo, as well as the noise in the electronics, were introduced based on actual measurements, as already stated in the paper draft. Finally, we applied the reconstruction algorithms developed specifically for DS-20k.

- Line 317: Is the reduction of 85Kr by a factor of 100 compared to DS-50 already demonstrated?

→ This (conservative) factor is an educated guess by the experts based on the UAr extraction plant design and simulations. This plant will have multiple distillation column system, allowing to get rid of 85Kr. There is unfortunately no reference for it. Actual measurements will be available next year..

- It looks like the spurious single electrons, whose origin is still unknown, are the dominant backgrounds at the lowest energies. Is there any strategy in place to understand and possibly mitigate this background?

→ There is work ongoing, both using DS-50 data and dedicated small LAr TPC set-up. The goal of this set-up is to understand the origin of this SE background. Hopefully, this will result in mitigation strategies that could be applied to DS-20k. To acknowledge these ongoing efforts, as already mentioned, we have added the following sentence at the end of the paper: "*Further improvements in the sensitivity of DarkSide-20k could be achieved through dedicated measurements aimed at quantifying the fluctuations in quenching for nuclear recoils or identifying the processes underlying the generation of spurious electrons, thereby enabling their effective suppression.*"

- Neutrons with energy in the MeV range can produce an energy deposit in the TPC that mimics a WIMP signal. In the paper, the nuclear recoil rate from radiogenic and cosmogenic neutrons is considered negligible compared to electron recoil backgrounds. The authors should quantify this statement in the paper.

→ As in DarkSide-50, we neglect neutrons, as their rate is approximately O(1-10) NRs/day from radiogenic and cosmogenic sources. This is significantly lower, more than 4 orders of magnitudes, compared to the expected rate of ERs, which justifies our statement. We have added this in the paper: "*the rate of NRs from radiogenic and cosmogenic neutrons is expected to be over four orders of magnitude lower than the ER one, and therefore not considered in this analysis.*"

- Improvement of figures: some colors used in Figures 2, 3, 4 and supplementary material are difficult to distinguish. The authors should add the names of the legend close to the lines in the plot.

→ We have tried to improve the readability of all Figures, changing colors for a few lines and increasing the font size of the x- and y-axis labels.

- Fit strategy: what is the reason to extend the fit up to 170 Ne- and not to restrict it to the Ne- region where the signal is expected? Is the information from this higher energy region helping somehow the fit?

→ Extending the fit to such high electron multiplicity allows to better constrain the background model, in particular the external X- and gamma-ray backgrounds which peak at high Ne. The same fitting strategy has been used for the DS-50 analyses (see PRD 107 (2023) 063001), for the same reason.

I would recommend this manuscript for approval once my previous comments and questions are addressed.

Rebuttal letter for manuscript COMMSPHYS-24-0785

Date: 27.11.2024

Dear reviewers and editor,

We thank you for your last comments regarding our manuscript titled "DarkSide-20k sensitivity to light dark matter particles". Please find below the one change made to the text following your last comment. All other changes are related to editorial requests that are detailed in the Editorial Requests Table.

Best regards,

the DarkSide-20k collaboration

Reviewer #1 (Remarks to the Author):

One remaining suggestion:

>> What exactly are the "expected limits" shown in all the plots?
> They are the usual 90% C.L. expected median limits, obtained on background-only pseudo-data.

I would recommend clarifying this in the manuscript, e.g. by changing 'expected' -> 'expected median' in an appropriate place. Without this, some readers may misunderstand "expected limit" as an average.

→ We have added the word ‘median’ when presenting the first results shown in Figure 3.