Global fits of axion-like particles to XENON1T and astrophysical data

Peter Athron,[a,b] Csaba Balázs,[b] Ankit Beniwal,[c] J. Eliel Camargo-Molina,[d]
Andrew Fowlie,[b] Tomás E. Gonzalo,[b] Sebastian Hoof,[e] Felix Kahlhoefer,[f]
David J.E. Marsh,[g,e] Markus Tobias Prim,[h] Andre Scaffidi,[i] Pat Scott,[j,d] Wei Su,[k]
Martin White,[k] Lei Wu[b] and Yang Zhang[b,l]

[a]*Department of Physics and Institute of Theoretical Physics, Nanjing Normal University,*
Wenyuan Road, Nanjing, Jiangsu 210023, China

[b]*School of Physics and Astronomy, Monash University,*
Wellington Rd, Melbourne, VIC 3800, Australia

[c]*Centre for Cosmology, Particle Physics and Phenomenology (CP3),*
Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

[d]*Department of Physics, Imperial College London,*
Blackett Laboratory, Prince Consort Road, London SW7 2AZ, U.K.

[e]*Institut für Astrophysik, Georg-August-Universität Göttingen,*
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

[f]*Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University,*
D-52056 Aachen, Germany

[g]*Department of Physics, King's College London,*
Strand, London WC2R 2LS, U.K.

[h]*Physikalisches Institut der Rheinischen Friedrich-Wilhelms-Universität Bonn,*
Nussallee 12, 53115 Bonn, Germany

[i]*Istituto Nazionale di Fisica Nucleare, Sezione di Torino,*
via P. Giuria 1, I-10125 Torino, Italy

[j]*School of Mathematics and Physics, The University of Queensland,*
St. Lucia, Brisbane, QLD 4072, Australia

[k]*ARC Centre of Excellence for Dark Matter Particle Physics & CSSM,*
Department of Physics, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

[l]*School of Physics, Zhengzhou University,*
No. 100 Science Avenue, ZhengZhou 450001, China

E-mail: peter.athron@coepp.org.au, csaba.balazs@monash.edu,
ankit.beniwal@uclouvain.be, eliel.camargo-molina@physics.uu.se,
andrew.j.fowlie@njnu.edu.cn, gonzalo@physik.rwth-aachen.de,
hoof@uni-goettingen.de, kahlhoefer@physik.rwth-aachen.de,
david.j.marsh@kcl.ac.uk, markus.prim@cern.ch,
andre-joshua.scaffidi@to.infn.it, pat.scott@uq.edu.au,
wei.su@adelaide.edu.au, martin.white@adelaide.edu.au,
leiwu@njnu.edu.cn, zhangyangphy@zzu.edu.cn

ABSTRACT: The excess of electron recoil events seen by the XENON1T experiment has been interpreted as a potential signal of axion-like particles (ALPs), either produced in the Sun, or constituting part of the dark matter halo of the Milky Way. It has also been explained as a consequence of trace amounts of tritium in the experiment. We consider the evidence for the solar and dark-matter ALP hypotheses from the combination of XENON1T data and multiple astrophysical probes, including horizontal branch stars, red giants, and white dwarfs. We briefly address the influence of ALP decays and supernova cooling. While the different datasets are in clear tension for the case of solar ALPs, all measurements can be simultaneously accommodated for the case of a sub-dominant fraction of dark-matter ALPs. Nevertheless, this solution requires the tuning of several *a priori* unknown parameters, such that for our choices of priors a Bayesian analysis shows no strong preference for the ALP interpretation of the XENON1T excess over the background hypothesis.

Contents

1 Introduction

The XENON Collaboration recently reported an excess of electronic recoil events over known backgrounds [1]. The statistically preferred explanation in the original analysis was that the excess is due to solar axion-like particles (ALPs) with a significance of about 3.5σ over the background-only hypothesis. This anomaly has already garnered considerable interest [2–40]. However, it was quickly noted that a solar ALP explanation is in conflict with astrophysical observations, including stellar evolution and cooling [8, 16, 41–43], SN1987A [8, 16], pulsating White Dwarfs (WDs) [8], and the predicted mass of astrophysical black holes [44], although this tension can be reduced in more complicated ALP scenarios [41, 45]. Interestingly, WD cooling presents a different anomaly that can also be

explained by axions, but the preferred axion couplings appear to be in conflict with the results of XENON1T.

A large number of physics scenarios have been put forward to explain the excess in the electronic recoil spectrum observed by XENON1T. One set of options is based around the existence of dark matter (DM) particles that either scatter inelastically in the detector [4, 5, 12, 21, 37, 46–48] or are boosted to semi-relativistic velocities via some other process before scattering elastically off electrons [10, 15, 17, 49–52]. A $2\,\mathrm{keV} - 3\,\mathrm{keV}$ dark photon with a small (10^{-15}) kinetic mixing with ordinary photons [2, 16, 53–55] or a massive dark photon produced from solar emission [6, 11, 54] (with the caveat of ref. [54]) may also explain the excess. Weakly-interacting relativistic bosons that are produced by the annihilation or decay of heavier dark particles have also been proposed to account for the data [9, 54, 56, 57]. Alternatively, the anomaly might result from new neutrino-lepton or non-standard neutrino interactions mediated by a light scalar or a vector particle [3, 13, 14, 45, 58–62]. Yet more potential explanations include exotic radioactivity affecting hydrogen decays [25], fermionic DM with an electric dipole moment (EDM) sourced by an oscillating axion-like field [63], and the resurrection of the solar ALP explanation via the postulate of a "stellar basin" of gravitationally-bound axions in the Sun [64]. Finally, tritium (^{3}H) [1, 65] or argon [66] present in the detector material have also been identified as possible causes, though the latter has since been excluded by the XENON Collaboration in a revised version of their initial submission [1].

Here we focus on ALP explanations for the excess similar to the ones originally considered by XENON1T. We consider solar ALPs and a scenario recently proposed by ref. [19], in which ALPs constitute a fraction of the local DM. This latter setup can potentially reconcile the different ALP-electron couplings favoured by XENON1T and by WD cooling hints, since the DM ALP signal in XENON1T scales with the DM fraction of ALPs. Hence, if only a fraction of the DM is made out of ALPs, the ALP-electron couplings favoured by XENON1T can be large enough to simultaneously explain the anomalous WD cooling.

In this work, we carefully investigate the impact of the XENON1T electronic recoil data on solar and DM ALPs using the GAMBIT global fitting software [67–69], considering Xe data and existing astrophysical constraints on ALPs previously considered in ref. [70], including a careful treatment of WD cooling hints. However, because of the systematic uncertainties inherent to these constraints, we also provide results without WD cooling hints. We include inverse Primakoff processes, as recently pointed out and examined by refs. [71, 72], and consider the potential ^{3}H background. We consider the impact of the Xe data on the ALP parameter space and the extent to which, when taken in combination with astrophysical data, it favours or disfavours ALP models.

We analyse the data by simply comparing best-fit likelihoods, by using the best-fit likelihoods in a frequentist analysis and by Bayesian methods. Reporting best-fit likelihoods and differences between them allows for a simple presentation of our findings without adopting either frequentist or Bayesian methods. In our frequentist analysis, we then use these results to compute confidence intervals and make estimates of p-values. Finally, we compute Bayes factors, which tell us how to update our belief in ALPs relative to the background model in light of all the data, and partial Bayes factors, which tell us how to

update our belief in light of the Xe data, supposing we already knew about the astrophysical data. A controversial aspect of any such analysis is that the findings are sensitive to our choices of prior (further discussion and details are provided in the appendices).

The paper is structured as follows: section 2 presents the ALP models that we analyse, section 3 discusses the various experimental results that enter into our analysis, along with their individual impacts, and section 4 describes the combined impact of all the constraints. Finally, we conclude in section 5. For the interested reader, we further discuss Bayes factors and our Bayesian numerical methods in appendix A, and our choices of prior and the prior dependence of our results in appendix B. In appendix C, we discuss and compute an alternative Bayesian measure, namely the Deviance Information Criterion. Lastly, we discuss the Monte Carlo simulations used in the frequentist analysis in appendix D.

2 Models

In order to examine the plausibility of ALP models explaining the excess events, we compare the "solar ALP" and "DM ALP" models introduced in detail below to a background-only model, in which we set all ALP couplings to zero. In the case of the XENON1T electronic recoil data, this corresponds to the known backgrounds included by the XENON Collaboration as described in section 3.1. We furthermore include an additional background contribution from tritium, which we discuss in section 2.3.

2.1 Solar ALP model

In order to compare directly to the analysis of ref. [1], and also to investigate the broadest possible parameter space, we consider ALPs with three independent couplings: to photons ($g_{a\gamma}$), electrons (g_{ae}), and nucleons (g_{aN}^{eff}).[1] The axion mass, m_a, is not a parameter in our solar ALP model, since the axions produced in the Sun are relativistic, $E_a \gg m_a$. We recall that for the QCD axion, all three couplings are linearly related to m_a. However, even for the QCD axion, there is considerable variation in the values of the couplings between different models, in particular for g_{ae} which can be loop suppressed [74]. A general ALP model is defined as one in which the couplings do not obey any particular relation to one another, and QCD axion models are a subset of the general ALP models considered here.

The XENON1T signal prediction for the solar ALP case consists of the Atomic recombination and de-excitation, bremsstrahlung, and Compton (ABC), Primakoff (denoted by "P"), and ^{57}Fe ("Fe") fluxes. These can either be deposited in the detector via the axio-electric effect ("aee") or, as pointed out in follow-up studies [71, 72], via the inverse Primakoff effect ("iP"). The latter was not considered in the original XENON1T analysis. The individual components are scaled by the effective axion couplings, i.e., we can calculate

---

[1]In this paper, we augment the GAMBIT GeneralALP model of ref. [70] with the ALP-nucleon coupling. We also add the new Xe likelihood functions and include corrections relevant at larger ALP masses in the R parameter and WD likelihoods. These will be made available in release version v2.0 of GAMBIT. YAML files for the scans performed in this work are available on Zenodo [73].

them at a reference scale. Schematically,

$$
\begin{aligned}
s = g_{ae}^2 \cdot \left(g_{ae}^2 \cdot s_{\text{ABC}}^{\text{aee}} + g_{a\gamma}^2 \cdot s_{\text{P}}^{\text{aee}} + (g_{aN}^{\text{eff}})^2 \cdot s_{\text{Fe}}^{\text{aee}} \right) \\
+ g_{a\gamma}^2 \cdot \left(g_{ae}^2 \cdot s_{\text{ABC}}^{\text{iP}} + g_{a\gamma}^2 \cdot s_{\text{P}}^{\text{iP}} + (g_{aN}^{\text{eff}})^2 \cdot s_{\text{Fe}}^{\text{iP}} \right) ,
\end{aligned}
\tag{2.1}
$$

where s is the signal, the subscripts denote the production channel in the Sun, and the superscripts denote the detection channel. We take the ABC, Primakoff and ^{57}Fe signal components, and backgrounds, from figure 1 of ref. [1]. We compute the inverse Primakoff contributions following ref. [72].

2.2 Dark matter ALP model

ALPs are viable DM candidates with a large parameter space spanning many orders of magnitude in mass and coupling [75, 76]. We consider three parameters for the DM ALP model: the coupling to electrons (g_{ae}), the ALP mass (m_a), and the fraction of the (local) DM around the Earth (η) that is made up of ALPs. We ignore the g_{aN}^{eff} coupling for the DM ALP model since it is not involved in the detection channels. While the $g_{a\gamma}$ coupling could potentially be relevant because of the inverse Primakoff process, it needs to effectively be zero as a result of the x-ray constraints discussed in section 3.4. We therefore ignore this possibility. We comment on the relation between solar ALPs and DM ALPs at the end of this section.

Given that the local DM moves non-relativistically with velocities of the order $10^{-3}c$, i.e., $E_a \simeq m_a$, we neglect the velocity effect in the DM ALP signal strength, which is given by

$$
s = 0.841 \, \text{t}^{-1}\text{yr}^{-1} \left(\frac{\eta \, \rho_0}{0.4 \, \text{GeV/cm}^3} \right) \left(\frac{m_a}{3 \, \text{keV}} \right) \left(\frac{\sigma_{\text{pe}}(m_a)}{1.68 \times 10^{-19} \, \text{cm}^2} \right) \left(\frac{g_{ae}}{10^{-14}} \right)^2 ,
\tag{2.2}
$$

where σ_{pe} is the photoelectric cross section which we adopt from ref. [77] (who use results from ref. [78]) and ρ_0 is the local DM density for which we use the constraints implemented in ref. [68], i.e., we use a log-normal distribution with a median value of $0.40(15) \, \text{GeV/cm}^3$. Note that this results (in general) in a value that is larger than the value of $\rho_0 = 0.3 \, \text{GeV/cm}^3$ considered in ref. [1].

The required relic abundance of cold keV DM ALPs can be produced in the early Universe by the non-thermal vacuum realignment mechanism if the scale of spontaneous symmetry breaking, f_a, is of order $10^{10} \, \text{GeV}$, assuming a temperature-independent ALP mass, and the standard thermal history up to $T \sim f_a$.

A keV ALP can also be produced thermally by the freeze-in mechanism, in which case it constitutes a warm DM component [79, 80]. The allowed abundance of warm DM is constrained by the observed Lyman-alpha flux power spectrum, which favours $\eta \lesssim 0.1$ for $m_a \sim 1 \, \text{keV}$ [81–83]. For further discussion of the production mechanisms relevant to this scenario, and the warm DM limits, see ref. [19]. Several explicit models for a DM ALP with the required mass and Standard Model couplings are given in refs. [19, 45].

In general, solar ALPs *could* be DM ALPs at the same time, but in our scenarios we assume that the solar and DM ALPs that explain XENON1T have rather different masses.

To explain the excess events in XENON1T, we must consider electron recoil energies of more than about $1 \, \mathrm{keV}$. As DM ALPs in the halo are non-relativistic, this implies that $m_a \gtrsim 1 \, \mathrm{keV}$ as well. Solar ALPs are usually considered to be much lighter since the production of ALPs at the keV scale or above, i.e. the typical energy scales of the processes inside the Sun, would be suppressed. For the Primakoff flux,[2] and a recent solar model [84], we estimate that the total integrated axion flux in the energy range relevant for XENON1T is reduced by about 27% (70%) for an ALP mass of $3 \, \mathrm{keV}$ ($5 \, \mathrm{keV}$) compared to effectively massless ALPs. While this would allow heavier DM ALPs to also be produced inside the Sun and influence the statistical inference on the values of the ALP couplings, we assume that we can treat the two hypotheses as distinct scenarios.

Lastly, we note that our DM ALP cannot be the QCD axion: among many constraints, a keV QCD axion has a lifetime shorter than the age of the Universe.

2.3 The tritium background hypothesis

Let us now comment on the possible presence of a relevant ^{3}H background in the XENON1T experiment, which could give rise to an excess of events at about $1 \, \mathrm{keV} - 15 \, \mathrm{keV}$. The XENON Collaboration quote a conservative upper limit of $4 \times 10^{-20} \, \mathrm{mol/mol}$ for the ^{3}H abundance relative to Xe from exposure to cosmic rays, but expect it to be reduced to at most $10^{-27} \, \mathrm{mol/mol}$ after purification. Ref. [65] suggests that these numbers should be considered uncertain by an order of magnitude. The XENON Collaboration furthermore discuss other mechanisms by which tritium may be introduced to the detector, possibly at most at the $10^{-26} \, \mathrm{mol/mol}$ level. For reference, fitting the anomaly with a tritium component requires about $5 \times 10^{-25} \, \mathrm{mol/mol}$.

In light of these uncertainties, in our Bayesian analyses we make a conservative treatment of the tritium level, employing a log-normal prior for the tritium fraction α_t

$$\log_{10}\left(\frac{\alpha_t}{1 \, \mathrm{mol/mol}}\right) = -27 \pm 3 \qquad (2.3)$$

with a central value at the upper estimate of the level of tritium and a moderate standard deviation. Of course, the XENON Collaboration itself would be better placed to construct a prior describing plausible levels of tritium in the detector. Another possible choice of prior would be a log-uniform prior between the expected ^{3}H levels after purification and the amount expected from cosmic rays, effectively encoding the assumption that the purification process was inefficient to an unknown degree. In our frequentist analyses, we permit an unconstrained tritium component, following the XENON1T methodology.

3 Experimental constraints and hints

As the ALP interpretation of the XENON1T anomaly is in tension with astrophysical observables, we now discuss some of these as well as our implementation of the relevant

---

[2]While the Primakoff contribution is not relevant in this case due to the x-ray constraints, we would have to revise the calculation for all parts of the ABC flux and the ^{57}Fe flux in order to consider heavier solar ALPs. The extension of the signal calculation is beyond the scope of this work.

Data	GAMBIT capability
Astrophysical	
R parameter [86] with He abundance from [87]	`lnL_RParameter`
WD cooling (G117-B15A) [88]	`lnL_WDVar_G117B15A`
WD cooling (R548) [89]	`lnL_WDVar_R548`
WD cooling (L19-2) [90]	`lnL_WDVar_L192`
WD cooling (PG 1351+489) [91]	`lnL_WDVar_PG1351489`
XENON1T	
Eq. (3.1) for binned data from electronic recoils [1]	`lnL_XENON1T_Anomaly`

Table 1. Likelihoods included in the analysis. Note that the Xe likelihood and the high-mass corrections for the astrophysical likelihoods will be made available in a future release of the GAMBIT software.

likelihood functions. We later show the impact of the observables in their entirety in section 4. For a discussion and recent global analyses considering astrophysical constraints, see e.g., refs. [70, 85]. The likelihood functions used in this analysis are summarised in table 1.

3.1 XENON1T

We implement the XENON1T likelihood (hereafter "Xe likelihood") from the binned data between 1 and 30 keV as made available by the XENON Collaboration on Zenodo [92]. We infer an exposure of about 0.6473 t yr from the Zenodo data. Our likelihood is the product of Poisson distributions,

$$\mathcal{L}_{\text{Xe}} = \prod_{i=1}^{29} \frac{\lambda_i^{o_i} e^{-\lambda_i}}{o_i!}, \qquad \lambda_i = \epsilon \cdot (\alpha_b b_i + \alpha_t t_i + s_i), \qquad (3.1)$$

where o_i are the observed counts; s_i are the binned signal predictions; b_i are the binned backgrounds other than tritium, which are scaled by a factor α_b; and t_i is the binned tritium background, scaled by the tritium fraction α_t (relative number of atoms w.r.t. Xe atoms; in units of mol/mol). The overall expected events are scaled by the efficiency ϵ. The efficiency and the background scale α_b are varied with Gaussian uncertainties 0.03 and 0.026, respectively, which were estimated from refs. [1, 93]. We note that additional possible contributions to the background [94] are not included.

We calculate the binned signals using the energy resolution for XENON1T as determined in ref. [95] and the detection efficiencies from Zenodo [92]. We emphasise that XENON1T in fact perform an unbinned analysis, which is expected to have higher statistical power than what can be done with publicly available data.

In figure 1, we compare our implementation to the 90% CL curves as shown in ref. [1]. Note that we do not include the inverse Primakoff contribution for the purpose of this comparison. The similarity between the curves validates that our binned likelihood approximates the unavailable, unbinned Xe likelihood. We find a best-fit point at $\hat{g}_{ae} =$
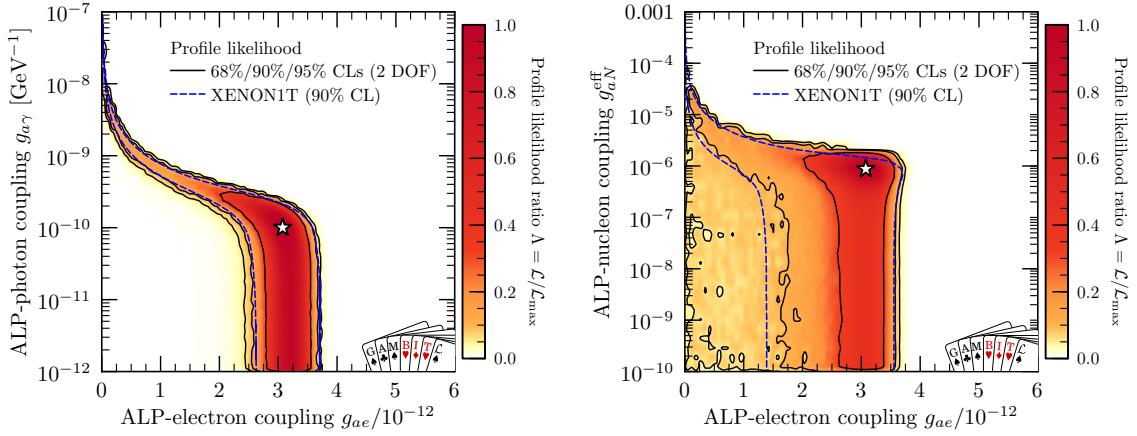
Figure 1. Comparison between our 68%/90%/95% CL regions (solid black lines) and the XENON1T 90% CL region (dotted blue lines) for the solar ALP effective couplings. The inverse Primakoff contribution is not included. Stars denote our best-fit point.

3.07×10^{-12}, $\hat{g}_{a\gamma} = 1.07 \times 10^{-10}\,\mathrm{GeV}^{-1}$, $\hat{g}_{aN}^{\mathrm{eff}} = 9.08 \times 10^{-7}$, $\hat{\alpha}_b = 0.98$ and $\hat{\epsilon} = 0.98$, which corresponds to $\chi^2 = 29.5$, compared to $\chi^2 = 44.0$ for the background model (see section 4).[3]

3.2 Horizontal and Red Giant Branch stars

One of the most robust constraints on axions and ALPs comes from the lifetime of stars [96] such as Horizontal Branch (HB) and Red Giant Branch (RGB) stars. The stellar plasma is transparent to weakly-coupled ALPs, so that once they are produced, they easily escape the star, leading to an additional cooling channel. These theoretical constraints can be turned into a likelihood by counting the number of HB and the number of RGB stars in e.g. Galactic globular clusters. The ratio of their numbers, the so-called R parameter, has been used to place strong constraints on the ALP-photon and ALP-electron coupling. We use a likelihood based on results from ref. [86], which was first implemented in ref. [70]. Note that we include the rather small correction for this likelihood that arises for higher ALP masses [97]. We hereafter refer to this as the R likelihood.

The R likelihood is the most robust astrophysical constraint that we consider in the sense that it is not affected by large systematic uncertainties. The measurement itself is based upon a large number of systems, which show good agreement with each other, suggesting that the measurement error is statistic dominated [98]. The theory prediction relies on the helium abundance, which introduces an additional potential source of error. However, ref. [87] argues that systematic uncertainties related to this measurement are

---

[3]After including the inverse Primakoff contribution, we find best-fit values of $\hat{g}_{ae} = 2.86 \times 10^{-12}$, $\hat{g}_{a\gamma} = 1.28 \times 10^{-10}\,\mathrm{GeV}^{-1}$, $\hat{g}_{aN}^{\mathrm{eff}} = 8.55 \times 10^{-7}$, $\hat{\alpha}_b = 0.98$ and $\hat{\epsilon} = 0.98$ with $\chi^2 = 29.2$. It has been observed in follow-up works [71, 72] that after including the inverse Primakoff contribution, the XENON1T best-fit regions move closer to the region in which astrophysical constraints are satisfied. This has been interpreted as a reduction of the tension. However, that fact that the value of $\chi^2 = 43.4$ of the best-fit point in the Xe + R likelihood combination analysis stays the same — with or without including the inverse Primakoff contribution — already indicates that the tension is not relieved significantly after including it.

under control. Hence, the R likelihood allows for a consistent statistical interpretation in the context of global fits. Our results from combining XENON1T and the R likelihood show that the latter puts the solar ALP interpretation of the XENON1T anomaly into question. Indeed, the R likelihood dominates over the XENON1T likelihood in the g_{ae}-$g_{a\gamma}$ plane of the parameters. This forces the solar ALP best-fit couplings to occupy a degenerate line away from values that can explain the excess. Note that not combining the R parameter with the XENON1T likelihood in the solar ALP case would be inconsistent since both consider ALP interactions with stellar systems.

For an impression of the importance of the R likelihood, consider the solar ALP model. With only XENON1T, we find a best-fit of $\chi^2 = 29.3$. With XENON1T and the R parameter, we find a best-fit of $\chi^2 = 43.4$, a considerable increase, despite only adding one additional data point to the fit. Indeed, the increase by about $\Delta\chi^2 = 14$ indicates that although g_{aN}^{eff} (with a best-fit at $\hat{g}_{aN}^{\text{eff}} = 1.08 \times 10^{-5}$) is not constrained by the R parameter, it cannot alleviate the tension between XENON1T and the R parameter, mostly because the ^{57}Fe signal associated with it is monochromatic and only contributes to the spectrum near $14.4\,\text{keV}$, whereas the excess is observed below $7\,\text{keV}$.

3.3 White Dwarf cooling hints

Similar to the XENON1T anomaly, observations of anomalous cooling in WDs can be interpreted as being due to ALPs with non-vanishing ALP-electron coupling g_{ae}. Indeed, measurements of the period increase in a number of pulsating WDs show anomalous cooling that is consistent with an ALP-electron coupling of $g_{ae} \sim \mathcal{O}(10^{-13})$ [86]. Another observable that can be used to infer the WD cooling is the white dwarf luminosity functions, see e.g., ref. [99].

Here we use a likelihood based on the findings of refs. [88–91], first implemented and described in ref. [100]. Similar to the R likelihood, we have to include correction terms for higher ALP masses (see e.g., ref. [101]). We estimate the WDs internal temperature from their astroseismological properties in refs. [88–90, 102] using Kramer's opacity law [96, section 2.2.2]. The typical corrections at $m_a = 3\,\text{keV}$ increase the ALP-electron coupling by a factor of 1.4 (except for PG1351+489, which has a higher internal temperature than the others and the correction is less than 1.1).

More specifically, for light ALPs (e.g., the solar ALPs that we consider) "WD cooling hints" therefore point to a value of $g_{ae} \sim 3.4 \times 10^{-13}$, which is an order or magnitude *lower* than the coupling expected to fit the XENON1T anomaly with solar ALPs [1]. This fact, together with the importance of the R parameter constraint mentioned above, takes our combined best fit point for solar ALPs to a region in significant tension with the XENON1T anomaly.

The situation is reversed for the DM ALP case. Here, the required value of $g_{ae} \sim 3.7 \times 10^{-13}$ to fit the cooling hints[4] is *larger* than the constraints placed by the XENON1T experiment (assuming that ALPs make up all of the local DM) [1]. This constraint can be evaded if ALPs are allowed to only constitute a fraction of the local DM.

---

[4]This value is slightly bigger due to the aforementioned corrections for higher masses around $m_a \sim 3\,\text{keV}$.

It should be noted that the WD cooling hints are somewhat speculative due to the difficulties involved in both the measurement and the modelling of WD evolution. Nonetheless, just as with the XENON1T anomaly, it is interesting to consider the consequences of the WD cooling anomaly. When included, the combined likelihood of the four WDs considered constitutes a strong constraint that can dominate the statistical inference on the ALP-electron coupling.

3.4 DM ALP decays

If ALPs constitute some or all of DM, their decays into photons would lead to potentially observable x-ray lines. The strongest constraints in the mass range of interest stem from observations of M31 [103] and from NuSTAR [104] and require

$$g_{a\gamma} \lesssim 10^{-16} \, \text{GeV}^{-1} \left(\frac{m_a}{1 \, \text{keV}} \right)^{-3/2} \eta^{-1/2} \, , \tag{3.2}$$

which is many orders of magnitude stronger than the constraints from stellar cooling. However, the ALP-photon coupling is not necessary to explain either the XENON1T anomaly or the WD cooling hint, so we will assume that the ALP-photon coupling for the DM ALP case is sufficiently suppressed (see refs. [19, 80] for how to obtain the necessary suppression in explicit models) to satisfy this constraint, and we set $g_{a\gamma} = 0$ explicitly.

Note that ignoring the $g_{a\gamma}$ coupling will have an impact on the statistical inference compared to including it together with the x-ray constraints mentioned above. In both frequentist (via the model DOFs) and Bayesian (via an Occam's penalty) analyses, the presence of a strongly constrained parameter will disfavour the ALP hypothesis — unless that parameter vanishes or can at least be suppressed sufficiently in a given model. By ignoring the ALP-photon coupling for the DM ALP case, we therefore assume that such models can be constructed in a natural way. This might be overly optimistic but also means that if the DM ALP model is disfavoured in this setting, it would perform even worse with $g_{a\gamma}$ as a free parameter.

3.5 SN1987A cooling

We end this section with a note on the possible impact of further cooling constraints from supernova SN1987A, though as later explained, we do not include SN1987A in our statistical analysis. SN1987A constrains axions and ALPs in numerous ways [96] such as the conversion of axions in interstellar magnetic fields, the decay of ALPs, and the neutrino cooling time. For the present work, the latter would be the most relevant one as it presents one of the strongest constraints on the ALP-nucleon coupling, g_{aN}^{eff}.

Unfortunately, the usually cited theoretical cooling bound has so far not been cast into a statistical framework. One of the difficulties is that the uncertainties on the observed neutrino flux from SN1987A are large and ALPs hardly affect its value, as noted in ref. [105].[5] It thus appears that statistical statements about the bound would require full

---

[5]In fact, the authors of ref. [105] challenge the validity of the bound itself by appealing to an alternative model for the delayed neutrino burst; see ref. [106] for further discussion.

supernova simulations including ALPs to find which ALP-nucleon couplings are consistent with SN1987A happening at all.

Although the lack of a likelihood function prevents us from including this bound in our analysis, let us now nonetheless discuss its possible implications on the XENON1T anomaly. The bound, usually quoted as

$$|g_{aN}^{\mathrm{SN}}| \lesssim 0.9 \times 10^{-9}\,, \tag{3.3}$$

appears to exclude the XENON1T anomaly best-fit ALP-nucleon coupling, $\hat{g}_{aN}^{\mathrm{eff}} \approx 10^{-6}$, by several orders of magnitude.[6] However, the ALP-nucleon coupling only impacts the ^{57}Fe component of the signal, a monochromatic feature at around $14.4\,\mathrm{keV}$, whereas it is the ABC and Primakoff components that could explain the XENON1T excess. Setting $g_{aN}^{\mathrm{eff}} = 0$, we find that the minimum χ^2 value for the solar ALP case would change from $\chi^2 = 29.2$ to 30.9, i.e., the effect of the SN1987A constraint on the solar ALPs hypothesis would be small.

Importantly, though, the SN1987A bound could prevent the QCD axion from playing the role of a solar ALP that explains the XENON1T excess. This is most easily seen by noting that the SN1987A limit on g_{aN}^{eff} places the strongest upper bound on the QCD axion mass, $m_a \lesssim 10^{-2}\,\mathrm{eV}$. This is incompatible with the smallest value of the QCD axion mass allowed by the XENON1T 90% CL region, which occurs for the DFSZ model with $g_{a\gamma} \approx 10^{-11}\,\mathrm{GeV}^{-1}$, leading to $m_a \approx 4 \times 10^{-2}\,\mathrm{eV}$. However, variations of the QCD axion model [108] with smaller $g_{a\gamma}$ remain compatible.

4 Results

We now turn to the detailed discussion of the results from our frequentist and Bayesian analyses. Our choice of priors and the prior sensitivity of our conclusions for the Bayesian analyses are summarised in appendix B. In our frequentist analyses the priors reflect the sampling strategy of the parameter scans and we anticipate limited impact.

4.1 Solar ALPs

Frequentist results. First, we consider frequentist results for the solar ALP model for only Xe data, when adding the R parameter and finally when adding WD hints. In table 2, we show the $\chi^2 \equiv -2 \log \max \mathcal{L}$ at every step and the log-likelihood ratio test statistic,

$$\Delta\chi^2 \equiv 2 \log \frac{\max \mathcal{L}_{s+b}}{\max \mathcal{L}_b}, \tag{4.1}$$

where $\max \mathcal{L}_{s+b}$ is the maximum likelihood in the ALP + background hypothesis and $\max \mathcal{L}_b$ is the maximum likelihood in the background hypothesis. In each case, this involves maximising the likelihood over several parameters. With just Xe data and without

---

[6]Note that the effective ALP-nucleon coupling for SN1987A [107],

$$|g_{aN}^{\mathrm{SN}}| = \sqrt{|g_{an}^2 + 0.53\, g_{an}\, g_{ap} + 0.61\, g_{ap}^2|}\,, \tag{3.4}$$

written in terms of the ALP-neutron (g_{an}) and ALP-proton (g_{ap}) couplings, differs from the effective ALP-nucleon coupling g_{aN}^{eff} considered by XENON1T. Despite these subtleties, we can still compare the effective ALP-nucleon couplings in order of magnitude.

Model	Xe		Xe + R		Xe + R + WD	
	χ^2	$\Delta\chi^2$	χ^2	$\Delta\chi^2$	χ^2	$\Delta\chi^2$
Background	44.0	0.0	45.0	0.0	66.6	0.0
Solar ALP	29.3	14.7	43.4	1.5	55.5	11.1
DM ALP	27.2	16.8	27.3	17.7	43.5	23.1
Background + ^{3}H	34.5	0.0	35.4	0.0	57.1	0.0
Solar ALP + ^{3}H	29.3	5.2	33.6	1.8	45.7	11.4
DM ALP + ^{3}H	25.9	8.5	26.0	9.4	42.2	14.9

Table 2. The χ^2 values associated with the best-fit points in our models for the Xe data, when adding the R parameter, and finally when adding the WD hints. The $\Delta\chi^2$ columns show the test statistic in eq. (4.1). For each model (solar or DM ALP) the test statistic is computed using the corresponding background (with or without ^{3}H) as the null hypothesis.

tritium, we see a $\Delta\chi^2 \simeq 15$ preference for solar ALPs over the background model. With an unconstrained tritium component, this reduces to only $\Delta\chi^2 \simeq 5$. When we include the R parameter, the incompatibility between XENON1T and the R parameter in the solar ALP model destroys the preference for solar ALPs and we find $\Delta\chi^2 \simeq 1.5$. Finally, adding the WD hints restores some preference for solar ALPS, $\Delta\chi^2 \simeq 11$. As we see in the recoil energy spectra in figure 2, in the latter case, the solar ALP model is barely distinguishable from the background model in XENON1T, as the ABC and Primakoff contributions must be suppressed to satisfy the R parameter constraint. The ^{57}Fe component, however, remains visible above the background. As discussed in section 3.5, even the visible ^{57}Fe component could be ruled out by SN1987A.

As mentioned in section 3.2 and section 3.3, fitting both the R parameter and WD cooling hints requires ALP couplings away from the XENON1T preferred region as depicted in figure 1. This can be readily seen in figure 3, where we show profile likelihoods for the solar ALP model with constraints from XENON1T, the R parameter and WD cooling hints. For comparison, we also show the 90% CL from the results of a fit including only Xe data (with the inverse Primakoff contribution [72]). We note that our best fit point lies outside of the Xe 90% CL in both the g_{ae}-$g_{a\gamma}$ and g_{ae}-$g_{aN}^{\rm eff}$ planes.[7] This tension is significantly driven by stellar cooling, which imply $g_{ae} \lesssim 10^{-12}$ and $g_{a\gamma} \lesssim \mathcal{O}(10^{-10})$. For our best-fit point, the larger value of $g_{aN}^{\rm eff}$, outside the 90% CL for Xe only, can be understood by the need to compensate the very small value of $g_{a\gamma}$ in order to reproduce the right signal around 14.4 keV.

Bayesian results. For our Bayesian results, in table 3 we show the Bayes factors in favour of our solar ALP model when omitting a tritium component, including it in the ALP and background models, and when including it only in the background model. For example, the Bayes factor in the third row and third column ($B = 0.25$) is obtained by

---

[7]The best-fit point in figure 3a is depicted at $g_{a\gamma} \sim 10^{-15}$. This is just indicative since the likelihood flattens for smaller values of $g_{a\gamma}$. The actual best-fit point/region can be anywhere below e.g. the corresponding limit of $g_{a\gamma} < 1.7 \times 10^{-11}\,{\rm GeV}^{-1}$ (95% CL; 1 DOF).
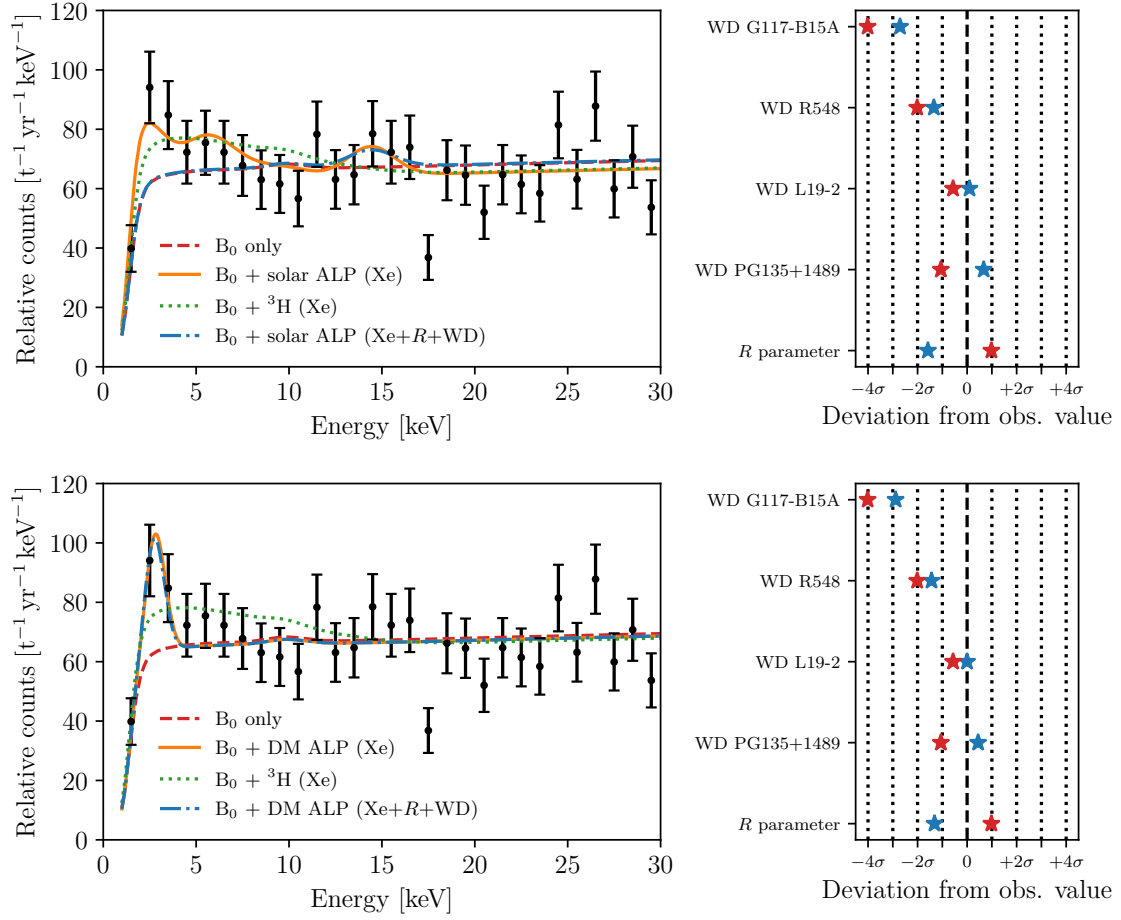
Figure 2. The best-fitting solar (top) and DM (bottom) ALP models. *Left:* we show the observed Xe data (black points and error bars) and the best-fit background (B$_0$; dashed red), B$_0$ + ^{3}H (dotted green), and B$_0$ + ALP models (solid orange). We also show the best-fit ALP + B$_0$ model for the XENON1T and astrophysical data (dashed-dotted blue). *Right:* a comparison of the best-fit predictions for the WD period decrease and R parameter for no ALP (red stars) and including ALPs (blue stars; with XENON1T and astrophysical data).

considering the ratio of probabilities

$$B = \frac{p(\text{Xe} + R + \text{WD} \,|\, B_0 + \text{solar ALP})}{p(\text{Xe} + R + \text{WD} \,|\, B_0 + {}^3\text{H})}, \tag{4.2}$$

i.e., for that entry, we consider the probability of the XENON1T, R parameter and WD data in the solar ALP model with the XENON1T B_0 background versus the probability of that data in the B_0 background model plus the added tritium component.

With only Xe data — and when omitting tritium backgrounds — we find a Bayes factor of about 3 in favour of the solar ALP scenario. This is "barely worth mentioning" on the Jeffreys' scale [109] and corresponds to a Z-score of about 0.6σ. With tritium backgrounds, the ALP model and backgrounds are barely distinguished, with a slight preference for the

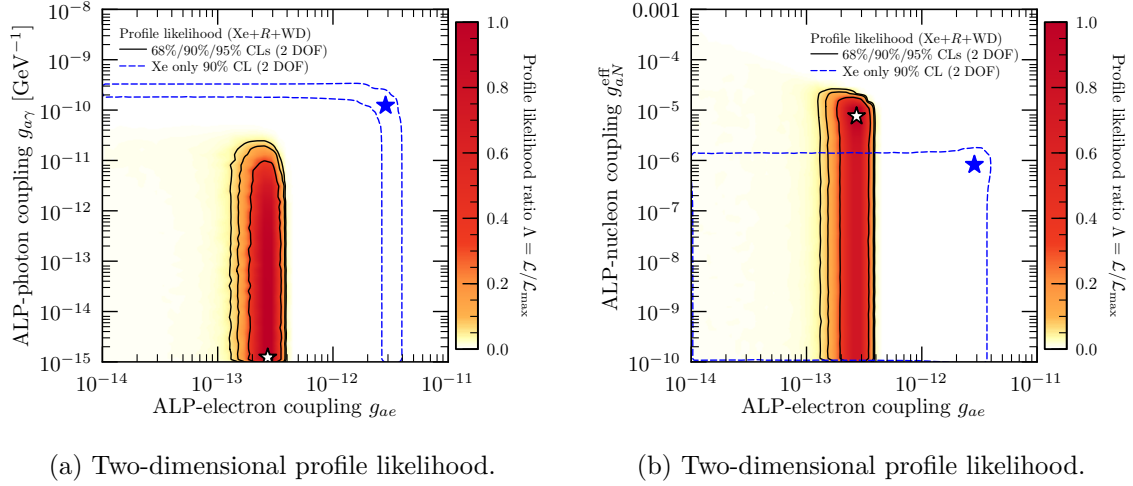(a) Two-dimensional profile likelihood. (b) Two-dimensional profile likelihood.

Figure 3. Profile likelihoods of solar ALP models for Xe + R + WD data for (a) the ALP-electron vs ALP-photon coupling and (b) the ALP-electron vs ALP-nucleon coupling. Blue dashed lines show the 90% CL limit from Xe data only. White (blue) stars denote the combined (Xe data only) best-fit point.

	Xe	(Xe + R)	(Xe + R + WD)	(Xe | R)	(Xe | R + WD)
No ^{3}H	2.7	0.26	1.3	0.99	0.92
^{3}H	0.64	0.27	1.0	1.0	0.73
^{3}H background only	0.52	0.051	0.25	0.19	0.18

Table 3. Bayes factor in favour of solar ALP versus background, with no tritium background (first row), tritium contributions to the background in both models (second row) and tritium contributions only in the background model (third row). The first three columns show results considering Xe, adding the R parameter and finally adding WD hints. The final two columns show partial Bayes factor for the Xe data given the R parameter, and given the R parameter and WD hints.

background only model. Lastly, the Xe data slightly favours the background-plus-tritium hypothesis over a signal without a tritium component.

The R parameter removes the slight preference for the solar ALP model. Indeed, with XENON1T and the R parameter, the solar ALP model is disfavoured by a factor of about 4. The inclusion of WD cooling hints slightly reverses the impact of the R parameter, but even then we find at most a tiny preference for the solar ALP model of about 1.3. Lastly, the partial Bayes factors for the Xe data given the astrophysical data were less than about one; meaning that given we knew the astrophysical data, the XENON1T excess in fact told us little about the solar ALP model. This happened because the astrophysical data forced the solar ALP couplings into regions that could not explain the XENON1T excess, making its predictions for the data observed by XENON1T no better than the background model, and in fact worse than a tritium component.

We emphasize that the Bayes factors that we obtain depend sensitively on our choice of priors. A detailed discussion of how our results would change for different priors is given in
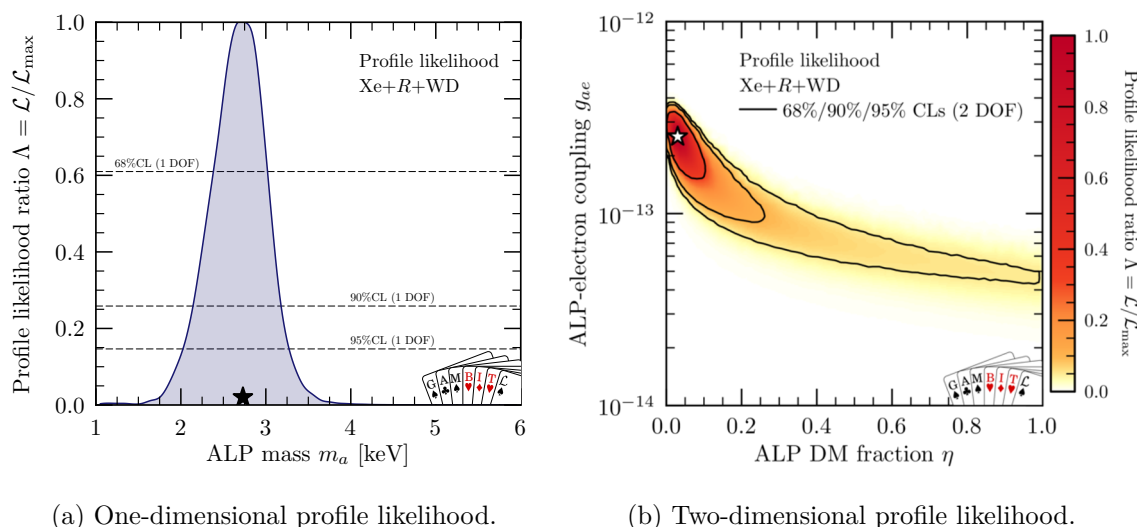
(a) One-dimensional profile likelihood. (b) Two-dimensional profile likelihood.

Figure 4. The profile likelihoods with Xe + R + WD data for (a) the ALP DM mass and (b) the ALP DM fraction and ALP-electron coupling. Stars indicate the best-fit point.

appendix B. We find that the maximum Bayes factor obtainable for any priors for the solar ALP couplings is roughly $e^{\Delta\chi^2/2} \approx 1500$, which corresponds to a Z-score of about 3.2σ. As an alternative approach for model comparison not directly dependent on the priors, we discuss the Deviance Information Criterion in appendix C.

4.2 DM ALPs

Frequentist results. Turning to DM ALPs, our frequentist results for the DM ALP mass, electron coupling and DM fraction — when combining the Xe, R, and WD likelihoods — are shown in figure 4. In figure 4a, we see that only ALP masses close to the best-fit point $\hat{m}_a = 2.7\,\mathrm{keV}$ are favoured. Figure 4b shows that smaller DM fractions permit greater electron couplings, with $\eta \lesssim 0.1$ favoured, and the best-fit point at $\hat{\eta} = 0.030$. While this has been appreciated before [19], here we show the confidence intervals compatible with the Xe likelihood, which might be of particular interest for model builders. Even at 90% CL, ALP DM is not the dominant DM component, where all our results include the uncertainties on the local DM density.

In figure 5, we show the DM ALP frequentist results and the various individual observables considered in this work. If the local DM density consists entirely of ALPs ($\eta = 1$) we can derive a bound $g_{ae} \lesssim 10^{-13}$ for any given ALP mass. By allowing the ALP DM abundance η to vary, greater couplings are allowed, and we can obtain a decent fit to the Xe data, while at the same time fitting the combined cooling hints and respecting the R parameter constraint ($\chi^2 = 43.5$). Note that the combination of the four WD cooling hints prefers ALP masses smaller than about $3\,\mathrm{keV}$, even though heavier ALPs can also contribute to WD cooling for larger values of g_{ae}. To illustrate this, we include the profile likelihood for only the WD G117-B15A in figure 5 as an example.
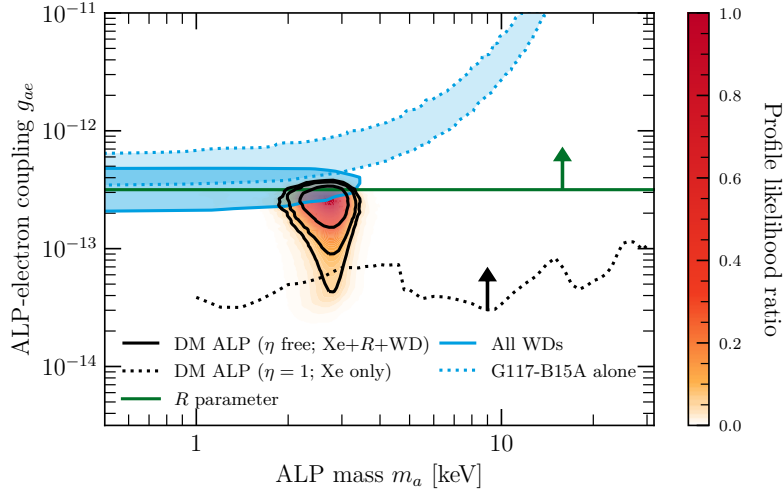
Figure 5. The profile likelihood (yellow-red density plot) and 68%/90%/95% CL regions (solid black; 2 DOF) for the DM ALP mass and DM ALP-electron coupling with Xe + R + WD data. For context, we show the 90% CL constraints from Xe assuming $\eta = 1$ (dotted black), the R parameter (green), and the region hinted by a combination of WDs (shaded blue region, solid line). We also show one WD (G117-B15A) alone for comparison (shaded blue region, dotted line). The arrows point towards the excluded regions.

In table 2, we show best-fit χ^2 and $\Delta\chi^2$ for the DM ALP model when adding the astrophysical data step-by-step. With only Xe data, the DM ALP model best-fit improves on the background model by $\Delta\chi^2 \simeq 17$, which is slightly greater than in the solar ALP case. Even with a tritium component, the DM ALP model is preferred by $\Delta\chi^2 \simeq 8.5$. For the solar ALP model, we saw severe tension between the R parameter and XENON1T. Here, however, we see that adding the R parameter in fact slightly increases the preference for DM ALPs to $\Delta\chi^2 \simeq 18$. Indeed, the DM ALP model successfully reconciles XENON1T and the R parameter by tuning the ALP DM fraction. Lastly, adding WD cooling hints further increases the preference for DM ALPs, reaching $\Delta\chi^2 \simeq 23$ without tritium, and $\Delta\chi^2 \simeq 15$ with tritium. The success of the model can be seen in figure 2 by the similarity between the best-fit spectra for XENON1T only, and to XENON1T and astrophysical data. Rather than reduce the allowed signal, the astrophysical data in fact push the amplitude of the signal slightly higher.

To make a rough estimate of the significance of the combined signals in the DM ALP model from XENON1T, R likelihood, and WD cooling we use Wilks' theorem [110], which states that in certain cases the log-likelihood ratio test statistic in eq. (4.1) should follow a χ^2 distribution. We compute local significances assuming 2 degrees of freedom, corresponding to the number of independent parameters of the DM ALP model except for the DM ALP mass. When including a tritium component and considering all data simultaneously, the observed value $\Delta\chi^2 = 14.9$ corresponds to a local significance of about 3.3σ. With no contribution from tritium the observed value $\Delta\chi^2 = 23.1$ corresponds to a local significance of 4.3σ.

(a) One-dimensional posterior probability.

(b) Two-dimensional posterior probability.

(c) Two-dimensional posterior probability.

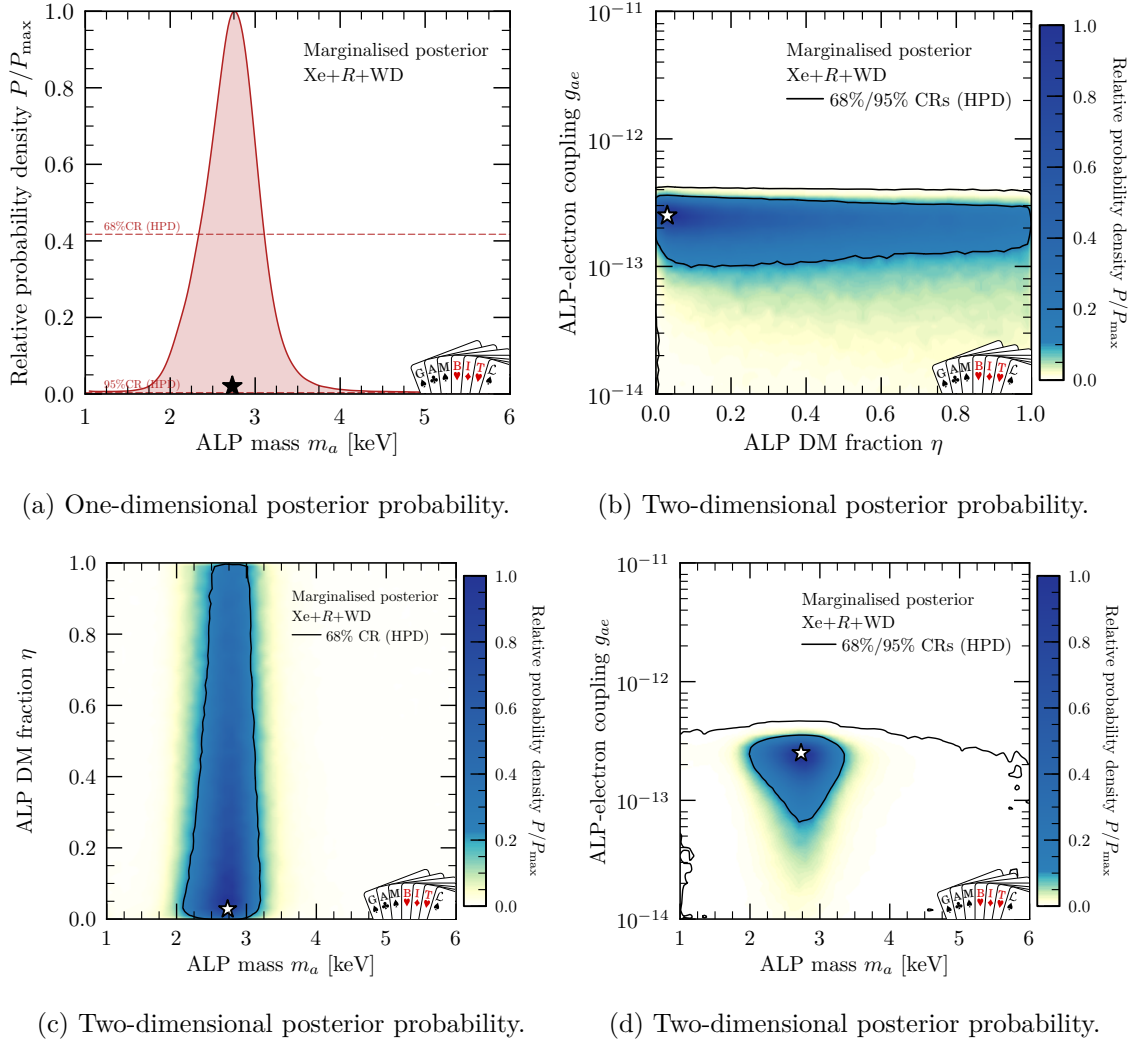(d) Two-dimensional posterior probability.

Figure 6. Posteriors with Xe + R + WD data for (a) the ALP DM mass, (b) the DM fraction and ALP-electron coupling, (c) the ALP DM mass and DM fraction and (d) the ALP DM mass and ALP-electron coupling. We should credible regions of highest posterior density (HPD), while stars indicate the best-fit point.

However, we warn that although our models are nested, important assumptions in Wilks' theorem are violated [111] — for example, the values we obtain are local significances that do not include a trial factor to account for the look-elsewhere effect for the DM ALP mass. Moreover, the background only model lies on the boundary of the ALP models, meaning that even our computation of the local p-values violate conditions in Wilks' theorem. In appendix D we provide a robust estimate of the global p-values using Monte Carlo simulations. The simulations largely confirm our estimates above, implying that our approximate local p-values are in fact accidentally closer to the global p-values (that account for the look-elsewhere effect) than one might naively have expected.

Bayesian results. Lastly, we present our Bayesian results for the DM ALP model. The posterior probability densities and 68% and 95% credible regions for the DM ALP parameters are shown in figure 6 for XENON1T, the R parameter and WD data. As expected from the profile likelihood results in figure 4, the DM ALP mass is strongly constrained to be around $m_a \sim 2.7\,\text{keV}$. However, in contrast to the frequentist results, DM ALP masses are allowed throughout the explored ranges, $1\,\text{keV} < m_a < 30\,\text{keV}$ at 95% credibility. Furthermore, the preference for a small DM fraction $\eta < 0.1$ is minimal in the Bayesian case (see figures 6b and 6c), compared to the frequentist results. The one dimensional marginalised posterior on η is in fact largest near $\eta \sim 0.1$, although this is barely visible in either of the two-dimensional plots. This can be understood as the marginalisation over either m_a or g_{ae} hides the enhancement due to the XENON1T results or the WD cooling hints, respectively. The last figure 6d shows, as we expected from the frequentist results, an enhancement of the posterior probability for the largest allowed ALP-electron coupling $g_{ae} \sim 10^{-13}$ and DM ALP mass around $m_a \sim 3\,\text{keV}$.

We show in table 4 the Bayes factors in favour of the DM ALP model versus the background model. Without tritium, the Bayes factors from the Xe data alone in fact slightly disfavours the DM ALP model. In other words, for our choices of prior, the DM ALP model does not predict the observed data better than the background model. The reward that the DM ALP model obtains for being able to in principle explain the excess for specific combinations of couplings does not outweigh the penalty incurred for making much broader, less specific predictions for the data, including predicting potential signals that are much bigger than the one observed. The combination of XENON1T, the R parameter and WD hints, however, favoured the DM ALP model by about 2.

The partial Bayes factors prefer the DM ALP model more than the Bayes factors, as once we take into account the astrophysical constraints, the remaining viable DM ALP couplings make better predictions for the excess observed at XENON1T than the background model. In particular, the extra freedom in the model allows it to reconcile the WD cooling anomaly and XENON1T, leading to a partial Bayes factor of about 3 in favour of DM ALPs over the background with no tritium for XENON1T given astrophysical data. Thus we again see that the DM ALP model is partially successful in explaining the XENON1T and WD anomalies, whilst avoiding R parameter constraints.

Discussion. We now comment on the differences in the preference for ALPs in our Bayesian and frequentist analyses. The approaches are mathematically different; there is no reason for them to agree and the fact that they differ does not imply that either one is wrong. Nevertheless, we see three reasons why the Bayesian approach yields so much lower preference for the DM ALP hypothesis than the frequentist approach.

First, we did not correctly account for a look-elsewhere effect in our computation of the p-values for the DM ALP, which might reduce the significance.The Bayesian approach, on the other hand, automatically accounts for the fact that the ALP mass is unknown and hence the signal could have appeared anywhere in the search window.

Second, the Bayesian evidence includes an automatic Occam penalty [112]. We chose a very broad prior range for the coupling strength g_{ae}, which in particular includes values

	Xe	(Xe + R)	(Xe + R + WD)	(Xe ∣ R)	(Xe ∣ R + WD)
No ^{3}H	0.76	0.83	1.8	1.5	3.0
^{3}H	0.47	0.51	0.66	0.90	1.1
^{3}H background only	0.15	0.16	0.35	0.29	0.59

Table 4. Bayes factor in favour of DM ALP versus background, with no tritium background (first row), tritium contributions to the ALP model and background-only model (second row) and tritium contributions only in the background model (third row). We show in the first three columns the results considering Xe, adding the R parameter and finally adding WD hints. The final two columns show partial Bayes factor for the Xe data given the R parameter, and given the R parameter and WD hints.

much larger than those favoured by the Xe data. This is penalised by the Bayesian evidence; the fact that the signal is not actually larger than what is observed counts as evidence against the DM ALP model. Of course, XENON1T is not the first experiment to search for ALPs and hence it was already known that g_{ae} cannot be much larger than the XENON1T preferred value. Including this previous information in the prior range for g_{ae} would slightly increase the Bayes factor.

Finally, the Bayesian computations include only the observed data. The frequentist approach, on the other hand, requires that we instead consider data as extreme or more extreme than that observed; see e.g., ref. [113] for further discussion.

5 Conclusions

The recently observed anomalous signal at low energies in the XENON1T experiment, consisting of an excess of 53 events between $1\,\mathrm{keV} - 7\,\mathrm{keV}$ with a significance of 3.5σ above the background only hypothesis, has generated considerable interest due to its possible interpretation as evidence for physics beyond the Standard Model. In this work we focus on two possible explanations: solar ALPs and a DM ALP model in which ALPs constitute a fraction of the local DM density. We used the GAMBIT software to perform global fits of the models to a combination of XENON1T data and astrophysical data[8] using both Bayesian and frequentist statistics to assess the ability of each model to explain the excess.

When the XENON1T data are considered on their own, we find that solar ALPs are favoured by about $\Delta\chi^2 \approx 15$ in our frequentist analysis, and a Bayes factor of about 3 for our choices of prior. When including the lifetime of Horizontal Branch and Red Giant Branch stars, we find that the solar ALP model cannot explain the XENON1T data. Including, in addition, the anomalous cooling of white dwarfs, however, the combined data nevertheless favour solar ALPs over the background only hypothesis by $\Delta\chi^2 \approx 11$ and by a Bayes factor of 1.3. The evidence in favour of the solar ALP is driven largely by the strength of the anomalous white dwarf data, and not by the XENON1T data.

Concerning DM ALPs that constitute a fraction of the local DM density we find that they give a better fit to XENON1T than the solar ALP, with $\Delta\chi^2 \approx 17$. The freedom

---

[8]We provide YAML input files for reproducing our analysis on Zenodo [73].

to lower the DM fraction furthermore allows the model to explain the XENON1T data with a larger axion-electron coupling, which also has the ability to explain the WD cooling hints, resulting in a $\Delta\chi^2 \approx 23$ preference for the DM ALPs. The results of our Bayesian analysis, however, slightly disfavour DM ALPs by about 0.75 with only XENON1T data and slightly favour it but only by about 2 with XENON1T, the R parameter and WD data. For our choices of prior, the background model better predicts the XENON1T data, since, although specific DM ALP couplings could fit the data, the DM ALP model makes much broader, less specific predictions for the data, including predicting potential signals that are much bigger than the one observed. We stress, however, that different choices of prior for the ALP DM fraction and the DM ALP couplings could increase the evidence for the DM ALP model. Moreover we find that the partial Bayes factors in favour of the DM ALP model for the XENON1T data, given known astrophysical constraints and hints, can be as large as 3 even for conservative priors, which, however, still corresponds to an evidence "barely worth mentioning" [109].

The DM ALP model that is favoured by the combination of the XENON1T and WD cooling anomalies consists of a particle mass $\hat{m}_a = 2.7\,\mathrm{keV}$, an axion-electron coupling $\hat{g}_{ae} = 2.5 \times 10^{-13}$, and constitutes a fraction of the local DM of about $\hat{\eta} = 0.030$ ($\eta \lesssim 0.2$ at 95% CL; 1 DOF). Further evidence for or against this model could come from the anticipated XENON-nT data and other electron-recoil direct DM searches, from further study of WD cooling with improved modelling and observations of the period decrease. Another promising strategy is to search for DM ALPs through their inevitable decay into two photons in future x-ray observatories like ATHENA [114].

There is no straight-forward way for the best-fit DM ALP to be the QCD axion, which would require circumventing constraints on hot DM and DM stability. Hence the DM ALP model offers no explanation for the non-observation of the neutron electric dipole moment [115]. Such an ALP is invisible to axion DM haloscopes, such as ADMX [116], which are sensitive only to the ALP-photon coupling, and masses $m \ll 1\,\mathrm{eV}$. It is also invisible to the QUAX haloscope [117], which probes the axion-electron coupling, but is also only sensitive to very low ALP masses. If the XENON1T DM ALP exists, however, it could be part of a larger ALP sector.

Our analysis is insensitive to the DM ALP temperature, and is consistent with either a thermal or non-thermal production channel in the early Universe. In the case of a thermal production channel, evidence for the DM ALP could appear in future precision measurements of the matter power spectrum, such as *Euclid* [118].

Lastly, we consider a possible tritium component in the background. In both the frequentist and Bayesian analyses, a tritium component reduces the preference for the ALP models and is preferred over the background only model by $\Delta\chi^2 \approx 10$ and a Bayes factor of about 5. In the frequentist analysis, however, the combination of XENON1T and the astrophysical data still favours the ALP models by as much as $\Delta\chi^2 \approx 15$, owing to the WD hints. We emphasise that we have allowed the level of tritium to vary by several orders of magnitude. It will be interesting to see what the additional cross checks planned for XENON-nT will reveal about the potential presence of tritium in the detector.

In summary, we have shown that the preference for either a solar or DM ALP explanation of the XENON1T excess strongly depends on the inclusion of astrophysical axion constraints. These generically lower the preference for the solar ALP explanation over the background-only model, and raise it for the DM ALP. Further interesting conclusions result from employing complementary statistical approaches (i.e. both Bayesian and frequentist), which allow one to determine whether it is justified to increase the number of parameters of a theory in order to bring different measurements into better agreement. The combination of growing datasets and advanced statistical methods will inevitably shed more light on the XENON1T anomaly in the near future.

Acknowledgments

A Bayes factors

To understand the impact of the Xe data on the models, we compute Bayes factors and partial Bayes factors [109]. For a review of Bayes factors, see ref. [126]. The Bayes factor,

$$B_{10} = \frac{p(D \mid M_1)}{p(D \mid M_0)},$$ (A.1)

updates our relative belief in two hypotheses (here M_1 and M_0) in light of experimental data, D. The factors in the Bayes factor are Bayesian evidences, which may be found by marginalising the likelihood over a prior for the model's parameters $\boldsymbol{\theta}$ i.e.

$$p(D \mid M) = \int p(D \mid \boldsymbol{\theta}, M) \, p(\boldsymbol{\theta} \mid M) \, \mathrm{d}^n\theta.$$ (A.2)

For a pedagogical discussion of priors, see refs. [127, 128]. The dependence of the above integrals on the choice of prior is a major problem in such analyses. Even among adherents of Bayesian statistics, Bayes factors remain controversial. They are considered by e.g., ref. [129] to be the primary and indisputable tool for model testing to the extent that science should abandon p-values and adopt Bayes factors. On the other hand, ref. [130] emphasizes their dependence on the choice of prior and that the choice may be unjustifiable and untestable. In the Bayesian framework, there are alternative tests that use predictive checks and depend only the posterior distribution of the model's parameters, ameliorating prior dependence. We present one such check, the Deviance Information Criterion (DIC, see appendix C). We also consider the impact of varying the priors within classes of alternatives.

For our main results, we compute all evidence integrals with MultiNest v3.11 [120, 121] inside ScannerBit [119] in GAMBIT v1.4.5. MultiNest is known to efficiently calculate the evidence for up to about 30-dimensional problems [131], whereas our models had at most 7 parameters. We use 5000 live points (nlive), a stopping tolerance (tol) of 0.001 and sampling efficiency (efr) of 0.05. The estimated fractional statistical errors in the evidence estimates were always less than 5%. We furthermore used the state-of-the-art cross-checks recently proposed in ref. [132] and found no evidence of a systematic error in the evidence estimates. To investigate prior dependence in figure 7, the Bayes factors were recomputed by reweighting the evidence estimates, after cross-checks to ensure that the procedure produced reliable results.

We furthermore consider a partial Bayes factor for data D given data D',

$$P_{10} = \frac{p(D \mid D', M_1)}{p(D \mid D', M_0)} = \frac{p(D, D' \mid M_1)}{p(D, D' \mid M_0)} \left(\frac{p(D' \mid M_1)}{p(D' \mid M_0)} \right)^{-1},$$ (A.3)

which is itself a ratio of Bayes factors. The partial Bayes factor tells us how to update our relative belief in two models, supposing that we already knew about the data D'. For independent datasets, the factors in the partial Bayes factor may in fact be written as

$$p(D \mid D', M) = \int p(D \mid \boldsymbol{\theta}, M) \, p(\boldsymbol{\theta} \mid D', M) \, \mathrm{d}^n\theta.$$ (A.4)

Unlike in eq. (A.2) where we averaged over the prior, here we average over a posterior distribution, weakening the dependence on the prior.

Parameter	Prior
Solar ALP	
g_{ae}	Log, $(10^{-20}, 10^{-3})$
$g_{a\gamma}/\mathrm{GeV}^{-1}$	Log, $(10^{-20}, 10^{-3})$
g_{aN}^{eff}	Log, $(10^{-20}, 10^{-3})$
DM ALP	
m_a/keV	Uniform, $(1, 30)$
g_{ae}	Log, $(10^{-20}, 10^{-3})$
XENON1T nuisance parameters	
α_b	Gaussian, 1 ± 0.026
ϵ	Gaussian, 1 ± 0.03
XENON1T tritium component	
α_t	Log-normal, $\log_{10}\left(\dfrac{\alpha_t}{1\,\mathrm{mol/mol}}\right) = -27 \pm 3$
DM ALP nuisance parameters	
ρ_0	Log-normal, $\log_{10}\left(\dfrac{\rho_0}{1\,\mathrm{GeV/\,cm^3}}\right) = \log_{10}(0.4) \pm 0.138$ [68]
η	Uniform, $(0, 1)$

Table 5. Choices of prior for the solar and DM ALP models and nuisance parameters. We discuss these choices at the beginning of section 4.

Lastly, to compare the Bayesian results with the significances reported by XENON1T (e.g., 3.5σ), we compute Z-scores corresponding to the posterior probability of the background models assuming equal prior odds,

$$Z = \Phi^{-1}\left(\frac{1}{1 + 1/B_{10}}\right), \tag{A.5}$$

where Φ is the standard normal CDF. It is known that they are often less than the frequentist significances [133].

B Choice of priors and prior sensitivity

The priors that we choose are summarised in table 5. For the ALP-photon coupling, we choose logarithmic priors in the range $10^{-20}\,\mathrm{GeV}^{-1}$ – $10^{-3}\,\mathrm{GeV}^{-1}$. The lower bound corresponds roughly to the ALP-photon coupling expected from an anomalous global symmetry broken at the GUT scale. Even smaller ALP-photon couplings can be realised if the global symmetry is anomaly-free [80] or if there is a cancellation between different contributions, but such small couplings would be indistinguishable from the background-only hypothesis even for far-future experiments. The upper bound for this range stems from the very conservative assumption that the ALP-photon coupling should be smaller than the pion-photon coupling to satisfy constraints on $e^+e^- \to \gamma + \mathrm{invisible}$ from LEP [134–136], which

are valid for all sub-MeV ALP masses. Of course, astrophysical constraints place much stronger bounds on the ALP-photon coupling. For instance, for $g_{a\gamma} = 10^{-8}\,\text{GeV}^{-1}$ the axion luminosity of the Sun would exceed its photon luminosity by an order of magnitude, which would be in clear contradiction with helioseismology. Rather than encoding astrophysical constraints in the prior ranges, we implement them via the likelihoods discussed above and therefore allow much broader prior ranges.

The ALP-nucleon coupling can be written as $g_{aN}^{\text{eff}} = m_N\,C_N/\Lambda$, where m_N is the nucleon mass, C_N is a coupling constant and Λ is the unknown scale where the effective interaction is generated. Assuming Λ to lie between the electroweak scale and the Planck scale, and C_N to be of order unity, we choose a logarithmic prior in the range 10^{-20} – 10^{-3} to encode our ignorance of the scale of new physics. Naively, the ALP-electron coupling should be smaller than the ALP-nucleon coupling by a factor m_e/m_N. However, there are many examples of ALP models where one of the two couplings is strongly suppressed, for example, if the ALP couples dominantly to gluons [137]. We therefore treat the two couplings as completely independent and choose the same prior range for g_{ae} as for g_{aN}^{eff}. Again, we include astrophysical constraints in the likelihoods rather than in the prior ranges. Since our Bayesian results will depend on these choices, we later investigate the impact of varying the width of the logarithmic ranges for the couplings.
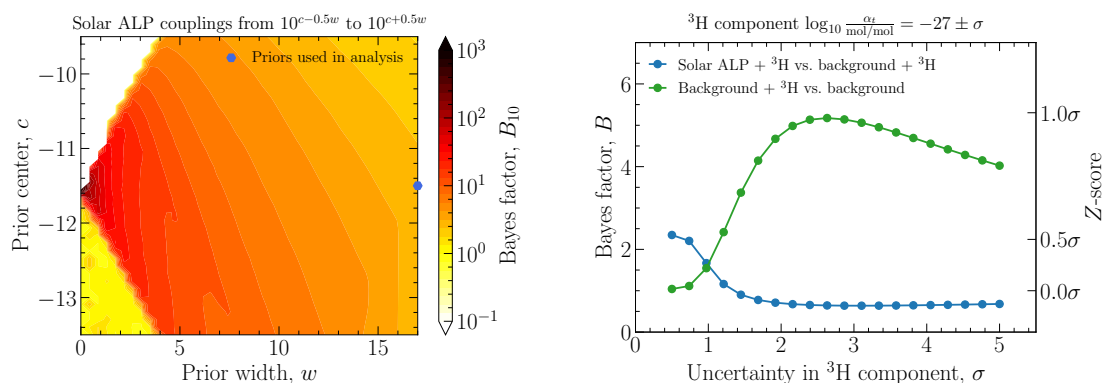
We picked a uniform prior on the ALP mass in the DM ALP scenario from $1\,\text{keV}$ – $30\,\text{keV}$, which spans the XENON1T low-energy energy spectrum. Enlarging this range could only weaken the evidence for a DM ALP signal.

The rationale behind the ^{3}H abundance prior is explained in section 2.3. The prior for the ALP DM fraction, η, is chosen so that ALPs are typically a sizeable fraction of the local DM, even if they could constitute none of it. Our priors for the XENON1T nuisance parameters, α_b and ϵ, are based on the constraints given by XENON1T, while our prior for the local density of DM [68], ρ_0, comes from astrophysical estimates. In our frequentist analyses, we instead include these constraints via equivalent likelihood functions. As discussed in section 2.3, we allow for the tritium component to be unconstrained in our frequentist analyses.

To check the dependence of the Bayes factors that we obtain for solar ALPs in section 4.1 on these choices, we varied the priors for the solar ALP couplings and the uncertainty on the tritium component. We re-computed the Bayes factor for the XENON1T data for logarithmic priors for the solar ALP couplings between $10^{c-0.5w}$ and $10^{c+0.5w}$ for different choices of width, w and centre, c.[9] The results in figure 7a show that the Bayes factor could favour the solar ALP scenario by as much as about 1000 for a narrow prior centred about the best-fitting electron coupling, 3×10^{-12}, but that the preference is usually less than about 10. The Bayes factor falls rapidly once the best-fitting couplings lie outside the prior range, tending towards one if the couplings are too small, as the solar ALP model resembles the background model (lower left yellow triangular region), and zero if they are too large (upper left white triangular region).

We find that a tritium component in the background is favoured over the background only by about 5. To check the dependence of this result on the uncertainty in the tritium

---

[9]See appendix A for details about the re-computation of the Bayes factors.

(a) B_0 + solar ALP versus B_0.

(b) B_0 + solar ALP + ^{3}H versus B_0 + ^{3}H, and B_0 + ^{3}H versus B_0.

Figure 7. Dependence of the Bayes factor with only Xe data on (a) the prior for the solar ALP couplings and (b) the width of the prior for the ^{3}H component of the background. The alternative y-axis on the right-hand side of (b) shows the Bayes factor converted into a Z-score by eq. (A.5).

component, we re-calculated it for alternative choices of prior. We used log-normal priors, $\log_{10} \alpha_t = -27 \pm \sigma$, for σ in the range 0 to 5. The results in figure 7b show that the Bayes factor depends weakly on the uncertainty on the tritium component, ranging from about one when $\sigma \approx 0$, to about five for $\sigma \approx 3$. The latter is preferred as it accommodates the best-fit tritium level, 5×10^{-25} mol/mol, without giving much support to much higher levels of tritium that are strongly disfavoured by the data.

Lastly, we checked the dependence of the Bayes factors and subsequent conclusions in section 4.2 on our choices of priors. We found that had we for example reduced the prior range for the DM ALP coupling g_{ae} from $10^{-20} - 10^{-3}$ to $10^{-20} - 10^{-11}$, all Bayes factors would favour the DM ALP model by approximately a factor of two more. We anticipate, however, that the partial Bayes factors, which take the astrophysical constraints as background knowledge, would not be significantly affected, as the astrophysical constraints already strongly disfavour $g_{ae} \gtrsim 10^{-11}$. Likewise, we find that in order to reconcile the WD hints and XENON1T requires an ALP DM fraction $\eta \lesssim 0.1$, which makes our results sensitive to our choice of prior for η. When using a logarithmic prior from $0.01 - 1$ instead of a uniform prior from $0 - 1$, we find that the preference for the DM ALP model in table 4 increases from $B = 1.8$ to $B = 2.3$.

C Deviance information criterion

Given a parameter point $\boldsymbol{\theta}$ we define the deviance for a model by

$$\mathcal{D}(\boldsymbol{\theta}) \equiv -2 \ln \mathcal{L}(\boldsymbol{\theta}), \tag{C.1}$$

where $\mathcal{L}(\boldsymbol{\theta})$ is the likelihood function. The deviance measures the model's ability to predict the observed data. We can compare several different models by computing the Deviance

Information Criterion (DIC) [138] for each model,

$$\mathrm{DIC} \equiv \mathcal{D}(\langle \boldsymbol{\theta} \rangle) + 2\, p_{\mathcal{D}} \tag{C.2}$$

where $\langle \cdot \rangle$ indicates the posterior mean:

$$\langle \boldsymbol{\theta} \rangle = \int \boldsymbol{\theta} \, p(\boldsymbol{\theta} \,|\, D, M) \, \mathrm{d}\boldsymbol{\theta} \tag{C.3}$$

with D denoting the available data and M the model under consideration. This is an estimate of the deviance that would be obtained with new data given the posterior mean estimate of the parameters, $\langle \boldsymbol{\theta} \rangle$ [130]. The term

$$p_{\mathcal{D}} \equiv \frac{1}{2} \left(\langle \mathcal{D}(\boldsymbol{\theta})^2 \rangle - \langle \mathcal{D}(\boldsymbol{\theta}) \rangle^2 \right) \tag{C.4}$$

corrects bias from over-fitting and is motivated by an analytic result for Gaussian posteriors. It may be interpreted as the number of model parameters constrained by the data, such that the DIC is a Bayesian version of the Akaike Information Criterion (see ref. [139] for further discussion). The model with the smallest DIC is expected to best predict future data and is therefore preferred.

In contrast to the Bayesian evidence, the DIC depends on the posterior and not directly on the prior. Whilst this could reduce the dependence on the choice of prior that is inherent in Bayesian model comparison, we find that this is not the case in the present context, because our posteriors are highly non-Gaussian. More specifically, the posteriors for the ALP couplings and tritium fraction exhibit fat tails which stem from the fact that vanishing couplings cannot be excluded. As a result, the posterior mean estimate of the parameters $\langle \theta \rangle$ can be very different from the most probable value and hence $\mathcal{D}(\langle \theta \rangle)$ can be very different from $\mathcal{D}_{\min}$. Moreover, the fat tails lead to an over-estimation in the number of constrained parameters and therefore an excessively large penalty for over-fitting (see e.g., the Cauchy example in ref. [139]). We therefore find that the DIC generally disfavours models with more parameters, even if they substantially improve the goodness of fit.[10]

We present the DIC results for the solar ALP and DM ALP models in table 6. In each case we show the difference in DIC between the background-only and ALP model such that positive differences indicate that the ALP model is preferred. For the reason explained above our findings are however quite counter-intuitive. For example, adding a tritium component typically reduces the preference for the background model or even leads to a preference for the ALP model. We will therefore refrain from attempting a more detailed interpretation of the numbers in table 6.

Nevertheless, the DIC presented here complements our frequentist and Bayesian results by suggesting that we should not expect ALP models to make better predictions for new data than the background only model. However, given the highly non-Gaussian posteriors in our analysis, our results are very sensitive to the priors and to the specific form of the

---

[10]We note that this pathological behaviour could be rectified (at least partially) by reducing the parameter ranges or by considering linear rather than logarithmic priors for the various couplings, which reduces the fat tails in the posterior and hence their impact on the DIC.

	Xe	(Xe + R)	(Xe + R + WD)
Solar ALP			
No ^{3}H	$-4.7 = 11.7 - 16.4$	$0.002 = 0.186 - 0.184$	$-7.6 = 7.3 - 14.9$
^{3}H	$0.59 = -0.35 + 0.94$	$0.53 \approx 0.57 - 0.03$	$-13 = 4 - 17$
^{3}H background only	$-1.5 \approx 1.9 - 3.3$	$3.2 = -9.6 + 12.8$	$-4.4 = -2.5 - 1.9$
DM ALP			
No ^{3}H	$-27 = 11 - 38$	$-40.5 = 0.2 - 40.7$	$-44 = 11 - 55$
^{3}H	$-1.7 = 0.1 - 1.8$	$-2.8 \approx -0.4 - 2.5$	$-17 = 2 - 19$
^{3}H background only	$-24 = 1 - 25$	$-37 \approx -10 - 28$	$-41 = 1 - 42$

Table 6. Difference in Deviance Information Criterion (DIC) between the solar and DM ALP models and the background only model for the Xe data, when adding the R parameter, and finally when adding the WD hints. Positive (negative) differences indicate that the ALP (background-only) model was favoured. To emphasize the importance of the overfitting term p_D we write our results as $\Delta\text{DIC} = \Delta\mathcal{D} + \Delta(2p_\mathcal{D})$, where $\Delta\mathcal{D}$ is the difference between the two models in the mean deviance and $\Delta(2p_\mathcal{D})$ is the difference in the overfitting penalty.

over-fitting term in the definition of the DIC. We note that alternative forms of the over-fitting term exist; we chose the form in eq. (C.4) because it is always positive, whereas we found that alternatives could become negative for our data and models.

D   Monte Carlo simulations for frequentist results

We largely avoid quoting *p*-values in section 4 and, where we do, we caution that the assumptions of the asymptotic formulae derived from Wilks' theorem are violated in our study. For example, the ALP signals in the XENON1T experiment must be positive, and the DM ALP mass parameter does not exist (is unidentifiable) for $g_{ae} = 0$.

There are several ways to overcome this issue, such as Monte Carlo (MC) simulations and semi-analytic approaches [140]. Monte Carlo simulations require the fewest assumptions to compute *p*-values and the coverage of our confidence regions (see ref. [111] for a more general discussion). In particular, MC simulations can account for the look-elsewhere effect, which affects the DM ALP mass parameter. Such simulations are, however, computationally expensive if the ensuing *p*-values are small.

To estimate the required computational effort, we can use Wilks' theorem, according to which the statistical significances under consideration are between 3σ and 4.5σ (similar for the WD cooling hints). Since *p*-values encode how extreme an observation is compared to the null hypothesis, we are interested in events that occur with odds between $1 : 400$ and $1 : 150,000$. Reliably calculating the probability of such rare events then requires significantly more simulations than the inverse of these odds.

The first step for performing MC simulations is to generate pseudo-data from the background-only hypothesis. Here, we make two simplifying assumptions to render the problem more tractable. First, when considering the DM ALP, we fix the local DM density
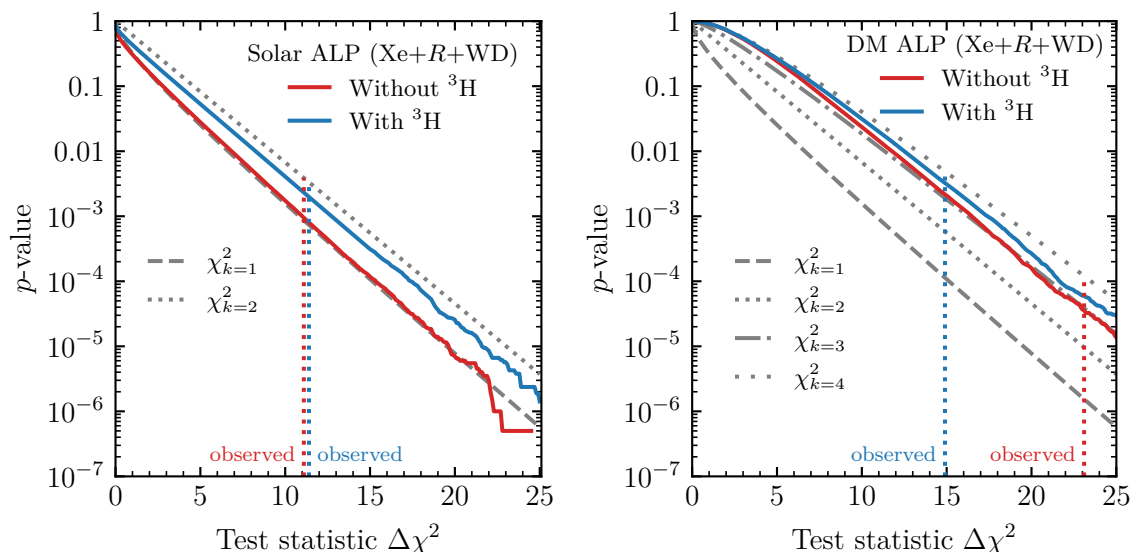
Figure 8. Relationship between p-value and $\Delta\chi^2$ found from Monte Carlo simulations for the solar ALP (*left*) and the DM ALP (*right*) cases, using Xe + R + WD data and with (blue) or without (red) a ^{3}H component. Vertical dotted lines in the corresponding colours denote the observed value of the test statistic. We also show analytically calculated p-value for various χ^2_k distributions with k degrees of freedom (grey lines).

nuisance parameter ρ_0 to its maximum-likelihood value, which only has a mild effect when $\eta \simeq 1$. Second, we always set both $\alpha_{\rm b} = 1$ and $\epsilon = 1$ as these nuisance parameters only have a negligible influence on the result.

We do, however, take into account the nuisance parameters for the systematic uncertainties in the R and WD measurements by averaging over them, which corresponds to computing prior predictive p-values and coverages that average over the unknown nuisance parameters (see ref. [141] for further discussion). Since we consider the background-only model, we only need to vary the nuisance parameters and uncertainties associated with the *measurements* (R parameter and WD pulsation period increase), and not those associated with the theoretical predictions in the ALP models.

The pseudo-data then consists of: counts in each bin for XENON1T (using a Poisson distribution), a measurement of the R parameter (Gaussian with the cut $R > 0$), and the WDs (Gaussian; not imposing a positivity cut since a period decrease is not inherently unphysical, even though it would be a surprising observation).

To obtain a p-value from the simulations, one needs to compute the $\Delta\chi^2$ test-statistic in eq. (4.1) using the pseudo-data. Then, the p-value can be estimated from the fraction of the simulated $\Delta\chi^2$ values that are more extreme than the observed $\Delta\chi^2$ (quoted in the main text in table 2).[11] We show the relationship between p-value and $\Delta\chi^2$ found from

---

[11]To validate our implementation we have performed a goodness-of-fit test using the pseudo-data for XENON1T, finding that the observed data corresponds to a 2σ fluctuation (for the energy range of $1\,\mathrm{keV} - 30\,\mathrm{keV}$), consistent with the expectation for a χ^2 distribution with 29 degrees of freedom. We have also performed MC simulations for a hypothesis test involving the background only model and the solar ALP model (no ^{3}H background) for the Xe likelihood alone, where we find a significance of 3.6σ. This agrees well with the 3.4σ stated by the XENON1T Collaboration, who perform an unbinned analysis and

simulations in figure 8 for the combinations of both ALP models (left and right panels) and ^{3}H hypotheses (red and blue colours), using the Xe + R + WD likelihoods. For reference, we show results from chi-squared distributions with k degrees of freedom (χ^2_k; grey lines) and indicate the *actually observed* $\Delta\chi^2$ values as vertical dotted lines. We generated and optimised about one million pseudo-data sets for each DM ALP case, and two million pseudo-data sets for each solar ALP case.

We can see that the simulated distribution of the test statistics seems to — at least in part — track some of the reference distributions shown in figure 8. For example, the solar ALP hypothesis (without ^{3}H) closely follows a χ^2_1 distribution while the DM ALP is similar to a χ^2_3 distribution. Introducing an additional ^{3}H component for both models increases the p-value for a given value of the test statistic, even though the extra parameter appears not to contribute a full additional degree of freedom.

The fact that the p-value falls more rapidly with increasing value of the test statistic than expected from a naive count of the number of model parameters (three for the solar ALP model) is not unexpected since e.g. Chernoff's theorem [142] predicts a similar deviation for the one-parameter case, where the true value of the parameter of interest lies on the boundary of the parameter space. Similar relations for multi-parameter cases also exist, as e.g. shown in ref. [143]. An additional issue is that g_{aN}^{eff} actually becomes unidentifiable on the background, as a non-zero value of this parameter would not have any effect on the signal in eq. (2.1). Again, this is in violation of the assumptions behind Wilks' theorem.

The situation for the DM ALP scenario is even more involved. Here, we also have three parameters, m_a, g_{ae}, and η (plus α_t for the ^{3}H case). For the background hypothesis of $g_{ae} = 0$ (where these parameters are on the boundary of the parameter space), the ALP mass m_a and efficiency η are not identifiable as can be seen from eq. (2.2). For non-zero values of g_{ae} on the other hand, we have to consider the look-elsewhere effect for the parameter m_a.[12] Further note that, for Xe data alone, the signal prediction is degenerate in the combination of ηg_{ae}^2, which could consequently be treated as a single free parameter. Only the inclusion of other data such as the R parameter and WD data breaks this degeneracy.

While the boundaries of the parameter space and the degeneracies are expected to *decrease* the p-value compared to the naïve expectation, the look-elsewhere effect should *increase* the p-value. The fact that the p-value curves in the right panel of figure 8 approximately follow a χ^2_3 is therefore to some degree coincidental. Nevertheless, these findings imply that there is no large difference between the local significances estimated in section 4 based on Wilks' theorem and the global significances obtained from MC simulations.

Regarding the $\Delta\chi^2$ values in table 2, the calibrated p-values for Xe + R + WD likelihoods from figure 8 imply a significance of 3.3σ for the solar ALP hypothesis (3.1σ when a ^{3}H background is included). Recall that the small difference between the p-values is because the outcome of this analysis mostly determined by the WD likelihoods. For the DM ALP hypothesis, we obtain a significance of 4.1σ (2.9σ when a ^{3}H background is included).

---

take into account additional nuisance parameters.

[12]The trials factor for the look-elsewhere effect on the DM ALP mass is hard to quantify, since our estimate of the local significance ignored the fact that the background model lies at the boundary of two of the DM ALP parameters. With that in mind, we estimate a trials factor of about 20, fairly consistent with rough rules of thumb based on range over resolution of the DM ALP mass.
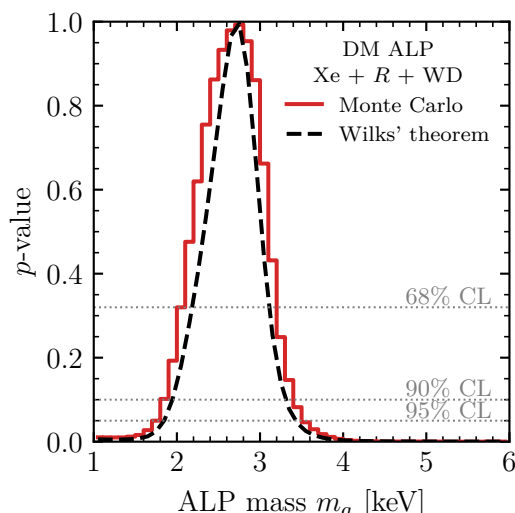
Figure 9. Monte Carlo simulations of *p*-values for the DM ALP mass (red contours) compared to the confidence interval from figure 4 obtained using Wilks' theorem (dashed black contours).

We can also use MC simulations to investigate the accuracy of the confidence regions shown in figure 4. To do so, we consider the profile likelihood ratio test statistic for different values of m_a and generate pseudo-data for the background only hypothesis. Since we are at most interested in the 95% CL region, we only need to accurately estimate *p*-values above 5% and therefore only need to generate a few thousand pseudo-data sets. For every set of pseudo-data, we first find the values of g_{ae} and η that maximise the likelihood for each value of m_a and then find the value of m_a that maximises the likelihood *globally*. We then compare the distribution of profile likelihood ratios for each value of m_a to the actually observed profile likelihood ratio as obtained from our scans of the full parameter space. The *p*-value is then defined as the fraction of simulations that yield a value of the test statistic larger than the one observed. The confidence interval is then given by all values of m_a for which $p > 1 - \text{CL}$.

Figure 9 shows the resulting confidence interval compared to the one from figure 4. We find that the confidence intervals (obtained from the red histogram) are slightly wider than their counterparts from Wilks' theorem (dashed black line). This is not surprising given that Wilks' theorem does not take into account the look-elsewhere effect, which is expected to increase *p*-values, because random fluctuations in all parts of the Xe spectrum can be fitted.

In conclusion, we find various differences between the *p*-values and confidence regions obtained from MC simulations and those derived from asymptotic results such as Wilks' theorem. This result is expected, given the various conditions for Wilks' theorem — or extensions thereof — to hold. Indeed, it is well-known that the MC simulations may both under- or over-predict asymptotic results even for the simple choice of a likelihood ratio as test statistic (see e.g. ref. [111] for a more detailed discussion). In our concrete case, we deal with parameters on the boundary of the parameters space, unidentifiable parameters, and the look-elsewhere effect. In spite of all these complications, we find no qualitative

differences between results obtained from asymptotic expressions and MC simulations and therefore conclude that our results are robust. Indeed, it appears that some of the additional effects included in the MC simulations partially cancel, leading to a surprisingly good agreement between our simple estimates and the results from more elaborate simulations.

References

[1] XENON collaboration, *Excess electronic recoil events in XENON*1*T*, *Phys. Rev. D* **102** (2020) 072004 [`arXiv:2006.09721`] [INSPIRE].

[2] G. Choi, M. Suzuki and T.T. Yanagida, *XENON*1*T anomaly and its implication for decaying warm dark matter*, *Phys. Lett. B* **811** (2020) 135976 [`arXiv:2006.12348`] [INSPIRE].

[3] D. Aristizabal Sierra, V. De Romeri, L.J. Flores and D.K. Papoulias, *Light vector mediators facing XENON*1*T data*, *Phys. Lett. B* **809** (2020) 135681 [`arXiv:2006.12457`] [INSPIRE].

[4] N.F. Bell, J.B. Dent, B. Dutta, S. Ghosh, J. Kumar and J.L. Newstead, *Explaining the XENON*1*T excess with luminous dark matter*, *Phys. Rev. Lett.* **125** (2020) 161803 [`arXiv:2006.12461`] [INSPIRE].

[5] G. Paz, A.A. Petrov, M. Tammaro and J. Zupan, *Shining dark matter in XENON*1*T*, *Phys. Rev. D* **103** (2021) L051703 [`arXiv:2006.12462`] [INSPIRE].

[6] Y. Chen, M.-Y. Cui, J. Shu, X. Xue, G.-W. Yuan and Q. Yuan, *Sun heated MeV-scale dark matter and the XENON*1*T electron recoil excess*, *JHEP* **04** (2021) 282 [`arXiv:2006.12447`] [INSPIRE].

[7] R. Primulando, J. Julio and P. Uttayarat, *Collider constraints on a dark matter interpretation of the XENON*1*T excess*, *Eur. Phys. J. C* **80** (2020) 1084 [`arXiv:2006.13161`] [INSPIRE].

[8] L. Di Luzio, M. Fedele, M. Giannotti, F. Mescia and E. Nardi, *Solar axions cannot explain the XENON*1*T excess*, *Phys. Rev. Lett.* **125** (2020) 131804 [`arXiv:2006.12487`] [INSPIRE].

[9] M. Du, J. Liang, Z. Liu, V.Q. Tran and Y. Xue, *On-shell mediator dark matter models and the XENON*1*T excess*, *Chin. Phys. C* **45** (2021) 013114 [`arXiv:2006.11949`] [INSPIRE].

[10] L. Su, W. Wang, L. Wu, J.M. Yang and B. Zhu, *Atmospheric dark matter and XENON*1*T excess*, *Phys. Rev. D* **102** (2020) 115028 [`arXiv:2006.11837`] [INSPIRE].

[11] A. Bally, S. Jana and A. Trautner, *Neutrino self-interactions and XENON*1*T electron recoil excess*, *Phys. Rev. Lett.* **125** (2020) 161802 [`arXiv:2006.11919`] [INSPIRE].

[12] K. Harigaya, Y. Nakai and M. Suzuki, *Inelastic dark matter electron scattering and the XENON*1*T excess*, *Phys. Lett. B* **809** (2020) 135729 [`arXiv:2006.11938`] [INSPIRE].

[13] C. Boehm, D.G. Cerdeno, M. Fairbairn, P.A.N. Machado and A.C. Vincent, *Light new physics in XENON*1*T*, *Phys. Rev. D* **102** (2020) 115013 [`arXiv:2006.11250`] [INSPIRE].

[14] D.W.P.d. Amaral, D.G. Cerdeno, P. Foldenauer and E. Reid, *Solar neutrino probes of the muon anomalous magnetic moment in the gauged* $U(1)_{L_\mu - L_\tau}$, *JHEP* **12** (2020) 155 [arXiv:2006.11225] [INSPIRE].

[15] B. Fornal, P. Sandick, J. Shu, M. Su and Y. Zhao, *Boosted dark matter interpretation of the XENON1T excess*, *Phys. Rev. Lett.* **125** (2020) 161804 [arXiv:2006.11264] [INSPIRE].

[16] G. Alonso-Álvarez, F. Ertas, J. Jaeckel, F. Kahlhoefer and L.J. Thormaehlen, *Hidden photon dark matter in the light of XENON1T and stellar cooling*, *JCAP* **11** (2020) 029 [arXiv:2006.11243] [INSPIRE].

[17] K. Kannike, M. Raidal, H. Veermäe, A. Strumia and D. Teresi, *Dark matter and the XENON1T electron recoil excess*, *Phys. Rev. D* **102** (2020) 095002 [arXiv:2006.10735] [INSPIRE].

[18] C.A.J. O'Hare, A. Caputo, A.J. Millar and E. Vitagliano, *Axion helioscopes as solar magnetometers*, *Phys. Rev. D* **102** (2020) 043019 [arXiv:2006.10415] [INSPIRE].

[19] F. Takahashi, M. Yamada and W. Yin, *XENON1T excess from anomaly-free axionlike dark matter and its implications for stellar cooling anomaly*, *Phys. Rev. Lett.* **125** (2020) 161801 [arXiv:2006.10035] [INSPIRE].

[20] G.B. Gelmini, V. Takhistov and E. Vitagliano, *Scalar direct detection: in-medium effects*, *Phys. Lett. B* **809** (2020) 135779 [arXiv:2006.13909] [INSPIRE].

[21] M. Baryakhtar, A. Berlin, H. Liu and N. Weiner, *Electromagnetic signals of inelastic dark matter scattering*, arXiv:2006.13918 [INSPIRE].

[22] L. Zu, G.-W. Yuan, L. Feng and Y.-Z. Fan, *Mirror dark matter and electronic recoil events in XENON1T*, *Nucl. Phys. B* **965** (2021) 115369 [arXiv:2006.14577] [INSPIRE].

[23] M. Lindner, Y. Mambrini, T.B. de Melo and F.S. Queiroz, *XENON1T anomaly: a light Z′ from a two Higgs doublet model*, *Phys. Lett. B* **811** (2020) 135972 [arXiv:2006.14590] [INSPIRE].

[24] K. Zioutas, G. Cantatore, M. Karuza, A. Kryemadhi, M. Maroudas and Y.K. Semertzidis, *Response-suggestion to the XENON1T excess: an overlooked dark matter signature?*, arXiv:2006.16907 [INSPIRE].

[25] D. McKeen, M. Pospelov and N. Raj, *Hydrogen portal to exotic radioactivity*, *Phys. Rev. Lett.* **125** (2020) 231803 [arXiv:2006.15140] [INSPIRE].

[26] P. Coloma, P. Huber and J.M. Link, *Telling solar neutrinos from solar axions when you can't shut off the sun*, arXiv:2006.15767 [INSPIRE].

[27] H. An and D. Yang, *Direct detection of freeze-in inelastic dark matter*, arXiv:2006.15672 [INSPIRE].

[28] S.-F. Ge, P. Pasquini and J. Sheng, *Solar neutrino scattering with electron into massive sterile neutrino*, *Phys. Lett. B* **810** (2020) 135787 [arXiv:2006.16069] [INSPIRE].

[29] C. Dessert, J.W. Foster, Y. Kahn and B.R. Safdi, *Systematics in the XENON1T data: the 15 keV anti-axion*, arXiv:2006.16220 [INSPIRE].

[30] W. Chao, Y. Gao and M.j. Jin, *Pseudo-Dirac dark matter in XENON1T*, arXiv:2006.16145 [INSPIRE].

[31] C. Cai, H.H. Zhang, M.T. Frandsen, M. Rosenlyst and G. Cacciapaglia, *XENON1T solar axion and the Higgs boson emerging from the dark*, *Phys. Rev. D* **102** (2020) 075018 [arXiv:2006.16267] [INSPIRE].

[32] P. Ko and Y. Tang, *Semi-annihilating Z_3 dark matter for XENON1T excess*, *Phys. Lett. B* **815** (2021) 136181 [arXiv:2006.15822] [INSPIRE].

[33] Y. Gao and T. Li, *Lepton number violating electron recoils at XENON1T by the $U(1)_{B-L}$ model with non-standard interactions*, arXiv:2006.16192 [INSPIRE].

[34] J. Sun and X.-G. He, *DFSZ axion couplings revisited*, *Phys. Lett. B* **811** (2020) 135881 [arXiv:2006.16931] [INSPIRE].

[35] S. Baek, J. Kim and P. Ko, *XENON1T excess in local Z_2 DM models with light dark sector*, *Phys. Lett. B* **810** (2020) 135848 [arXiv:2006.16876] [INSPIRE].

[36] R. Budnik, H. Kim, O. Matsedonskyi, G. Perez and Y. Soreq, *Probing the relaxed relaxion and Higgs-portal with $S1\&S2$*, arXiv:2006.14568 [INSPIRE].

[37] H.-J. He, Y.-C. Wang and J. Zheng, *EFT approach of inelastic dark matter for xenon electron recoil detection*, *JCAP* **01** (2021) 042 [arXiv:2007.04963] [INSPIRE].

[38] M. Chala and A. Titov, *One-loop running of dimension-six Higgs-neutrino operators and implications of a large neutrino dipole moment*, *JHEP* **09** (2020) 188 [arXiv:2006.14596] [INSPIRE].

[39] F. Arias-Aragón, F. D'eramo, R.Z. Ferreira, L. Merlo and A. Notari, *Cosmic imprints of XENON1T axions*, *JCAP* **11** (2020) 025 [arXiv:2007.06579] [INSPIRE].

[40] C. Han, M.L. López-Ibáñez, A. Melis, O. Vives and J.M. Yang, *Anomaly-free leptophilic axionlike particle and its flavor violating tests*, *Phys. Rev. D* **103** (2021) 035028 [arXiv:2007.08834] [INSPIRE].

[41] I.M. Bloch, A. Caputo, R. Essig, D. Redigolo, M. Sholapurkar and T. Volansky, *Exploring new physics with O(keV) electron recoils in direct detection experiments*, *JHEP* **01** (2021) 178 [arXiv:2006.14521] [INSPIRE].

[42] O.G. Miranda, D.K. Papoulias, M. Tórtola and J.W.F. Valle, *XENON1T signal from transition neutrino magnetic moments*, *Phys. Lett. B* **808** (2020) 135685 [arXiv:2007.01765] [INSPIRE].

[43] S. Chigusa, M. Endo and K. Kohri, *Constraints on electron-scattering interpretation of XENON1T excess*, *JCAP* **10** (2020) 035 [arXiv:2007.01663] [INSPIRE].

[44] D. Croon, S.D. McDermott and J. Sakstein, *Missing in axion: where are XENON1T's big black holes?*, *Phys. Dark Univ.* **32** (2021) 100801 [arXiv:2007.00650] [INSPIRE].

[45] T. Li, *The KSVZ axion and pseudo-Nambu-Goldstone boson models for the XENON1T excess*, arXiv:2007.00874 [INSPIRE].

[46] H.M. Lee, *Exothermic dark matter for XENON1T excess*, *JHEP* **01** (2021) 019 [arXiv:2006.13183] [INSPIRE].

[47] J. Bramante and N. Song, *Electric but not eclectic: thermal relic dark matter for the XENON1T excess*, *Phys. Rev. Lett.* **125** (2020) 161805 [arXiv:2006.14089] [INSPIRE].

[48] D. Borah, S. Mahapatra, D. Nanda and N. Sahu, *Inelastic fermion dark matter origin of XENON1T excess with muon $(g-2)$ and light neutrino mass*, *Phys. Lett. B* **811** (2020) 135933 [arXiv:2007.10754] [INSPIRE].

[49] Q.-H. Cao, R. Ding and Q.-F. Xiang, *Searching for sub-MeV boosted dark matter from xenon electron direct detection*, *Chin. Phys. C* **45** (2021) 045002 [arXiv:2006.12767] [InSPIRE].

[50] Y. Jho, J.-C. Park, S.C. Park and P.-Y. Tseng, *Leptonic new force and cosmic-ray boosted dark matter for the XENON1T excess*, *Phys. Lett. B* **811** (2020) 135863 [arXiv:2006.13910] [InSPIRE].

[51] L. Delle Rose, G. Hütsi, C. Marzo and L. Marzola, *Impact of loop-induced processes on the boosted dark matter interpretation of the XENON1T excess*, *JCAP* **02** (2021) 031 [arXiv:2006.16078] [InSPIRE].

[52] H. Alhazmi, D. Kim, K. Kong, G. Mohlabeng, J.-C. Park and S. Shin, *Implications of the XENON1T excess on the dark matter interpretation*, arXiv:2006.16252 [InSPIRE].

[53] K. Nakayama and Y. Tang, *Gravitational production of hidden photon dark matter in light of the XENON1T excess*, *Phys. Lett. B* **811** (2020) 135977 [arXiv:2006.13159] [InSPIRE].

[54] H. An, M. Pospelov, J. Pradler and A. Ritz, *New limits on dark photons from solar emission and keV scale dark matter*, *Phys. Rev. D* **102** (2020) 115022 [arXiv:2006.13929] [InSPIRE].

[55] C.-W. Chiang and B.-Q. Lu, *Evidence of a simple dark sector from XENON1T excess*, *Phys. Rev. D* **102** (2020) 123006 [arXiv:2007.06401] [InSPIRE].

[56] J. Buch, M.A. Buen-Abad, J. Fan and J.S.C. Leung, *Galactic origin of relativistic bosons and XENON1T excess*, *JCAP* **10** (2020) 051 [arXiv:2006.12488] [InSPIRE].

[57] U.K. Dey, T.N. Maity and T.S. Ray, *Prospects of Migdal effect in the explanation of XENON1T electron recoil excess*, *Phys. Lett. B* **811** (2020) 135900 [arXiv:2006.12529] [InSPIRE].

[58] A.N. Khan, *Can nonstandard neutrino interactions explain the XENON1T spectral excess?*, *Phys. Lett. B* **809** (2020) 135782 [arXiv:2006.12887] [InSPIRE].

[59] N. Okada, S. Okada, D. Raut and Q. Shafi, *Dark matter Z' and XENON1T excess from* U(1)$_X$ *extended standard model*, *Phys. Lett. B* **810** (2020) 135785 [arXiv:2007.02898] [InSPIRE].

[60] G. Arcadi, A. Bally, F. Goertz, K. Tame-Narvaez, V. Tenorth and S. Vogl, *EFT interpretation of XENON1T electron recoil excess: neutrinos and dark matter*, *Phys. Rev. D* **103** (2021) 023024 [arXiv:2007.08500] [InSPIRE].

[61] D. Choudhury, S. Maharana, D. Sachdeva and V. Sahdev, *Dark matter, muon anomalous magnetic moment, and the XENON1T excess*, *Phys. Rev. D* **103** (2021) 015006 [arXiv:2007.08205] [InSPIRE].

[62] S. Karmakar and S. Pandey, *XENON1T constraints on neutrino non-standard interactions*, arXiv:2007.11892 [InSPIRE].

[63] J. Davighi, M. McCullough and J. Tooby-Smith, *Undulating dark matter*, *JHEP* **11** (2020) 120 [arXiv:2007.03662] [InSPIRE].

[64] K. Van Tilburg, *Stellar basins of gravitationally bound particles*, arXiv:2006.12431 [InSPIRE].

[65] A.E. Robinson, *XENON1T observes tritium*, arXiv:2006.13278 [InSPIRE].

[66] M. Szydagis, C. Levy, G.M. Blockinger, A. Kamaha, N. Parveen and G.R.C. Rischbieter, *Investigating the XENON1T low-energy electronic recoil excess using NEST*, *Phys. Rev. D* **103** (2021) 012002 [arXiv:2007.00528] [InSPIRE].

[67] GAMBIT collaboration, *GAMBIT: the global and modular beyond-the-standard-model inference tool*, *Eur. Phys. J. C* **77** (2017) 784 [*Addendum ibid.* **78** (2018) 98] [arXiv:1705.07908] [InSPIRE].

[68] GAMBIT Dark Matter Workgroup collaboration, *DarkBit: a GAMBIT module for computing dark matter observables and likelihoods*, *Eur. Phys. J. C* **77** (2017) 831 [arXiv:1705.07920] [InSPIRE].

[69] A. Kvellestad, P. Scott and M. White, *GAMBIT and its application in the search for physics beyond the standard model*, *Prog. Part. Nucl. Phys.* **113** (2020) 103769 [arXiv:1912.04079] [InSPIRE].

[70] S. Hoof, F. Kahlhoefer, P. Scott, C. Weniger and M. White, *Axion global fits with Peccei-Quinn symmetry breaking before inflation using GAMBIT*, *JHEP* **03** (2019) 191 [*Erratum ibid.* **11** (2019) 099] [arXiv:1810.07192] [InSPIRE].

[71] C. Gao, J. Liu, L.-T. Wang, X.-P. Wang, W. Xue and Y.-M. Zhong, *Reexamining the solar axion explanation for the XENON1T excess*, *Phys. Rev. Lett.* **125** (2020) 131806 [arXiv:2006.14598] [InSPIRE].

[72] J.B. Dent, B. Dutta, J.L. Newstead and A. Thompson, *Inverse Primakoff scattering as a probe of solar axions at liquid xenon direct detection experiments*, *Phys. Rev. Lett.* **125** (2020) 131805 [arXiv:2006.15118] [InSPIRE].

[73] P. Athron et al., *Supplementary material for "Global fits of axion-like particles to XENON1T and astrophysical data"*, Zenodo, (2021).

[74] L. Di Luzio, M. Giannotti, E. Nardi and L. Visinelli, *The landscape of QCD axion models*, *Phys. Rept.* **870** (2020) 1 [arXiv:2003.01100] [InSPIRE].

[75] P. Arias, D. Cadamuro, M. Goodsell, J. Jaeckel, J. Redondo and A. Ringwald, *WISPy cold dark matter*, *JCAP* **06** (2012) 013 [arXiv:1201.5902] [InSPIRE].

[76] D.J.E. Marsh, *Axion cosmology*, *Phys. Rept.* **643** (2016) 1 [arXiv:1510.07633] [InSPIRE].

[77] K. Arisaka et al., *Expected sensitivity to galactic/solar axions and bosonic super-WIMPs based on the axio-electric effect in liquid xenon dark matter detectors*, *Astropart. Phys.* **44** (2013) 59 [arXiv:1209.3810] [InSPIRE].

[78] W. Veigele, *Photon cross sections from* 0.1 *keV to* 1 *MeV for elements* $z = 1$ *to* $z = 94$, *Atom. Data Nucl. Data Tabl.* **5** (1973) 51.

[79] J. Jaeckel, J. Redondo and A. Ringwald, 3.55 *keV hint for decaying axionlike particle dark matter*, *Phys. Rev. D* **89** (2014) 103511 [arXiv:1402.7335] [InSPIRE].

[80] K. Nakayama, F. Takahashi and T.T. Yanagida, *Anomaly-free flavor models for Nambu-Goldstone bosons and the* 3.5 *keV X-ray line signal*, *Phys. Lett. B* **734** (2014) 178 [arXiv:1403.7390] [InSPIRE].

[81] V. Iršič et al., *New constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman-α forest data*, *Phys. Rev. D* **96** (2017) 023522 [arXiv:1702.01764] [InSPIRE].

[82] A.V. Maccio, O. Ruchayskiy, A. Boyarsky and J.C. Muñoz-Cuartas, *The inner structure of haloes in cold+warm dark matter models*, *Mon. Not. Roy. Astron. Soc.* **428** (2013) 882 [arXiv:1202.2858] [InSPIRE].

[83] A. Kamada and K. Yanagi, *Constraining FIMP from the structure formation of the universe: analytic mapping from* m_{WDM}, *JCAP* **11** (2019) 029 [arXiv:1907.04558] [INSPIRE].

[84] N. Vinyoles et al., *A new generation of standard solar models*, *Astrophys. J.* **835** (2017) 202 [arXiv:1611.09867] [INSPIRE].

[85] M. Giannotti, I.G. Irastorza, J. Redondo, A. Ringwald and K. Saikawa, *Stellar recipes for axion hunters*, *JCAP* **10** (2017) 010 [arXiv:1708.02111] [INSPIRE].

[86] M. Giannotti, I. Irastorza, J. Redondo and A. Ringwald, *Cool WISPs for stellar cooling excesses*, *JCAP* **05** (2016) 057 [arXiv:1512.08108] [INSPIRE].

[87] E. Aver, K.A. Olive and E.D. Skillman, *The effects of He I* $\lambda 10830$ *on helium abundance determinations*, *JCAP* **07** (2015) 011 [arXiv:1503.08146] [INSPIRE].

[88] A.H. Corsico et al., *The rate of cooling of the pulsating white dwarf star* $G117 - B15A$: *a new asteroseismological inference of the axion mass*, *Mon. Not. Roy. Astron. Soc.* **424** (2012) 2792 [arXiv:1205.6180] [INSPIRE].

[89] A.H. Corsico et al., *An independent limit on the axion mass from the variable white dwarf star* $R548$, *JCAP* **12** (2012) 010 [arXiv:1211.3389] [INSPIRE].

[90] A.H. Córsico et al., *An asteroseismic constraint on the mass of the axion from the period drift of the pulsating DA white dwarf star* $L19 - 2$, *JCAP* **07** (2016) 036 [arXiv:1605.06458] [INSPIRE].

[91] T. Battich, A.H. Córsico, L.G. Althaus, M.M. Miller Bertolami and M.M.M. Bertolami, *First axion bounds from a pulsating helium-rich white dwarf star*, *JCAP* **08** (2016) 062 [arXiv:1605.07668] [INSPIRE].

[92] XENON collaboration, *Data from: observation of excess electronic recoil events in* $XENON1T$, Zenodo, (2020).

[93] J.-W. Chen, H.-C. Chi, C.P. Liu and C.-P. Wu, *Low-energy electronic recoil in xenon detectors by solar neutrinos*, *Phys. Lett. B* **774** (2017) 656 [arXiv:1610.04177] [INSPIRE].

[94] B. Bhattacherjee and R. Sengupta, $XENON1T$ *excess: some possible backgrounds*, *Phys. Lett. B* **817** (2021) 136305 [arXiv:2006.16172] [INSPIRE].

[95] XENON collaboration, *Energy resolution and linearity of* $XENON1T$ *in the MeV energy range*, *Eur. Phys. J. C* **80** (2020) 785 [arXiv:2003.03825] [INSPIRE].

[96] G. Raffelt, *Stars as laboratories for fundamental physics: the astrophysics of neutrinos, axions, and other weakly interacting particles*, University Of Chicago Press, Chicago, IL, U.S.A. and London, U.K. (1996).

[97] D. Cadamuro and J. Redondo, *Cosmological bounds on pseudo Nambu-Goldstone bosons*, *JCAP* **02** (2012) 032 [arXiv:1110.2895] [INSPIRE].

[98] A. Ayala, I. Domínguez, M. Giannotti, A. Mirizzi and O. Straniero, *Revisiting the bound on axion-photon coupling from globular clusters*, *Phys. Rev. Lett.* **113** (2014) 191302 [arXiv:1406.6053] [INSPIRE].

[99] M.M. Miller Bertolami, B.E. Melendez, L.G. Althaus and J. Isern, *Revisiting the axion bounds from the galactic white dwarf luminosity function*, *JCAP* **10** (2014) 069 [arXiv:1406.7712] [INSPIRE].

[100] S. Hoof, F. Kahlhoefer, P. Scott, C. Weniger and M. White, *Axion global fits with Peccei-Quinn symmetry breaking before inflation using GAMBIT*, *JHEP* **03** (2019) 191 [*Erratum ibid.* **11** (2019) 099] [arXiv:1810.07192] [INSPIRE].

[101] L. Calibbi, D. Redigolo, R. Ziegler and J. Zupan, *Looking forward to lepton-flavor-violating ALPs*, arXiv:2006.04795 [INSPIRE].

[102] T. Battich, A.H. Córsico, L.G. Althaus, M.M. Miller Bertolami and M.M.M. Bertolami, *First axion bounds from a pulsating helium-rich white dwarf star*, *JCAP* **08** (2016) 062 [arXiv:1605.07668] [INSPIRE].

[103] S. Horiuchi, P.J. Humphrey, J. Onorbe, K.N. Abazajian, M. Kaplinghat and S. Garrison-Kimmel, *Sterile neutrino dark matter bounds from galaxies of the local group*, *Phys. Rev. D* **89** (2014) 025017 [arXiv:1311.0282] [INSPIRE].

[104] K. Perez, K.C.Y. Ng, J.F. Beacom, C. Hersh, S. Horiuchi and R. Krivonos, *Almost closing the νMSM sterile neutrino dark matter window with NuSTAR*, *Phys. Rev. D* **95** (2017) 123002 [arXiv:1609.00667] [INSPIRE].

[105] N. Bar, K. Blum and G. D'Amico, *Is there a supernova bound on axions?*, *Phys. Rev. D* **101** (2020) 123025 [arXiv:1907.05020] [INSPIRE].

[106] R. Bollig, W. DeRocco, P.W. Graham and H.-T. Janka, *Muons in supernovae: implications for the axion-muon coupling*, *Phys. Rev. Lett.* **125** (2020) 051104 [arXiv:2005.07141] [INSPIRE].

[107] P. Carenza, T. Fischer, M. Giannotti, G. Guo, G. Martínez-Pinedo and A. Mirizzi, *Improved axion emissivity from a supernova via nucleon-nucleon bremsstrahlung*, *JCAP* **10** (2019) 016 [*Erratum ibid.* **05** (2020) E01] [arXiv:1906.11844] [INSPIRE].

[108] L. Di Luzio, F. Mescia and E. Nardi, *Redefining the axion window*, *Phys. Rev. Lett.* **118** (2017) 031801 [arXiv:1610.07593] [INSPIRE].

[109] H. Jeffreys, *The theory of probability*, Oxford Classic Texts in the Physical Sciences, Oxford University Press, Oxford, U.K. (1939).

[110] S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *Annals Math. Statist.* **9** (1938) 60 [INSPIRE].

[111] S. Algeri, J. Aalbers, K.D. Morå and J. Conrad, *Searching for new phenomena with profile likelihood ratio tests*, *Nature Rev. Phys.* **2** (2020) 245.

[112] D.J.C. MacKay, *Information theory, inference & learning algorithms*, Cambridge University Press, U.S.A. (2002).

[113] J.O. Berger and M. Delampady, *Testing precise hypotheses*, *Statist. Sci.* **2** (1987) 317.

[114] A. Neronov and D. Malyshev, *Toward a full test of the νMSM sterile neutrino dark matter model with Athena*, *Phys. Rev. D* **93** (2016) 063518 [arXiv:1509.02758] [INSPIRE].

[115] NEDM collaboration, *Measurement of the permanent electric dipole moment of the neutron*, *Phys. Rev. Lett.* **124** (2020) 081803 [arXiv:2001.11966] [INSPIRE].

[116] ADMX collaboration, *Extended search for the invisible axion with the Axion Dark Matter Experiment*, *Phys. Rev. Lett.* **124** (2020) 101303 [arXiv:1910.08638] [INSPIRE].

[117] QUAX collaboration, *Axion search with a quantum-limited ferromagnetic haloscope*, *Phys. Rev. Lett.* **124** (2020) 171801 [arXiv:2001.08940] [INSPIRE].

[118] L. Amendola et al., *Cosmology and fundamental physics with the Euclid satellite*, *Living Rev. Rel.* **21** (2018) 2 [arXiv:1606.00180] [INSPIRE].

[119] GAMBIT collaboration, *Comparison of statistical sampling methods with ScannerBit, the GAMBIT scanning module*, *Eur. Phys. J. C* **77** (2017) 761 [arXiv:1705.07959] [INSPIRE].

[120] F. Feroz, M.P. Hobson and M. Bridges, *MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics*, *Mon. Not. Roy. Astron. Soc.* **398** (2009) 1601 [arXiv:0809.3437] [INSPIRE].

[121] F. Feroz, M.P. Hobson, E. Cameron and A.N. Pettitt, *Importance nested sampling and the MultiNest algorithm*, *Open J. Astrophys.* **2** (2019) 10 [arXiv:1306.2144] [INSPIRE].

[122] P. Scott, *Pippi — painless parsing, post-processing and plotting of posterior and likelihood samples*, *Eur. Phys. J. Plus* **127** (2012) 138 [arXiv:1206.2245] [INSPIRE].

[123] J.D. Hunter, *Matplotlib: a 2D graphics environment*, *Comput. Sci. Eng.* **9** (2007) 90 [INSPIRE].

[124] P. Virtanen et al., *SciPy 1.0: fundamental algorithms for scientific computing in Python*, *Nature Meth.* **17** (2020) 261.

[125] S. van der Walt, S.C. Colbert and G. Varoquaux, *The NumPy array: a structure for efficient numerical computation*, *Comput. Sci. Eng.* **13** (2011) 22.

[126] R.E. Kass and A.E. Raftery, *Bayes factors*, *J. Amer. Statist. Assoc.* **90** (1995) 773.

[127] R.E. Kass and L. Wasserman, *The selection of prior distributions by formal rules*, *J. Amer. Statist. Assoc.* **91** (1996) 1343.

[128] G. Consonni, D. Fouskakis, B. Liseo and I. Ntzoufras, *Prior distributions for objective Bayesian analysis*, *Bayesian Anal.* **13** (2018) 627.

[129] J. Berger and L. Pericchi, *Bayes factors*, in *Wiley StatsRef: statistics reference online*, John Wiley & Sons Inc., U.S.A. (2015), pg. 1.

[130] A. Gelman, J. Carlin, H. Stern and D. Rubin, *Bayesian data analysis*, third edition, Texts in statistical science, Chapman & Hall/CRC, (2004).

[131] W.J. Handley, M.P. Hobson and A.N. Lasenby, *PolyChord: nested sampling for cosmology*, *Mon. Not. Roy. Astron. Soc.* **450** (2015) L61 [arXiv:1502.01856] [INSPIRE].

[132] A. Fowlie, W. Handley and L. Su, *Nested sampling cross-checks using order statistics*, *Mon. Not. Roy. Astron. Soc.* **497** (2020) 5256 [arXiv:2006.03371] [INSPIRE].

[133] A. Fowlie, *Bayesian and frequentist approaches to resonance searches*, 2019 *JINST* **14** P10031 [arXiv:1902.03243] [INSPIRE].

[134] P.J. Fox, R. Harnik, J. Kopp and Y. Tsai, *LEP shines light on dark matter*, *Phys. Rev. D* **84** (2011) 014028 [arXiv:1103.0240] [INSPIRE].

[135] DELPHI collaboration, *Search for one large extra dimension with the DELPHI detector at LEP*, *Eur. Phys. J. C* **60** (2009) 17 [arXiv:0901.4486] [INSPIRE].

[136] M.J. Dolan, T. Ferber, C. Hearty, F. Kahlhoefer and K. Schmidt-Hoberg, *Revised constraints and Belle II sensitivity for visible and invisible axion-like particles*, *JHEP* **12** (2017) 094 [*Erratum ibid.* **03** (2021) 190] [arXiv:1709.00009] [INSPIRE].

[137] F. Ertas and F. Kahlhoefer, *On the interplay between astrophysical and laboratory probes of MeV-scale axion-like particles*, *JHEP* **07** (2020) 050 [arXiv:2004.01193] [INSPIRE].

[138] D.J. Spiegelhalter, N.G. Best, B.P. Carlin and A. van der Linde, *Bayesian measures of model complexity and fit*, *J. Roy. Statist. Soc. B* **64** (2002) 583.

[139] W. Handley and P. Lemos, *Quantifying dimensionality: Bayesian cosmological model complexities*, *Phys. Rev. D* **100** (2019) 023512 [`arXiv:1903.06682`] [ɪɴSPIRE].

[140] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, *Eur. Phys. J. C* **70** (2010) 525 [`arXiv:1005.1891`] [ɪɴSPIRE].

[141] C. Blocker et al., *Simple facts about p-values*, Tech. Rep. CDF/MEMO/STATISTICS/PUBLIC/8023, Fermilab, Batavia, IL, U.S.A. (2006).

[142] H. Chernoff, *On the distribution of the likelihood ratio*, *Ann. Math. Statist.* **25** (1954) 573.

[143] S.G. Self and K.-Y. Liang, *Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions*, *J. Amer. Statist. Assoc.* **82** (1987) 605.