Evidence for axion-like particles from XENON1T and astrophysical data: noise or Nobel?

Peter Athron et al. (July 2020). In: arXiv: 2007.05517 [astro-ph.CO]

Andrew Fowlie

12 January 2021

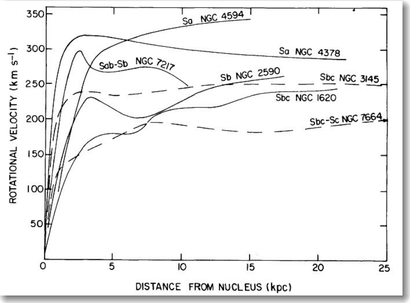
Nanjing Normal University

Table of contents

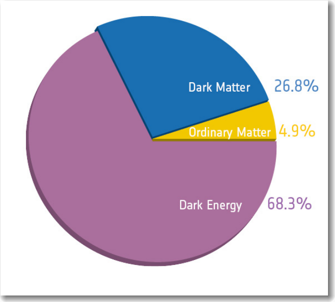
Dark matter & XENON1T

Dark matter experimental evidence

We all know the evidence for dark matter (DM) in gravitational interactions, e.g.



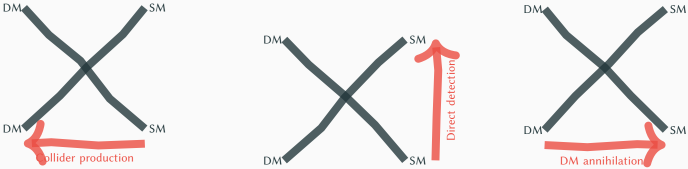
(I) Rotation curves (Rubin, Ford, and Thonnard 1980)



(II) CMB (Ade et al. 2016)

Make it, shake it, break it

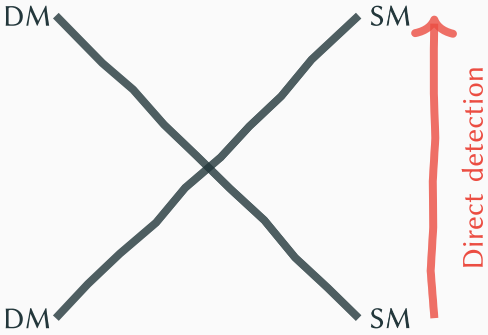
Three detection strategies



Motivated by crossing symmetries of amplitudes

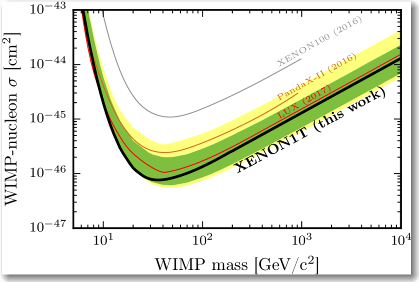
DM scatters with SM nucleons

We can search for DM in direct detection experiments. DM elastic scatters with nucleons in a detector on Earth.



There is a wind of WIMP particles from the Earth's motion in the dark matter halo.

The Panda (Cui et al. 2017), LUX (Akerib et al. 2016), XENON (Aprile et al. 2017) and PICO (Amole et al. 2017) experiments saw nothing, resulting in exclusion contours on the (mass, cross section) planes



Electron recoils

If you get bored of looking for WIMPs in nuclear recoils in your direct detection experiment, look for electronic ones!

Distinctive signature in XENON1T — different ionisation and scintillation characteristics mean that they can be distinguished from e.g., WIMP nuclear recoils

Axion-like particles

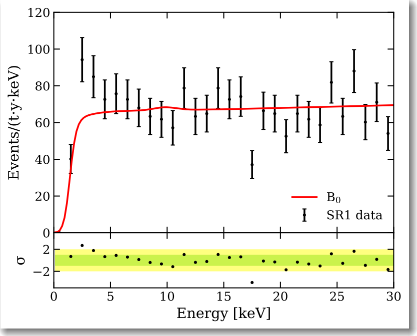
The QCD axion, a, is a neutral pseudo-scalar that could solve the strong CP problem through the PQ mechanism by the coupling

$$\mathcal{L} \propto \left(\frac{a}{f_a} - \Theta\right) F\tilde{F}$$

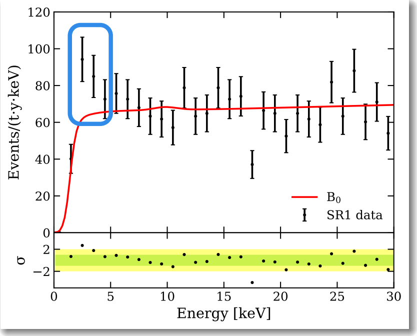
and a potential that generates the vacuum expectation value $\langle a \rangle = f_a \Theta$. More generally,

- Light neutral pseudo-scalars that couple to photons are called axion-like particles (ALPs)
- Usually too light to produce any observable nuclear recoil
- Might produce an electronic recoil

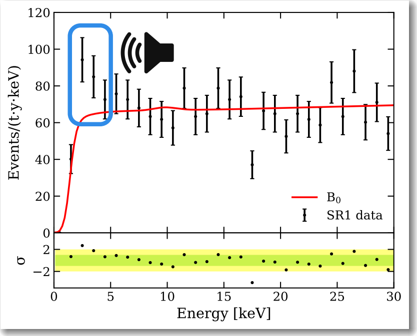
So XENON1T looked for electronic recoils (Aprile et al. 2020)



What's that?

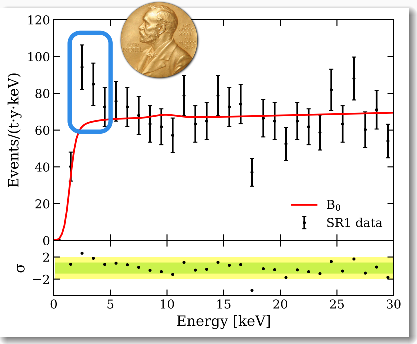

120
100
80
60
40
20
0
Events/(t·y·keV)
B0
SR1 data
2
−2
σ
0
5
10
15
20
25
30
Energy [keV]

Noise?

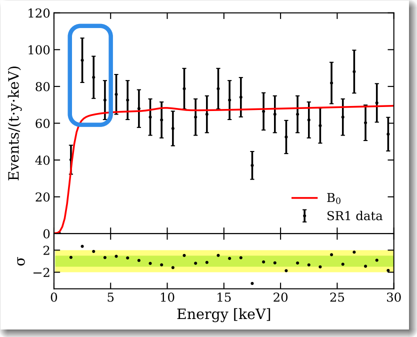


120
100
80
60
40
20
0
Events/(t·y·keV)
B0
SR1 data
2
−2
σ
0
5
10
15
20
25
30
Energy [keV]

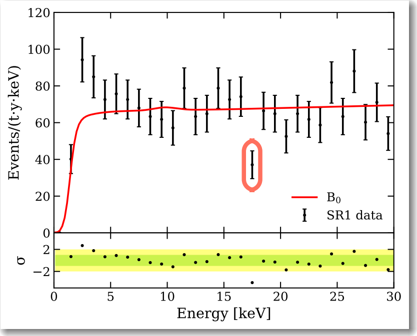
Nobel prize?


120
100
80
60
40
20
0
Events/(t·y·keV)
2
−2
σ
0
5
10
15
20
25
30
Energy [keV]
B0
SR1 data

Worth writing a paper about?

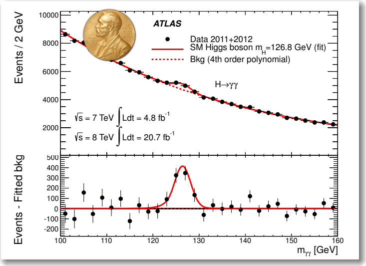


And what's that?

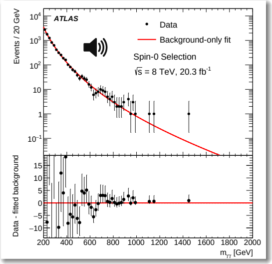


Noise or Nobel?

We've been here before



(I) Higgs discovery (Aad et al. 2013)



(II) 750 GeV bonanza! (Aad et al. 2015)

Recently, once it was Nobel (Higgs), many other times it was noise (750 GeV).

A global fit

In this case, XENON1T report 3.5σ evidence for solar ALPs. But is that compatible with what we already know about solar ALPs from other experiments?

We need a global fit — an analyis that looks at the model's whole parameter space and considers all relevant experimental constraints

Models

ALP models

We consider four models to explain the observations:

1. The background only model
2. The background only model + a tritium background
3. The background + solar ALPs
4. The background + DM ALPs

Background only

Unlike traditional WIMP searches involving nuclear recoils, significant backgrounds for electronic recoils from all kinds of radioactive isotopes in the detector

We take best-fit background from XENON1T paper.

Tritium

^{3}H background in the XENON1T experiment could give rise to an excess of events at about 1 keV – 15 keV. Fitting the anomaly with a tritium component requires about 5×10^{-25} mol/mol

In light of the uncertainties about the tritium level, we consider tritium fraction α_t

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

with a central value at the upper estimate of the level of tritium and a moderate standard deviation

Solar ALPs

Phenomenological solar ALP model with three independent couplings to

- photons ($g_{a\gamma}$)
- electrons (g_{ae})
- nucleons (g_{aN}^{eff})

The axion mass, m_a, is not a parameter in our solar ALP model, since the axions produced in the Sun are relativistic, $E_a \gg m_a$.

XENON1T solar ALP signal

Solar ALPs can be produced in the Sun by:

- Atomic recombination and de-excitation, Bremsstrahlung, and Compton (ABC)
- Primakoff (P)
- ^{57}Fe

They can interact in the detector by

- Axio-electric effect (aee)
- Inverse Primakoff effect (iP) (Gao et al. 2020; Dent et al. 2020)

The latter was not considered in the original XENON1T analysis.

XENON1T solar ALP signal

The individual components are scaled by the effective axion couplings. Schematically,

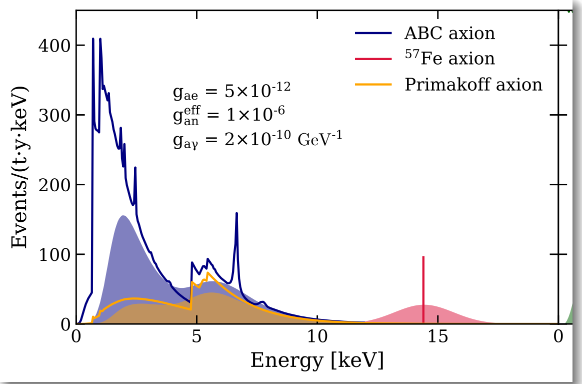
$$s = g_{ae}^2 \cdot \left(g_{ae}^2 \cdot s_{\text{ABC}}^{\text{aee}} + g_{a\gamma}^2 \cdot s_{\text{P}}^{\text{aee}} + (g_{aN}^{\text{eff}})^2 \cdot s_{\text{Fe}}^{\text{aee}} \right) +$$
$$g_{a\gamma}^2 \cdot \left(g_{ae}^2 \cdot s_{\text{ABC}}^{\text{iP}} + g_{a\gamma}^2 \cdot s_{\text{P}}^{\text{iP}}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

The notation is $s_{\text{Production}}^{\text{Detection}}$

We take the ABC, Primakoff and ^{57}Fe signal components from fig. 1 of (Aprile et al. 2020). We compute the inverse Primakoff contributions following (Dent et al. 2020).

XENON1T solar ALP signal

So, we scale the shaded areas up and down by the couplings and sum them together



DM ALPs

ALPs are viable DM candidates with a large parameter space spanning many orders of magnitude in mass and coupling (Arias et al. 2012; Marsh 2016).

We consider a phenomenological DM ALP model with 3 parameters:

- coupling to electrons (g_{ae})
- mass (m_a)
- fraction of the (local) DM around the Earth (η) that is made up of ALPs

We don't consider a photon coupling because of x-ray constraints or a neutron coupling as it doesn't impact the detection

XENON1T DM ALP signal

The DM ALP signal is given by

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

where σ_{pe} is the photoelectric cross section (Arisaka et al. 2013) and ρ_0 is the local DM density

A relic abundance of cold keV DM ALPs can be produced in the early Universe by the non-thermal vacuum realignment mechanism or thermally by the freeze-in mechanism

Several explicit models for a DM ALP with the required mass and Standard Model couplings (Takahashi, Yamada, and Yin 2020; Li 2020).

Are the solar and DM ALPs definitely different?

In general, solar ALPs *could* be DM ALPs at the same time but

- XENON1T needs electron recoil energies of more than about 1 keV
- DM ALPs in the halo are non-relativistic, so needs $m_a \gtrsim 1\,\mathrm{keV}$
- ALPs so heavy won't be produced in the Sun as typical energy scale keV

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.

Constraints

XENON1T & astrophysics

- The ALP interpretation of the XENON1T anomaly is in tension with astrophysical observables — we must include them in our analysis
- For each one, we construct a likelihood function. For a model with parameters $\boldsymbol{\Theta}$

  $$\mathcal{L}(\boldsymbol{\Theta}) = p\left(\text{Observed data} \,\middle|\, \text{Model}, \boldsymbol{\Theta}\right)$$

  This tells us the probability of the observed data assuming a particular model and set of parameters.
- The total likelihood function is just the product of XENON1T likelihood and the astrophysical ones

XENON1T

Construct XENON1T likelihood from binned data between 1 and 30 keV. A product of Poisson distributions,

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

where o_i are the observed counts; s_i are the signal predictions; b_i are the backgrounds, scaled by a factor α_b; and t_i is the tritium background, scaled by α_t

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 (Chen et al. 2017; Aprile et al. 2020).

Horizontal and Red Giant Branch stars

- Constraints on ALPs from the lifetime of Horizontal Branch (HB) and Red Giant Branch (RGB) stars (Raffelt 1996)
- ALPs easily escape the star, leading to an additional cooling channel
- Turned into a likelihood by measuring the ratio (R) of the number of HB and the number of RGB stars in e.g. Galactic globular clusters
- We use a likelihood based on (Giannotti et al. 2016), first implemented in (Hoof et al. 2019).
- Most robust and important astrophysical constraint

White Dwarf cooling hints

- Measurements of the period decrease in a number of pulsating WDs show anomalous cooling
- Consistent with an ALP-electron coupling of $g_{ae} \approx 3 \times 10^{-13}$ (Giannotti et al. 2016)
- Here we use a likelihood based on (Corsico et al. 2012b; Corsico et al. 2012a; Córsico et al. 2016; Battich et al. 2016), first implemented in (Hoof et al. 2019)
- Interpretation somewhat controversial; systematic uncertainties etc

For solar ALPs, WDs point to g_{ae} an order or magnitude *lower* than required by XENON1T

For DM ALPs, WDs point to g_{ae} an order or magnitude *higher* than required by XENON1T

DM ALP decays

- If ALPs constitute some or all of DM, their decays into photons would lead to potentially observable x-ray lines
- Strongest constraints stem from M31 (Horiuchi et al. 2014) and NuSTAR (Perez et al. 2017)

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

- Very strong constraint; we set $g_{a\gamma} = 0$ explicitly

SN1987A cooling

Further cooling constraints from supernova SN1987A, though we do not include SN1987A in our statistical analysis.

- SN1987A constrains axions and ALPs in numerous ways (Raffelt 1996) such as the neutrino cooling time
- Unfortunately, usually cited cooling bound not cast into a statistical statement
- Statistical statements would require full supernova simulations including ALPs

Methodology

What do we do?

We need a statistical methodology to judge evidence for a discovery taking into account the previous data

We consider:

1. Frequentist
2. Bayesian
3. Goodness-of-fit
4. Expected predictive accuracy

Error theoretic

Construct a rule so that you'd wrongly reject the null hypothesis at a pre-specified rate in the long-run in an ensemble of experiments

That is, control type-1 error rate. Type-1 error means we reject the null when it was in fact true.

We can't treat all models on equal footing — must specify a null — and doesn't consider only the evidence from this experiment — we have to think about an ensemble of repeats.

p-value

Compute a p-value

$$P\left(\text{data more or as extreme as that observed} \,\middle|\, \text{null hypothesis}\right)$$

Leap from the observed data to data at least as extreme. This can be challenging to compute.

To define more or as extreme, we introduce a test-statistic, λ, such that

$$p\text{-value} = P\left(\lambda \geq \lambda_{\text{Observed}} \,\middle|\, \text{null hypothesis}\right)$$

Well-motivated choice

$$\lambda = 2 \ln \frac{\max \mathcal{L}_1}{\max \mathcal{L}_0}$$

where we maximise over the model's parameters

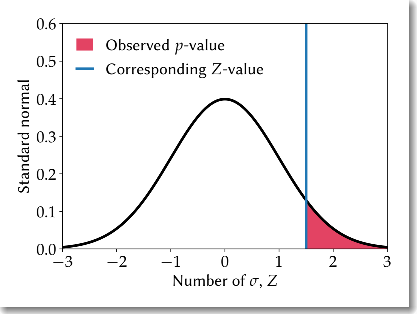
Interpreting p-values

Very popular in particle physics and elsewhere. Two possibly contradictory interpretations:

- p is a measure of evidence against H_0: small $p \Rightarrow H_0$ implausible.
- p is a means to control error rate: if we reject null when p-value ≤ 0.05, for example, becomes error theoretic approach with type 1 error rate 0.05.

5σ discovery threshold

In particle physics, it's common to translate p-values into Z-values. 5σ corresponds to about $p = 10^{-7}$. This is just a convention



5σ discovery threshold

In particle physics, it's common to translate p-values into Z-values. 5σ corresponds to about $p = 10^{-7}$. This is just a convention

It's common to hear that we require 5σ for a discovery. Two possibly contradictory interpretations:

- p is a measure of evidence against H_0: 5σ is a threshold on strength of evidence
- p is a means to control error rate: use 5σ as a desired type-1 long-run error rate

What the p-value isn't

The p-value itself isn't formally a measure of evidence or an error rate — it's a means to controlling an error rate. The error rate itself was specified prior to even collecting data (e.g., 5σ)

Yet it's widely misinterpreted as both of those things

The fact that academics don't know what p means is a symptom of the fact that p doesn't tell anything worth knowing (Wagenmakers 2020)

Bayesian approach

Compute the change in plausibility of the background only model relative to the signal model in light of the data

- If you like, you can compute the probability that you are making an error in the case at hand (cf. long-run error rates that are independent of the observed data)
- Typically shows weaker evidence for new physics than suggested by p-value (Fowlie 2019) — see also Lindley's paradox (Lindley 1957) and (Fowlie 2020)
- We just apply probability theory to the problem (Jeffreys 1939). Simple in theory; in practice there are difficulties.

Bayes factors

The Bayes factor (Kass and Raftery 1995) relates the relative plausibility of two models after data to their relative plausibility before data;

$$\textit{Posterior odds} = \textit{Bayes factor} \times \textit{Prior odds}$$

where

$$\textit{Bayes factor} = \frac{p\left(\textit{Observed data} \,|\, \textit{Model a}\right)}{p\left(\textit{Observed data} \,|\, \textit{Model b}\right)}$$

A nice result — by applying laws of probability, we see that models should be compared by nothing other than their ability to predict the observed data.

Bayesian evidence

The factors in the ratio are Bayesian evidences

$$\mathcal{Z} \equiv p\,(D \,|\, M) = \int_{\Omega_{\boldsymbol{\Theta}}} \mathcal{L}(\boldsymbol{\Theta})\, \pi(\boldsymbol{\Theta})\, \mathrm{d}\boldsymbol{\Theta},$$

where D is the observed data, $\mathcal{L}(\boldsymbol{\Theta}) = p\,(D \,|\, \boldsymbol{\Theta}, M)$ is the likelihood and $\pi(\boldsymbol{\Theta}) = P\,(\boldsymbol{\Theta} \,|\, M)$ is our prior, and $\boldsymbol{\Theta}$ are the model's parameters.

The prior describes what we knew about the parameters before seeing the data

The evidence is the likelihood averaged over the prior — the averaging penalises fine-tuned models

Priors

Many consider the dependence of the Bayes factor on the priors to be a major problem.

No priors, no predictions

I need to compare your model's predictions with data. If you don't tell the plausible parameters, how am I to know what it predicts?

Sensitive to arbitrary choices

If the inference changes dramatically within a class of reasonable priors, we can't draw reliable conclusions.

Science is hard; it's hard to get reliable knowledge about the world. We often disagree about the consequences of experimental data.

How could it be any other way?

Goodness of fit

Just compare goodness of fit by looking at $\Delta\chi^2$. In this context

$$\chi^2 \equiv -2 \max \ln \mathcal{L}$$

Pros:

- Simple to compute
- Simple to communicate

Cons:

- How to interpret it?
- How to calibrate it without computing *p*-value?

Box, Keynes & Spiegelhalter

Lastly, we will look at the DIC

- All models are wrong (Box 1976)!
- And who cares about the long run; in the long run, we're all dead (Keynes 1923)!

So (Spiegelhalter et al. 2002) doesn't care about which model is right or long-run error rates; only cares about which model likely makes the best predictions for future data!

DIC

So (Spiegelhalter et al. 2002) propose to measure the expected predictive accuracy of a model through the Deviance Information Criterion (DIC)

$$\mathrm{DIC} \equiv -2 \ln \mathcal{L}(\langle \boldsymbol{\Theta} \rangle) + 2p_{\mathcal{D}}$$

where $\langle \cdot \rangle$ indicates a posterior mean,

$$\langle \theta \rangle = \int \theta \, p(\theta \mid D, M) \, \mathrm{d}\theta,$$

where D is the data that we already observed and M denotes the model.

DIC

Notice the double use of the data in the term $\mathcal{L}(\langle \mathbf{\Theta} \rangle)$; first in the likelihood function and second in the estimate of the model's parameters.

The term,

$$p_{\mathcal{D}} \equiv 2\left(\langle \ln \mathcal{L}(\mathbf{\Theta})^2 \rangle - \langle \ln \mathcal{L}(\mathbf{\Theta}) \rangle^2\right)$$

corrects bias from over-fitting and is motivated by an analytic result for Gaussian posteriors

Computational methods

Computation performed with GAMBIT!

GAMBIT: The Global And Modular BSM Inference Tool

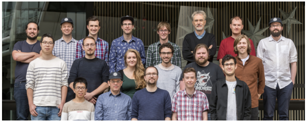
gambit.hepforge.org EPJC **77** (2017) 784 arXiv:1705.07908

- Extensive model database – not just SUSY
- Extensive observable/data libraries
- Many statistical and scanning options (Bayesian & frequentist)
- *Fast* LHC likelihood calculator
- Massively parallel
- Fully open-source

Members of:
ATLAS, Belle-II, CLiC, CMS, CTA, *Fermi*-LAT, DARWIN, IceCube, LHCb, SHiP, XENON

Authors of:
DarkSUSY, DDCalc, Diver, FlexibleSUSY, gamlike, GM2Calc, IsaTools, nulike, PolyChord, Rivet, SoftSUSY, SuperISO, SUSY-AI, WIMPSim

- Fast definition of new datasets and theories
- Plug and play scanning, physics and likelihood packages



Recent collaborators:
F Agocs, V Ananyev, P Athron, C Balázs, A Beniwal, J Bhom, S Bloor, T Bringmann, A Buckley, J-E Camargo-Molina, C Chang, M Chrzaszcz, J Conrad, J Cornell, M Danninger, J Edsjö, B Farmer, A Fowlie, T Gonzalo, P Grace, W Handley, J Harz, S Hoof, S Hotinli, F Kahlhoefer, N Avis Kozar, A Kvellestad, P Jackson, A Ladhu, N Mahmoudi, G Martinez, MT Prim, F Rajec, A Raklev, J Renk, C Rogan, R Ruiz, I Sáez Casares, N Serra, A Scaffidi, P Scott, P Stöcker, W Su, J Van den Abeele, A Vincent, C Weniger, M White, Y Zhang

70+ participants in 11 experiments and 14 major theory codes

Bayesian

- Numerical challenge — high-dimensional integration for Bayesian evidence
- Nested sampling algorithm for Bayesian computation (Skilling 2006)
- Implemented in `MultiNest` (Feroz, Hobson, and Bridges 2009) — ellipsoidal rejection sampling efficient for $d \lesssim 20$
- We performed state of the art cross checks (Fowlie, Handley, and Su 2020)

$\Delta\chi^2$

- Numerical challenge — find minimum of multi-dimensional likelihood function
- Used samples from nested sampling — for moderate dimensional problems and strict stopping conditions, reasonable performance as optimiser
- Meta-heuristic differential evolution algorithm, implemented in `Diver` (Martinez et al. 2017)

p-values

- Numerical challenge — compute a tiny tail probability
- Asymptotic approximations — Wilks' theorem (Wilks 1938) etc) — aren't strictly valid
- Ideally, requires expensive Monte Carlo simulations
- In some case, semi-analytic asymptotics for look-elsewhere effect are possible (Gross and Vitells 2010)

DIC

- Numerical challenge — compute posterior expectations
- Used weighted posterior samples returned by nested sampling algorithm

$$\langle \theta \rangle \approx \frac{\sum w_i \theta_i}{\sum w_i}$$

Results

Results

We consider data one by one

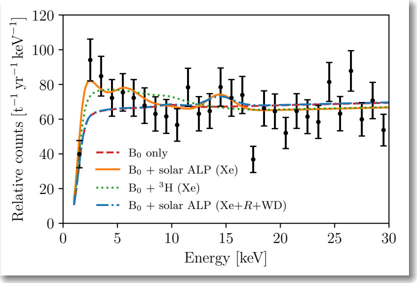
- XENON1T (Xe)
- Xe + R
- Xe + R + WD

for our solar ALP and DM ALP models and

- No ^3H
- ^3H in the signal and background models
- ^3H background only

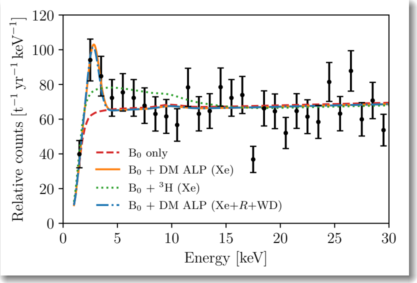
Best-fit Xe spectra

Solar ALPs fit Xe anomaly but once astrophysical data added, best-fit signal vanishes



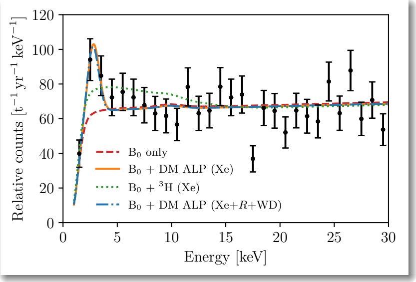
Best-fit Xe spectra

DM ALPs fit Xe anomaly even when simultaneously fit to astrophysical data



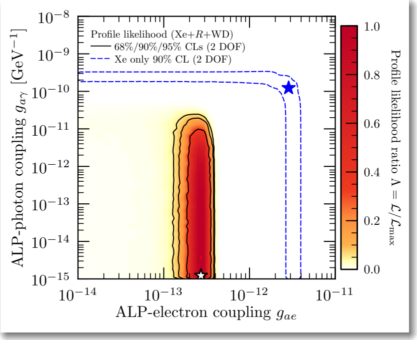
Best-fit Xe spectra

Tritium improves fit



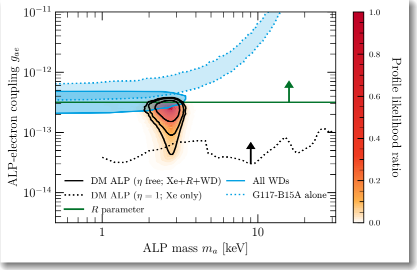
Solar ALP tension

The astrophysical data require considerably smaller couplings than Xe and rule it out



DM ALP resolution

DM ALP evade *R* constraint. By lowering DM fraction and increasing couplings, Xe signal unaffected but fits WD hints



Model comparison

From looking at the fits, we expect:

- DM ALP and solar ALP both favoured by Xe
- Tritium better than background but worse than ALPs
- But preference for solar ALP quashed by astrophysics
- DM ALP preference possibly enhanced by WD hints?

Let's see what happened.

$\Delta\chi^2$

Positive differences favor ALP models

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	45	0	67	0
Solar ALP	29	15	43	2	56	11
DM ALP	27	17	27	18	43	23
Background + ^{3}H	34	0	35	0	57	0
Solar ALP + ^{3}H	29	5	34	2	46	11
DM ALP + ^{3}H	26	9	26	9	42	15

$\Delta\chi^2$

Xe indeed favors ALPs

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	45	0	67	0
Solar ALP	29	15	43	2	56	11
DM ALP	27	17	27	18	43	23
Background + ^{3}H	34	0	35	0	57	0
Solar ALP + ^{3}H	29	5	34	2	46	11
DM ALP + ^{3}H	26	9	26	9	42	15

$\Delta\chi^2$

R parameter destroys preference for solar ALP but not DM ALP

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	45	0	67	0
Solar ALP	29	15	43	2	56	11
DM ALP	27	17	27	18	43	23
Background + ^{3}H	34	0	35	0	57	0
Solar ALP + ^{3}H	29	5	34	2	46	11
DM ALP + ^{3}H	26	9	26	9	42	15

$\Delta\chi^2$

WD hints favor ALPS and Xe and WD hints add together for DM ALP but not solar ALP

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	45	0	67	0
Solar ALP	29	15	43	2	56	11
DM ALP	27	17	27	18	43	23
Background + ^{3}H	34	0	35	0	57	0
Solar ALP + ^{3}H	29	5	34	2	46	11
DM ALP + ^{3}H	26	9	26	9	42	15

$\Delta\chi^2$

ALPs still preferred in presence of tritium

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	45	0	67	0
Solar ALP	29	15	43	2	56	11
DM ALP	27	17	27	18	43	23
Background + ^3H	34	0	35	0	57	0
Solar ALP + ^3H	29	5	34	2	46	11
DM ALP + ^3H	26	9	26	9	42	15

p-values

Too challenging to reliably compute!

- Asymptotic approximations that provide simple relations between $\Delta\chi^2$ and p aren't strictly applicable
- There is a look-elsewhere effect e.g., in the DM ALP mass
- Correction would require expensive simulations

Local significance for DM ALP from Xe + R + WD roughly 4.3σ

Bayes factors for solar ALP

Bayes factors show change in plausibility of solar ALP relative to background only. Partial Bayes factors (Xe | ·) show change induced by Xe given the astrophysical data. Greater (less) than one means ALP (background only) favored

	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 bkg only	0.52	0.051	0.25	0.19	0.18

Bayes factors for solar ALP

Xe mildly favored solar ALPs

	Xe	+= R	+= WD	(Xe $\mid$ R)	(Xe $\mid$ 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 bkg only	0.52	0.051	0.25	0.19	0.18

Bayes factors for solar ALP

But preference wiped out by 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 bkg only	0.52	0.051	0.25	0.19	0.18

Bayes factors for solar ALP

Partial Bayes factors show that Xe didn't tell us much new given the astrophysical data

	Xe	+= R	+= WD	$(\text{Xe} \mid R)$	$(\text{Xe} \mid R + \text{WD})$
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 bkg only	0.52	0.051	0.25	0.19	0.18

Bayes factors for solar ALP

Tritium slightly preferred over ALPs by Xe; more economical explanation

	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 bkg only	0.52	0.051	0.25	0.19	0.18

Bayes factor for DM ALP

Xe favors background only! DM ALP requires tuning mass and DM fraction etc

	Xe	+= R	+= WD	(Xe | R)	(Xe | R + WD)
No ^{3}H	0.8	0.8	2	1	3
^{3}H	0.5	0.5	0.7	0.9	1
^{3}H bkg only	0.1	0.2	0.4	0.3	0.6

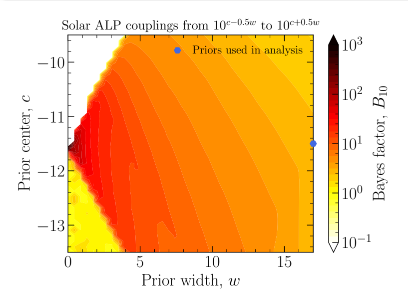
Bayes factor for DM ALP

Astrophysics mildly increase preference for DM ALP

	Xe	+= R	+= WD	(Xe | R)	(Xe | R + WD)
No ^{3}H	0.8	0.8	2	1	3
^{3}H	0.5	0.5	0.7	0.9	1
^{3}H bkg only	0.1	0.2	0.4	0.3	0.6

Prior dependence

Only weak evidence for any ALPs. How strong was the prior dependence? Vary the prior ranges for the solar ALP couplings.



Whilst the maximum is about 1500, that occurs only when the prior is carefully centred around the observed best-fits. Typically less than about 10

Differences in DIC

Finally, we turn to DIC. Positive differences favor ALP models

	Xe	(Xe + R)	(Xe + R + WD)
Solar ALP			
No ^{3}H	−5	0.002	−8
^{3}H	0.6	0.5	−10
^{3}H bkg only	−1	3	−4
DM ALP			
No ^{3}H	−30	−40	−40
^{3}H	−2	−3	−20
^{3}H bkg only	−20	−40	−40

Differences in DIC

Finally, we turn to DIC.

We found the DIC incoherent in this setting. E.g., R increases preference for solar ALPs?!

	Xe	(Xe + R)	(Xe + R + WD)
Solar ALP			
No ^{3}H	−5	0.002	−8
^{3}H	0.6	0.5	−10
^{3}H bkg only	−1	3	−4
DM ALP			
No ^{3}H	−30	−40	−40
^{3}H	−2	−3	−20
^{3}H bkg only	−20	−40	−40

Differences in DIC

Finally, we turn to DIC.

DIC was motivated by simple Gaussian case. Our problem deviated from that by the fat-tails in the distributions of allowed couplings

Couplings close to zero weren't completed ruled out.

Summary

- XENON1T anomaly could be explained by solar or DM ALPs
- Examined anomaly through several statistical methodologies
- Agreement that, given astrophysics, Xe provides at best weak evidence for solar ALP model
- Agreement that DM ALPs evade astrophysical constraints and even fit WD hints
- Disagreement about strength of evidence for DM ALP model — Bayesian analysis suggesting weaker evidence
- p-values were computationally challenging to compute

References

Aad, Georges et al. (2013). "Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC." In: *Phys. Lett. B* 726. [Erratum: Phys.Lett.B 734, 406–406 (2014)], pp. 88–119. arXiv: 1307.1427 [hep-ex].

– (Dec. 2015). "Search for resonances decaying to photon pairs in 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector." In:

Ade, P. A. R. et al. (2016). "Planck 2015 results. XIII. Cosmological parameters." In: *Astron. Astrophys.* 594, A13. arXiv: 1502.01589 [astro-ph.CO].

Akerib, D. S. et al. (2016). "Results on the Spin-Dependent Scattering of Weakly Interacting Massive Particles on Nucleons from the Run 3 Data of the LUX Experiment." In: *Phys. Rev. Lett.* 116.16, p. 161302. arXiv: 1602.03489 [hep-ex].

Amole, C. et al. (2017). "Dark Matter Search Results from the PICO-60 C_3F_8 Bubble Chamber." In: *Phys. Rev. Lett.* 118.25, p. 251301. arXiv: 1702.07666 [astro-ph.CO].

Aprile, E. et al. (2017). "First Dark Matter Search Results from the XENON1T Experiment." In: arXiv: 1705.06655 [astro-ph.CO].

– (2020). "Excess electronic recoil events in XENON1T." In: *Phys. Rev. D* 102.7, p. 072004. arXiv: 2006.09721 [hep-ex].

Arias, Paola et al. (2012). "WISPy Cold Dark Matter." In: *JCAP* 06, p. 013. arXiv: 1201.5902 [hep-ph].

Arisaka, K. et al. (2013). “Expected Sensitivity to Galactic/Solar Axions and Bosonic Super-WIMPs based on the Axio-electric Effect in Liquid Xenon Dark Matter Detectors.” In: *Astropart. Phys.* 44, pp. 59–67. arXiv: 1209.3810 [astro-ph.CO].

Athron, Peter et al. (July 2020). “Global fits of axion-like particles to XENON1T and astrophysical data.” In: arXiv: 2007.05517 [astro-ph.CO].

Battich, Tiara et al. (2016). “First axion bounds from a pulsating helium-rich white dwarf star.” In: *JCAP* 08, p. 062. arXiv: 1605.07668 [astro-ph.SR].

Box, George E. P. (1976). "Science and Statistics." In: *Journal of the American Statistical Association* 71.356, pp. 791–799. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1976.10480949. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1976.10480949.

Chen, Jiunn-Wei et al. (2017). "Low-energy electronic recoil in xenon detectors by solar neutrinos." In: *Phys. Lett. B* 774, pp. 656–661. arXiv: 1610.04177 [hep-ex].

Corsico, A.H. et al. (2012a). "An independent limit on the axion mass from the variable white dwarf star R548." In: *JCAP* 12, p. 010. arXiv: 1211.3389 [astro-ph.SR].

Corsico, Alejandro H. et al. (2012b). "The rate of cooling of the pulsating white dwarf star G117−B15A: a new asteroseismological inference of the axion mass." In: *Mon. Not. Roy. Astron. Soc.* 424, p. 2792. arXiv: 1205.6180 [astro-ph.SR].

Córsico, Alejandro H. et al. (2016). "An asteroseismic constraint on the mass of the axion from the period drift of the pulsating DA white dwarf star L19-2." In: *JCAP* 07, p. 036. arXiv: 1605.06458 [astro-ph.SR].

Cui, Xiangyi et al. (2017). "Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment." In: arXiv: 1708.06917 [astro-ph.CO].

Dent, James B. et al. (2020). "Inverse Primakoff Scattering as a Probe of Solar Axions at Liquid Xenon Direct Detection Experiments." In: *Phys. Rev. Lett.* 125.13, p. 131805. arXiv: 2006.15118 [hep-ph].

Feroz, F., M. P. Hobson, and M. Bridges (2009). "MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics." In: *Mon. Not. Roy. Astron. Soc.* 398, pp. 1601–1614. arXiv: 0809.3437 [astro-ph].

Fowlie, Andrew (2019). "Bayesian and frequentist approaches to resonance searches." In: *JINST* 14.10, P10031. arXiv: 1902.03243 [hep-ph].

– (Dec. 2020). "Objective Bayesian approach to the Jeffreys-Lindley paradox." In: arXiv: 2012.04879 [stat.ME].

Fowlie, Andrew, Will Handley, and Liangliang Su (June 2020). "Nested sampling cross-checks using order statistics." In: *Mon. Not. Roy. Astron. Soc.* 497.4, pp. 5256–5263. arXiv: 2006.03371 [stat.CO].

Gao, Christina et al. (2020). "Reexamining the Solar Axion Explanation for the XENON1T Excess." In: *Phys. Rev. Lett.* 125.13, p. 131806. arXiv: 2006.14598 [hep-ph].

Giannotti, Maurizio et al. (2016). "Cool WISPs for stellar cooling excesses." In: *JCAP* 05, p. 057. arXiv: 1512.08108 [astro-ph.HE].

Gross, Eilam and Ofer Vitells (2010). "Trial factors for the look elsewhere effect in high energy physics." In: *Eur. Phys. J. C* 70, pp. 525–530. arXiv: 1005.1891 [physics.data-an].

Hoof, Sebastian et al. (2019). “Axion global fits with Peccei-Quinn symmetry breaking before inflation using GAMBIT.” In: *JHEP* 03. [Erratum: JHEP 11, 099 (2019)], p. 191. arXiv: 1810.07192 [hep-ph].

Horiuchi, Shunsaku et al. (2014). “Sterile neutrino dark matter bounds from galaxies of the Local Group.” In: *Phys. Rev. D* 89.2, p. 025017. arXiv: 1311.0282 [astro-ph.CO].

Jeffreys, Harold (1939). *The Theory of Probability*. Oxford Classic Texts in the Physical Sciences. Oxford University Press. ISBN: 9780198503682.

Kass, Robert E. and Adrian E. Raftery (1995). “Bayes Factors.” In: *J. Am. Statist. Assoc.* 90.430, pp. 773–795.

Keynes, J.M. (1923). *A Tract on Monetary Reform*.

Li, Tianjun (July 2020). “The KSVZ Axion and Pseudo-Nambu-Goldstone Boson Models for the XENON1T Excess.” In: arXiv: 2007.00874 [hep-ph].

Lindley, D. V. (1957). “A Statistical Paradox.” In: *Biometrika* 44.1/2, pp. 187–192.

Marsh, David J. E. (2016). “Axion Cosmology.” In: *Phys. Rept.* 643, pp. 1–79. arXiv: 1510.07633 [astro-ph.CO].

Martinez, Gregory D. et al. (2017). “Comparison of statistical sampling methods with ScannerBit, the GAMBIT scanning module.” In: *Eur. Phys. J.* C77.11, p. 761. arXiv: 1705.07959 [hep-ph].

Perez, Kerstin et al. (2017). "Almost closing the νMSM sterile neutrino dark matter window with NuSTAR." In: *Phys. Rev. D* 95.12, p. 123002. arXiv: 1609.00667 [astro-ph.HE].

Raffelt, G.G. (1996). *Stars as laboratories for fundamental physics: The astrophysics of neutrinos, axions, and other weakly interacting particles.* Chicago & London: University Of Chicago Press, p. 686. ISBN: 9780226702728. URL: http://wwwth.mpp.mpg.de/members/raffelt/mypapers/199613.pdf.

Rubin, V. C., W. K. Ford Jr., and N. Thonnard (June 1980). "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/." In: *Ap.J.* 238, pp. 471–487.

Skilling, John (2006). "Nested sampling for general Bayesian computation." In: *Bayesian Analysis* 1.4, pp. 833–859.

Spiegelhalter, David J. et al. (2002). "Bayesian Measures of Model Complexity and Fit." In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64.4, pp. 583–639. ISSN: 13697412, 14679868. URL: http://www.jstor.org/stable/3088806.

Takahashi, Fuminobu, Masaki Yamada, and Wen Yin (2020). "XENON1T Excess from Anomaly-Free Axionlike Dark Matter and Its Implications for Stellar Cooling Anomaly." In: *Phys. Rev. Lett.* 125.16, p. 161801. arXiv: 2006.10035 [hep-ph].

Wagenmakers, E.J. (2020).

Wilks, S.S. (1938). “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses.” In: *Annals Math. Statist.* 9.1, pp. 60–62.