DAMNED Data analysis

Etienne Savalle[1], Aurélien Hees[2] and Peter Wolf[2]

[1] Université de Paris, CNRS, Astroparticule et Cosmologie, F-75013 Paris, France
[2] SYRTE, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, LNE, 75014 Paris, France

May 26, 2021

Based on **Phys. Rev. Lett. 126, 051301**

Scalar field theory action

The theory relies on an action where φ is the massive scalar field :

$$S = \int d^4x \frac{\sqrt{-g}}{c} \frac{c^4}{16\pi G} \underbrace{\left[R - 2g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - V(\varphi)\right]}_{\text{General Relativity + scalar field}}$$

$$+ \int d^4x \frac{\sqrt{-g}}{c} \underbrace{\left[\mathcal{L}_{SM}[g_{\mu\nu}, \Psi_i] + \mathcal{L}_{int}[g_{\mu\nu}, \varphi, \Psi_i]\right]}_{\text{Standard Model + scalar field}}$$

Deterministic scalar field

As we've seen in previous presentations, this leads to a deterministic scalar field oscillation :

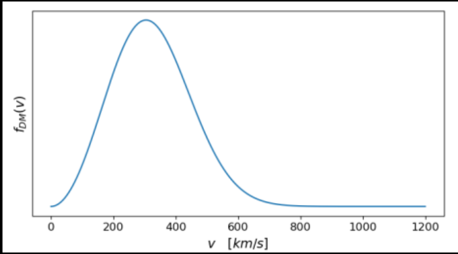
$$\varphi = \varphi_0 \cos(\omega_{\varphi_0} t + \delta_0) \tag{1}$$

Velocity distribution

During its formation, the dark matter halo has acquired a characteristic velocity distribution defined as :

$$f_{DM}(v) = \sqrt{\frac{2}{\pi}} \frac{v}{v_\odot \sigma_v} e^{-\frac{v^2+v_\odot^2}{2\sigma v}} \sinh\left(\frac{v v_\odot}{\sigma_v^2}\right) \tag{2}$$

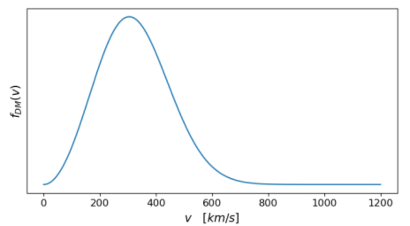
where σ_v the typical dispersion velocity and $v_\odot$ our Sun speed in the galactic halo frame.



Sum of multiple scalar fields

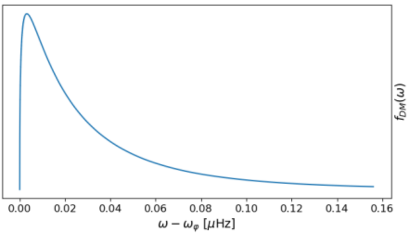
The scalar field acquires an additionnal kinetic energy following the velocity distribution.

$$\hbar\omega_\varphi \to \hbar\omega_\varphi + \frac{1}{2} m_\varphi v_{DM}^2$$



Oscillation frequency spread

This additionnal energy creates a shift in the scalar field oscillation frequency.



Stochastic scalar field

$$\varphi = \varphi_0 \sum_j \alpha_j \sqrt{f_{DM}(\omega_{\varphi j})\Delta\omega} \cos\left(\omega_{\varphi 0}\left(1 + \frac{v_j^2}{2c^2}\right) t + \delta_j\right)$$

where the amplitude α_j is following a Rayleigh distribution and the phase δ_j is drawn from an uniform distribution :

$$P\left[\alpha_j\right] = \alpha_j e^{-\alpha_j^2/2} \qquad P\left[\delta_j\right] = \frac{1}{2\pi} \quad \text{if} \quad 0 \leq \varphi_j < 2\pi. \tag{3}$$

Concise version

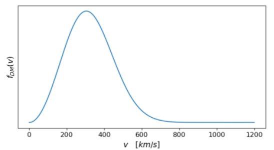
$$s(t) = \gamma \sum_j A_j \alpha_j \cos\left(\omega_j t + \delta_j\right) \tag{4}$$

with $A_j = M_j \frac{\varphi_0}{f_\varphi} \sqrt{f_{\mathrm{DM}}\left(f_j\right) \Delta f}$ and $\omega_j = 2\pi f_j = 2\pi f_\varphi \left(1 + \frac{v_j^2}{2c^2}\right)$

Distribution width

The dark matter distribution has a typical width δf such that :

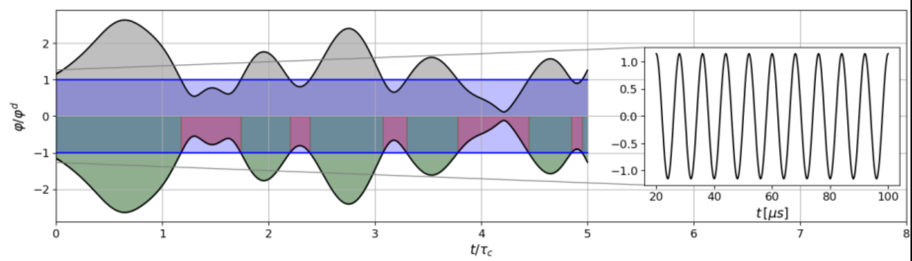
$$\delta_f = \frac{1}{2} f_\varphi \left[\frac{(v_{\max} + \delta_v)^2}{2c^2} - \frac{(v_{\max} - \delta_v)^2}{2c^2} \right] = f_\varphi \frac{v_{\max}\delta_v}{c^2} \simeq 10^{-6} f_\varphi \quad (5)$$



Coherence time

The scalar field coherence time is defined as :

$$\tau_c = (\delta f)^{-1} \simeq 10^6 P_\varphi \quad (6)$$



Scalar field amplitude enhancement/reduction

1. Green is good : we can expect a increase in the scalar field amplitude
2. Red is bad : we are not sensitive to the scalar field.
3. The coherence time depends on the frequency of the signal.
4. If not taken into account, the amplitude of the signal may be overestimated by almost an order of magnitude.
5. Looking in the PSD would reveal an asymmetric signature that is useful when there is suspicious peaks in the data.

How to search for the signal ?

1. Time domain frequentist approach : create multiple templates of the scalar field signal to fit in the data.
2. Frequency domain frequentist approach : simulate multiple versions of the experiment noise and evaluate the asymmetric scalar field signal coupling parameters.
3. Frequency domain bayesian approach : use bayesian inference using all the a-priori information on the signal to fit the true data.

Likelihood

$$\mathcal{L} = \frac{1}{\sqrt{\det(2\pi \mathrm{C})}} \exp\left(-\frac{1}{2N}(\tilde{\mathrm{d}} - \tilde{\mathrm{s}})^T \cdot \mathrm{C}^{-1} \cdot (\tilde{\mathrm{d}} - \tilde{\mathrm{s}})\right) \quad (7)$$

where C is the noise covariance matrix, $\tilde{d}$ is the Fourier transform of our data and $\tilde{s}$ is the Fourier transform of the dark matter signal.

Assuming white noise ($C = \sigma^2$), we have a likelihood that is linked to the χ^2 of our data :

$$\mathcal{L}(\mathrm{d} \mid \mathrm{s}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2} \sum_{k=\lfloor -N/2 \rfloor + 1}^{\lfloor N/2 \rfloor} \frac{\left|\tilde{d}(f_k) - \tilde{s}(f_k)\right|^2}{\sigma_k^2}\right) \quad (8)$$

χ_k^2 expression

In order to continue the calculation, we need to express χ_k^2 :

$$\tilde{\chi}_k^2 = \frac{\left|\tilde{d}_k - \tilde{s}_k\right|^2}{\sigma_k^2} = \frac{\left|\tilde{d}_k\right|^2}{N\sigma^2} + \frac{\gamma^2 N A_k^2 \alpha_k^2}{4\sigma^2} - \frac{\gamma A_k \alpha_k \left|\tilde{d}_k\right|}{\sigma^2} \cos\left(\delta_k + \theta_k\right) \quad (9)$$

with $\tilde{d}_k = \left|\tilde{d}_k\right| e^{i\theta_k}$.

Unknowns

This χ_k^2 depends on many unknowns :

1. γ the dark matter coupling parameter that we want to constrain,
2. N $\{\alpha_k\}$ the scalar field amplitude that are following the Rayleigh distribution.
3. N $\{\delta_k\}$ the scalar field random phases.

Marginalization

In order to only depend on γ, we need to marginalize the likelihood $\mathcal{L}(\tilde{d}|\gamma, \{^a lpha_k\}, \{\delta_k\}$ on the phases and amplitudes :

$$\tilde{\mathcal{P}}(\gamma \mid \tilde{\mathrm{d}}) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \prod_k \int_0^\infty d\alpha_k \alpha_k e^{-\alpha_k^2/2} \int_0^{2\pi} d\frac{\delta_k}{2\pi} e^{-\beta_k \tilde{\chi}_k^2} \tag{10}$$

All steps were analytical and we get a likelihood that depends only on the coupling parameter γ.

Posterior

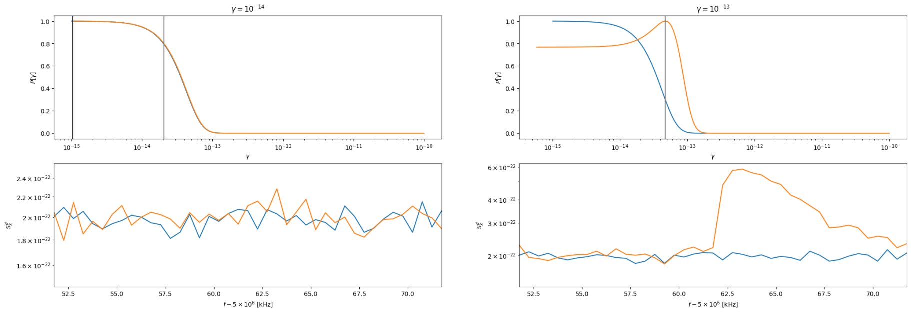
We are left with the logarithm of the posterior distribution of γ :

$$-\ln \mathcal{P}(\gamma \mid \tilde{d}) = \text{cst} + \sum_{k=0}^{N/2} \frac{\frac{S^d(f_k)}{S^n(f_k)}}{1 + \gamma^2 \frac{TA(f_k)^2}{2S^n(f_k)}} + \ln\left(1 + \gamma^2 \frac{TA(f_k)^2}{2S^n(f_k)}\right) \qquad (11)$$

Quantities

1. For a given frequency f_k :
   1. $S^d(f_k)$ is the experiment data power spectral density,
   2. $S^n(f_k)$ is the experiment noise power spectral density,
   3. $A(f_k)$ is the link between amplitude of the scalar field and your observable,
2. T is the experiment duration,
3. γ is what we are looking for.

Posterior plot



The posterior as a function of γ (top) is computed based on the power spectral density of the noise (blue) and signal (orange)

To extract the uncertainty, we compute :

$$\int_{-\gamma^{95}}^{\gamma^{95}} P(\gamma)\mathrm{d}\gamma = 2\int_{0}^{\gamma^{95}} P(\gamma)\mathrm{d}\gamma = 0.95 \quad (12)$$

N^2 scaling

The likelihood is a sum of N terms for one frequency.

$$\mathcal{L}(f_\varphi) \simeq \sum_k^N \mathcal{L}_k(f_\varphi) \tag{13}$$

With N frequencies, we expect N^2 computation ($N_{DAMNED} \simeq 10^7$). The computation time is increasing as the square of the experiment duration.

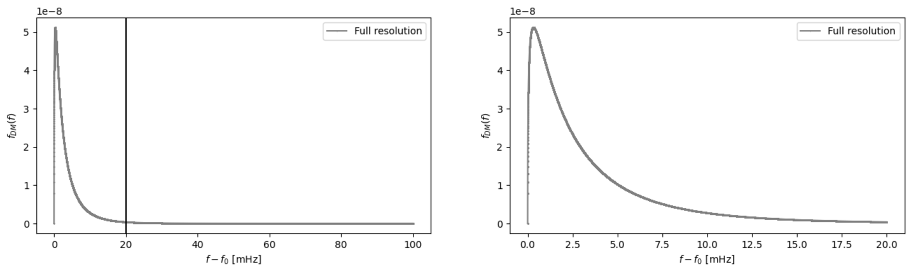
In the likelihood expression, the scalar fields amplitude are given by :

$$A_j = \frac{\varphi_0}{f_\varphi}\sqrt{f_{\mathrm{DM}}\left(f_j\right)\Delta f} \tag{14}$$

where the distribution $f_{\mathrm{DM}}\left(f_j\right)$ can be equal to zero.

To balance speed/physics, we choose to compute the likelihood in $\left[f_\varphi, f_\varphi + a\delta f\right]$

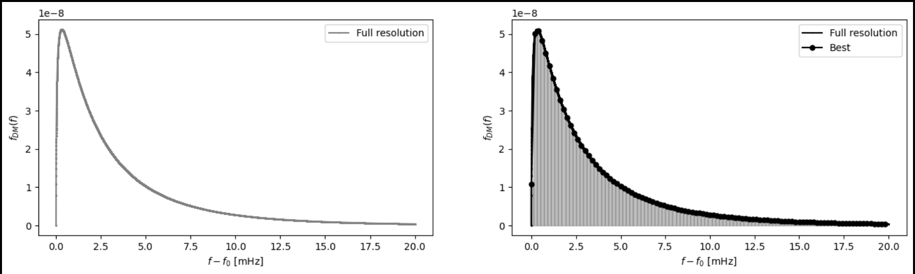
1. the lower bound comes naturally from the definition of f_{DM}
2. while the upper bound needs to be large enough to well represent the distribution.



Experiment resolution

For a given experiment duration T, we would expect to have a frequency resolution of $\Delta f = 1/T$. With this resolution, we can evaluate the number of frequencies N_j needed to best represent the distribution :

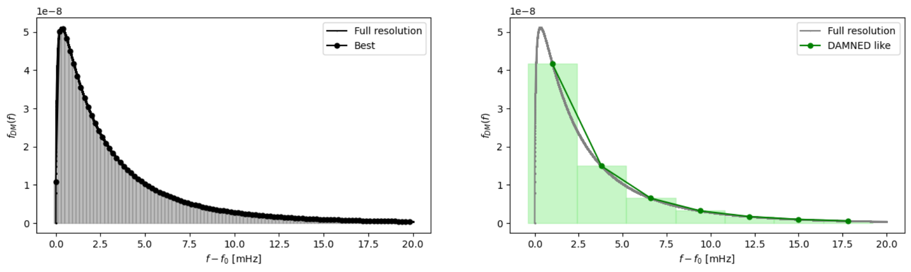
$$N_j = \left(\frac{f_\varphi + a\delta f}{\Delta f} - \frac{f_\varphi}{\Delta f} \right) = a\,\delta f\;T \qquad (15)$$



Memory

One drawback from increase duration is that we need to compute the PSD of the timeserie. Unfortunately, to perform the FFT method, we need the equivalent of at least $2 \times N$ points in memory.

In order to overcome this problem, we can compute the FFT by chunks which allows us to get the full duration but with a worse resolution (inverse of the chunk duration $1/T_b$).



Stochastic scalar field

It is important to take into account the stochastic nature of the scalar field in order to avoid any underestimation of the constraints put on some dark matter parameter.

Take home message

Life is about sacrifice...

1. Increase experiment duration :
   1. In order to improve sensitivity, a long experiment duration is needed.
   2. In order to well represent the dark matter velocity distribution, you need a fine frequency resolution ($\equiv$ long duration).
2. Decrease experiment duration :
   1. Overcome data storage issue,
   2. Acceleration of computing time,
   3. In some cases, make the actual FFT possible...