Improved hot dark matter bound on the QCD axion

Alessio Notari,[1, 2, *] Fabrizio Rompineve,[3, †] and Giovanni Villadoro[4, ‡]

[1]*Departament de Física Quàntica i Astrofisíca & Institut de Ciències del Cosmos (ICCUB),*
Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain
[2]*Galileo Galilei Institute for theoretical physics,*
Centro Nazionale INFN di Studi Avanzati Largo Enrico Fermi 2, I-50125, Firenze, Italy
[3]*CERN, Theoretical Physics Department, 1211 Geneva 23, Switzerland*
[4]*Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151, Trieste, Italy*

(Dated: November 9, 2022)

We strengthen the cosmological bound on the axion mass, by solving the momentum-dependent Boltzmann equations for axion-pion scatterings and by using a phenomenological production rate derived from pion-pion scattering data, overcoming the breakdown of chiral perturbation theory. Using present cosmological datasets we obtain $m_a \leq 0.24$ eV. To further improve the bound and exploit the reach of upcoming cosmological surveys, reliable non-perturbative calculations above the QCD crossover are needed.

I. INTRODUCTION

The Peccei-Quinn solution to the strong CP problem [1, 2] involves an extra field [3, 4], the axion a, coupled to QCD. In the cosmic evolution scatterings mediated by such couplings produce a background of "hot" relic axions [5, 6] (see also refs. [7–9] for early follow-ups), very much analogous to the neutrino background and distinct from cold axions produced via the misalignment mechanism [10–12] and the decay of topological defects [13–15].

Available cosmological datasets [16] already constrain light relics produced at temperatures slightly below the QCD crossover, i.e. $T \lesssim T_c \equiv 150$ MeV, while important sensitivity improvements of current [17–19] and upcoming [20–23] Cosmic Microwave Background (CMB) and Large Scale Structure (LSS) surveys will allow to probe production during the crossover. A QCD axion with mass $m_a \sim (0.01 - 1)$ eV would be produced precisely around these temperatures, and can thus be constrained/discovered with cosmology, independently of astrophysical probes (for SN1987A, see ref. [24] and refs. [25–27] for recent conflicting reassessments; for the most recent constraints from other sources see refs. [28–31]), in a region targeted also by direct searches [32, 33]. Extracting a reliable prediction for the relic abundance of such axions is thus of crucial observational relevance.

In this Letter, we significantly improve on previous such predictions in two aspects. First, the rapid and significant variation of the number of relativistic species $g_{*,S}$ around T_c implies that the axion spectrum can significantly deviate from the equilibrium distribution that has been employed so far. We account for this by solving the momentum-dependent Boltzmann equations for axion scatterings. This strategy is used for the relic neutrino abundance [34, 35], which however occurs when the variation of $g_{*,S}$ is much less relevant, thus causing a smaller distortion (see ref. [36] for a related work focused on heavy axions). Second, the axion production rate of interest for current datasets, dominantly due to scatterings with pions ($\pi\pi \leftrightarrow a\pi$), has been historically computed at leading order (LO) in chiral perturbation theory (χPT) [8] (see refs. [37–40] for the corresponding "hot dark matter" bound on m_a from CMB data). Nonetheless, recent work [41] finds that such a rate receives large one-loop corrections already at $T \gtrsim 60$ MeV, casting doubts on the validity of existing constraints. We overcome this obstacle by deriving the a-π rate from the experimental data on $\pi\pi \leftrightarrow \pi\pi$ scattering, via a simple rescaling of the corresponding cross sections, as dictated by the mixing between the axion and the pion. These two improvements allow us to place a robust upper bound on the axion mass using cosmological datasets.

Finally, we highlight previously neglected non-perturbative contributions at T_c and above, which hinder the reliability of perturbative approaches and call for a dedicated study to further improve the present bound and to correctly interpret the phenomenological implications of forthcoming experiments.

---

* notari@fqa.ub.edu

† fabrizio.rompineve@cern.ch

‡ giovanni.villadoro@ictp.it

II. BOLTZMANN EQUATIONS

We focus on the minimal model-independent axion interaction, $\mathcal{L}_{\text{int}} = \alpha_s a G\tilde{G}/(8\pi f_a)$, where G is the gluon field strength, with its dual $\tilde{G}$, α_s is the strong coupling constant and f_a is the axion decay constant, related to the axion mass by $m_a = 0.57$ eV$(10^7$ GeV$/f_a)$ [42].

An important novelty of our work is the use of the momentum-dependent Boltzmann equations to compute the actual spectrum of axions produced via scatterings. We find this to be necessary, for two reasons. First, the initial axion abundance may be negligible and production may always remain sub-thermal, such that the axion spectrum may always significantly differ from the equilibrium distribution. Second, even when reaching equilibrium, interaction rates depend on the axion momentum, so that high momenta decouple later than low momenta. Since $g_{*,S}$ decreases quickly and substantially with temperature around T_c, high momenta are less diluted with respect to photons than low momenta.

In the expanding Universe it is convenient to use comoving 3-momenta $\mathbf{p}$, such that the axion distribution function $f_{\mathbf{p}}$ remains constant in the absence of interactions. The Boltzmann equations then reads [6, 43]:

$$\frac{df_{\mathbf{p}}}{dt} = (1 + f_{\mathbf{p}})\,\Gamma^< - f_{\mathbf{p}}\,\Gamma^> \,, \tag{1}$$

where the axion, with negligible mass at the time of production, has physical energy $E = |\mathbf{p}|/R$ and four momentum $k^\mu = (E, \mathbf{k})$, R is the scale factor, t is cosmic time and the energy dependent rates $\Gamma^<$ and $\Gamma^>$ describe creation and destruction of an axion. Assuming thermal equilibrium of the QCD bath, the following non-perturbative relation holds:

$$\Gamma^> = e^{\frac{E}{T}}\Gamma^< = \frac{\Gamma^>_{\text{top}}}{2Ef_a^2}\,, \qquad \Gamma^>_{\text{top}} \equiv \int d^4x\, e^{ik^\mu x_\mu} \left\langle \frac{\alpha_s}{8\pi}G\tilde{G}(x^\mu)\frac{\alpha_s}{8\pi}G\tilde{G}(0) \right\rangle\,, \tag{2}$$

where $\langle \cdots \rangle$ stands for thermal average. At weak coupling the rates are dominated by $2 \leftrightarrow 2$ scatterings, with amplitude $\mathcal{M}$, leading to [43]:

$$\Gamma^< = \frac{1}{2E} \int \left(\prod_{i=1}^{3} \frac{d^3\mathbf{k}_i}{(2\pi)^3 2E_i} \right) f_1^{\text{eq}} f_2^{\text{eq}}(1 + f_3^{\text{eq}})(2\pi)^4\delta^{(4)}(k_1^\mu + k_2^\mu - k_3^\mu - k^\mu)|\mathcal{M}|^2\,, \tag{3}$$

where $f_i^{\text{eq}} \equiv (e^{E_i/T} \mp 1)^{-1}$ is the equilibrium distribution for bosons and fermions. For instance, for axion-pion scatterings, $\pi\pi \leftrightarrow \pi a$, the physical four momenta are k_1 and k_2 for the incoming pions, k_3 for the emitted pion (gluons or other particles can be treated analogously).

The axion relic abundance is then commonly expressed in terms of the effective number of (massless) neutrino species beyond the SM neutrinos $N_\nu \approx 3.044$ [44, 45]:

$$\Delta N_{\text{eff}} \equiv N_{\text{eff}} - N_\nu = \frac{8}{7}\left(\frac{11}{4}\right)^{\frac{4}{3}}\left(\frac{\rho_a}{\rho_\gamma}\right)_{\text{CMB}}\,, \tag{4}$$

where the energy density of axions, ρ_a, and photons, ρ_γ, should be evaluated at recombination, when the Cosmic Microwave Background (CMB) is generated. In the following we shall make contact with previous approaches in the literature, which use a momentum-independent Boltzmann equation (see Appendix A), through the averaged rate

$$\overline{\Gamma} \equiv \frac{1}{n^{\text{eq}}}\int \frac{d^3\mathbf{k}}{(2\pi)^3}\Gamma^<\,, \tag{5}$$

with $n^{\text{eq}} \equiv \int d^3\mathbf{k}/(2\pi)^3 f_{\mathbf{p}}^{\text{eq}}$. Furthermore, under the assumption of instantaneous decoupling of an initial equilibrium population of axions, eq. (4) is approximated as $\Delta N_{\text{eff}} \simeq 0.027\,[106.75/g_{*,S}(T_d)]^{4/3}$, with T_d defined by $\overline{\Gamma} = H|_{T=T_d}$ and $H \equiv \dot{R}/R$. As mentioned above, neither of these two approaches is justified in our case.

In our region of interest m_a is actually close to the temperature at recombination ($\lesssim$ eV). ΔN_{eff} should then be interpreted as the would-be abundance of axions, if they were massless. As we will show, the bound on the axion contribution to ΔN_{eff} will be significantly stronger than the massless species constraint, $\Delta N_{\text{eff}} \leq 0.28$ at 95% C.L. (Planck 18+BAO) [16].
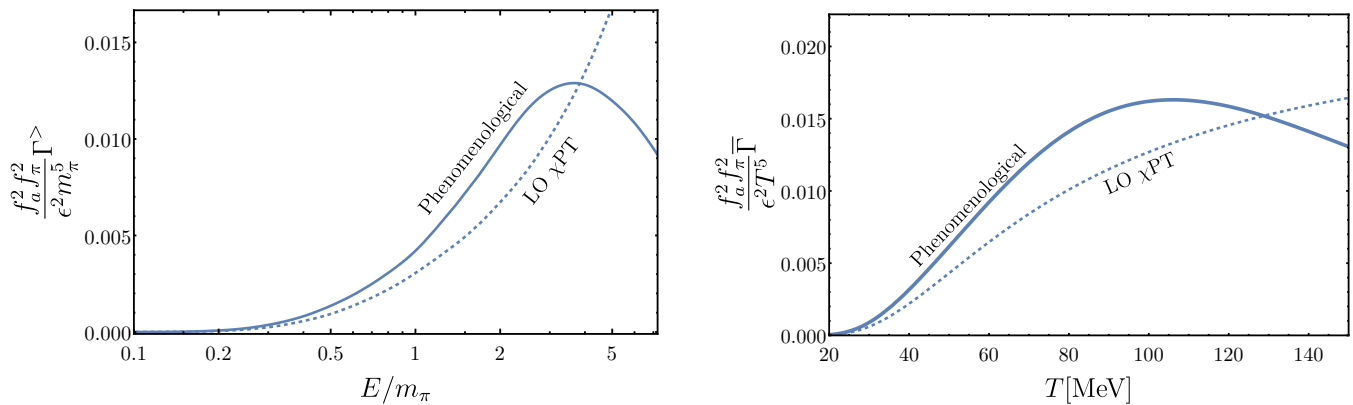
Figure 1. Scattering rates for $a\pi \leftrightarrow \pi\pi$ (solid curves for phenomenological rates, dashed for LO in χPT). *Left:* momentum dependent axion destruction rates at $T = 120$ MeV. *Right:* Averaged rates.

III.   AXION RATE BELOW T_c

At $T \lesssim T_c$ the QCD thermal bath is dominated by pions, all other hadrons being Boltzmann suppressed to some extent. As already stressed, recently ref. [41] pointed out that next-to-leading-order (NLO) corrections invalidate the LO χPT computation, already at $T \simeq 60$ MeV. In retrospect this is not surprising given that 1) the typical center of mass energy $\sqrt{s}$ for a pair of pions at such temperatures is already above 0.4 GeV and 2) the scattering amplitudes grow with energy, thereby pions with even higher energies, for which the χPT estimate is unreliable, are weighted more in the integral (5).

In fact the very same problem was encountered long ago for the pion damping rates [46–48], and it was circumvented by using experimental π-π scattering data directly, rather than relying on χPT. With this strategy, the pion rate was computed up to $T = T_c$ (beyond which the rate becomes rapidly comparable with the mass and pions cannot be considered as elementary particles anymore). In an attempt to rescue the χPT computation ref. [48] proposed also a unitarization approach. The idea is to apply the χPT expansion to (functions of) the scattering phase shifts rather than to the amplitude $\mathcal{M}$ itself. In this way unitarity is always respected and the growth at high energies tamed. The unitarization procedure is however not unique. While some choices agree well with experimental data, without the latter it would be hard to defend the use of one particular prescription over the others, or why it would be a good approximation to neglect even higher-order corrections.

Unfortunately, for axions there are no experimental data and at a first glance it seems that only unitarization could improve analytically on the fixed order χPT computation. However, we prove (see Appendix B) the following relation between a-π and π-π scattering amplitudes at all orders in χPT

$$\mathcal{M}_{a\pi^i \to \pi^j \pi^k} = \frac{\epsilon f_\pi}{2 f_a} \cdot \mathcal{M}_{\pi^0 \pi^i \to \pi^j \pi^k} + \mathcal{O}\left(\frac{m_\pi^2}{s}\right), \qquad \epsilon \equiv \frac{m_d - m_u}{m_d + m_u}, \tag{6}$$

where $f_\pi = 92.3$ MeV, $m_\pi = 138$ MeV (we use the average pion mass), m_u and m_d are up and down quark masses, $m_u/m_d \simeq 0.47$ [49]. We checked this explicitly at LO and NLO by comparing ref. [50] for π-π and refs.[8, 41] for a-π scattering amplitudes. The $\mathcal{O}(m_\pi^2/s)$ corrections near the two pions threshold can be computed directly at LO (see Appendix B) and they are at most $\mathcal{O}(10\%)$ and rapidly decrease at higher energies. While here we focused on the model-independent coupling of the QCD axion, eq. (6) can easily be extended to accommodate the most general axion couplings, see appendix B.

Thanks to eq. (6) phenomenological fits of π-π scattering data can be used to reconstruct a-π scattering amplitudes with a few percent precision up to T_c. To compute the momentum-dependent rate we applied eq. (6) to the phenomenological π-π partial wave amplitudes provided in refs. [51, 52] (specifically the S0, S2 and P waves, valid up to the two-kaon threshold $\sqrt{s} = 2m_K \simeq 1$ GeV, while we checked that higher partial waves give negligible contributions). The result is presented in Fig. 1 (solid blue curve, left panel), for a reference temperature $T = 120$ MeV. As expected, it decreases sharply at large momenta, in contrast to the LO a-π rate (dashed). The corresponding averaged rates are shown in Fig. 1 (right panel).

The decrease of the a-π rate at high temperatures is in part due to the opening of new scattering channels (see [47] for an analogous discussion for pions). Using the LO χPT Lagrangian, we checked for instance that scatterings with Kaons $\pi K \to a K$ are subleading below T_c, but would eventually dominate above $T \simeq 200$ MeV, see Appendix C (similar considerations may apply to scattering off nucleons). While in general a χPT calculation for these processes
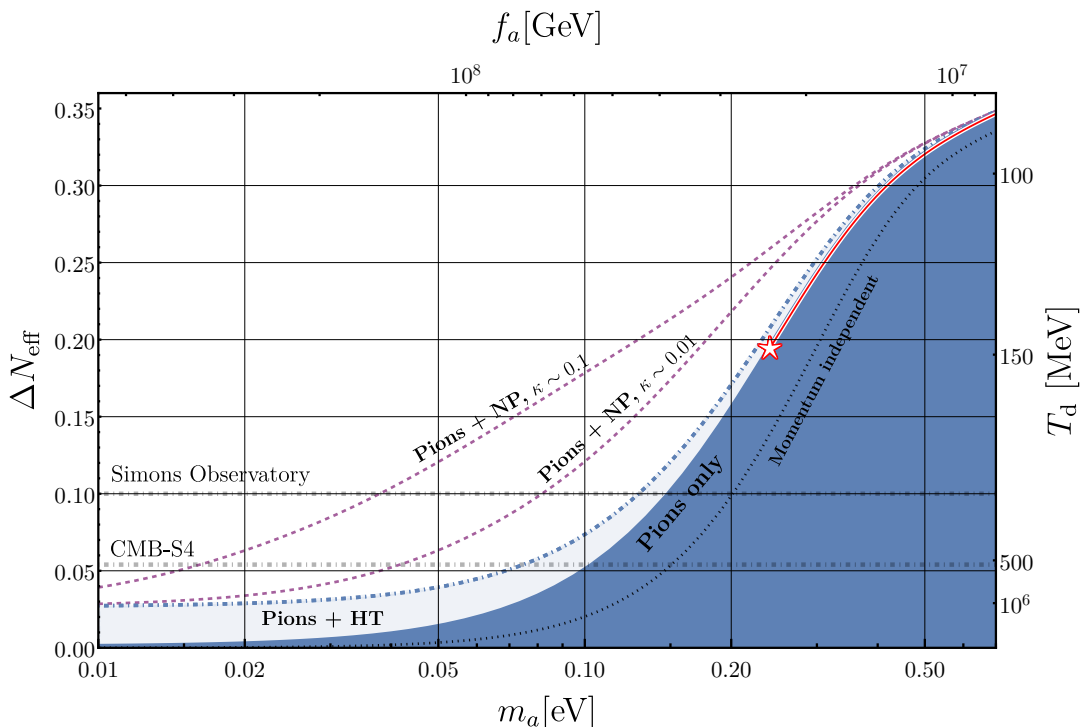
Figure 2. Relic axion abundance parameterized by $\Delta N_{\rm eff}$, for an axion minimally coupled to QCD. The blue region is our lower bound on $\Delta N_{\rm eff}$ obtained by using the full momentum-dependent Boltzmann equations with phenomenological $\pi\pi \leftrightarrow a\pi$ rate below T_c only, and with zero initial conditions at T_c. For comparison, results with the "standard" momentum-independent Boltzmann equation (dotted black) or assuming an equilibrium population of axions above $T = 1$ TeV ("Pions+HT", dot-dashed) are shown. Tentative dimensional analysis estimates for non-perturbative production up to $T = 2$ GeV are plotted as dashed purple curves ("Pions+NP"). On the right vertical axis, the temperature T_d corresponding to axions produced with instantaneous decoupling is plotted. The red starred point corresponds to the 95% C.L. bound that we obtain using CMB, BAO and Pantheon data. The 95% C.L. expected sensitivities of upcoming CMB surveys for massless species are shown by the dot-dashed gray lines.

is less reliable, such an estimate shows that the phenomenological a-π rate provides the most important contribution for the axion thermalization rate at low temperatures, representing a reliable lower bound at $T \lesssim T_c$.

Integrating numerically the momentum dependent Boltzmann equations (1) (see Appendix A) and assuming conservatively no extra production from $T > T_c$, we get the lower bound on $\Delta N_{\rm eff}$ represented by the boundary of the solid blue region in fig. 2. Because of the rapid change in $g_{*,S}$, during the axion decoupling, the resulting spectrum is distorted and ρ_a is enhanced compared to the standard calculation without momentum dependence (dotted black curve). A slightly less conservative assumption is to consider the axion in thermal equilibrium at some temperature $T > 1$ TeV (with $g_{*,S}$ saturated by the Standard Model). For this case, the bound is represented by the boundary of the light blue region in fig. 2.

Let us also mention that the axion coupling to pions can be slightly enhanced or suppressed with respect to the value considered here (i.e. corresponding to so-called hadronic models [53, 54]), in scenarios where the QCD axion also couples to quarks at high energies (e.g. [55, 56], see refs. [57–62] for relic abundance calculations). We do not consider these here, since they are generically more constrained by astrophysical observations [28] and the resulting $\Delta N_{\rm eff}$ is necessarily model-dependent.

IV.   ESTIMATES ABOVE T_c

While thermalization rates at the QCD confining scale are obviously non-perturbative, naively one would expect that at temperatures above a few GeV perturbative QCD should provide reliable rates. After all, at those temperatures the typical momentum of quarks and gluons is definitely in the perturbative regime. However, it has long been known that at finite temperature the convergence of perturbative QCD is worsened by various infrared divergences [63, 64]. One of them is manifest already at the leading order for the case at hand. The leading processes involve gluons

(g) and quarks (q): $ag \leftrightarrow gg$, $aq \leftrightarrow gq$ and $ag \leftrightarrow q\bar{q}$. The gluon exchange in the t channel leads to an infrared divergence [9, 65, 66] for the first two processes. At weak coupling ($g_s = \sqrt{4\pi\alpha_s} \ll 1$) the divergence is regulated by the gluon screening mass $m_g \sim g_s T$, leading to a logarithmic, but finite, enhancement in the rate proportional to $\log(T/m_g)$. However, g_s is still larger than unity at the electro-weak scale and only decreases logarithmically at higher energies. This makes the infrared logarithm and the total rate turn unphysically negative, unless the temperature is exponentially large.

This issue has been improved in ref. [66] (see also ref. [67]) using the full resummed thermal gluon propagator computed at LO in g_s. Such an approach however is not fully consistent, since vertex corrections are not properly resummed and gauge invariance not completely restored at the order considered. While ref. [66] used this method for temperatures $T \gtrsim 10^4$ GeV, where $g_s \lesssim 1$, ref. [68] pushed it all the way down to $T = 2$ GeV, where $g_s \simeq 2$.

It should be pointed out however that perturbation theory can really be trusted only for $g_s \ll 1$. One way to see this is to consider the ratio of the gluon thermal width over its electric mass, Γ_g/m_g. When this becomes $\mathcal{O}(1)$ the gluon is not a good weakly coupled degree of freedom. At leading order (see e.g. [43]) $\Gamma_g/m_g \sim 0.5g_s$, implying that perturbativity is really under control only at $T \gg$ TeV. Another requirement for the case at hand is the existence of a clear separation between T, the typical momentum scale of the thermal bath, and $g_s^2 T$, the momentum scale below which collective behaviors become important and gluon interactions become non perturbative. As long as $g_s \ll 1$ such soft momentum region is phase space suppressed, so that the perturbative expansion is meaningful up to a certain finite order [64]. When $g_s \gtrsim 1$ non-perturbative effects may become dominant and perturbation theory unreliable.

Accidentally, the non-perturbative axion damping rate at zero momentum has been studied in a different context. Such a quantity is indeed related via eq. (2) to the so-called strong-sphaleron rate $\Gamma_{\mathrm{sph}} \equiv \Gamma_{\mathrm{top}}^{>}(k^\mu = 0)$ [69]. While non-perturbative in nature, at small enough couplings its contribution has been inferred [70, 71] and checked numerically using classical field simulations [72], leading to the approximate scaling $\Gamma_{\mathrm{sph}} \simeq (N_c \alpha_s)^5 T^4$. If extrapolated at small energies this can become comparable to the putative perturbative computation at temperatures well above GeV (see Appendix E for more details).

We therefore conclude that axion-gluons perturbative calculations cannot be trusted below the electro-weak scale. In particular below the GeV scale, the most interesting region of temperatures for upcoming CMB experiments, the rate is definitely non-perturbative. Besides, between 150 MeV and 500 MeV the entropy and the energy densities have the largest variation associated to the QCD crossover [73]. This will enhance the effects of the spectral distortion possibly even more than that observed in our pion computation; it becomes therefore mandatory to study directly $\Gamma_{\mathrm{top}}^{>}(E)$ of eq. (2) with non-perturbative methods to further improve the bound and correctly assess the implications of forthcoming experiments on the axion parameter space.

At present we are not aware of any such computations in this regime, a challenging task beyond the aim of this work. All we can do is to attempt a very rough estimate of the averaged rate during the crossover, say for $T_c \lesssim T \lesssim 2$ GeV, using dimensional analysis, i.e. $\bar{\Gamma} \sim \kappa T^3/f_a^2$, where κ is an unknown coefficient. For illustrative purposes we choose two reference values $\kappa = 0.01$ and $\kappa = 0.1$, justified as follows.

First, at $T \gtrsim f_\pi$ the a-π rate just below T_c is $\sim 0.04\,\epsilon^2 T^3/f_a^2$ (see also Fig. 5 in Appendix C). The isospin breaking suppression ϵ is present only in the coupling to pions but not to other hadrons (see Appendix C for the unsuppressed Kaon cross section). Factoring that away and considering the contributions from other states above T_c, it is reasonable to expect $\kappa \sim \mathcal{O}(0.1)$ at around T_c and slightly above.

Second, recent attempts in pure $SU(3)$ lattice gauge theory estimate the rate for the zero mode $\Gamma_{\mathrm{sph}} \simeq \kappa_{\mathrm{latt}} T^4$ with κ_{latt} ranging from $\mathcal{O}(0.1)$ to $\mathcal{O}(0.01)$ as the temperature increases from T_c to a GeV [74, 75]. While the presence of light quarks would strongly suppress the contribution to the axion rate for the zero mode [69, 76], such a suppression is not expected to occur for the typical modes with $E \gtrsim T$.

The corresponding effect for ΔN_{eff}, assuming an averaged rate with $\kappa = 0.1$ or $\kappa = 0.01$, turned on only in the interval $T_c < T < 2$ GeV (in addition to the momentum dependent production below T_c), is represented by the dashed purple curves in fig. 2. We stress that our dimensional analysis estimate for the non-perturbative contribution is provided only to highlight its potential importance, especially for upcoming experiments.

V. CURRENT BOUND AND OUTLOOK

Using the "Pions only" curve in Fig. 2 we set a conservative upper bound on m_a from cosmological datasets. We employ the full Planck 2018 high-ℓ and low-ℓ TT, TE, EE and lensing likelihoods [77], and additionally Baryon Acoustic Oscillations (BAO) measurements from 6dFGS at $z = 0.106$ [78], from the MGS galaxy sample of SDSS at $z = 0.15$ [79], and from the CMASS and LOWZ galaxy samples of BOSS DR12 at $z = 0.38, 0.51$, and 0.61 [80], including the measurements of the rate of growth of structure [81], (we use the "consensus" likelihood of [80]), and
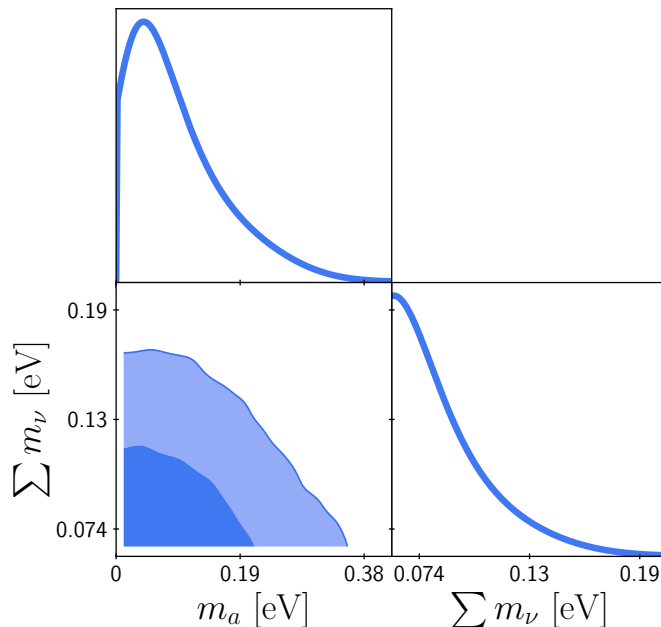
Figure 3. 1d and 2d posterior distributions for $\sum m_\nu$ and m_a. The darker/lighter shades correspond to 1σ and 2σ confidence level regions, extracted from the joint Planck18 TT+TE+EE+low ℓ+lensing+BAO+Pantheon likelihood.

the Pantheon Supernovae data [82].[1] We modified the Boltzmann solver `CLASS` [83, 84] to include the axion species and its actual distribution function, rather than using the Bose-Einstein distribution (see Appendix F for details). We ran a Markov Chain Monte Carlo (MCMC) analysis of the eight-parameter ΛCDM+$\sum m_\nu$+m_a cosmological model, where $\sum m_\nu \geq 0.06$ eV [49] is the sum of the neutrino masses, using the `MontePython` sampler [85, 86]. We find $m_a \leq 0.24$ eV, corresponding to $\Delta N_{\rm eff} \lesssim 0.19$ (red line with star in Fig. 2), and $\sum m_\nu \leq 0.14$, both at 95% C.L.; the 1d and 2d posterior distributions of m_a and $\sum m_\nu$ are reported in Fig. 3 (see Appendix F for results on all cosmological parameters). Our result for the axion mass is stronger than previous cosmological bounds in the literature, despite employing more conservative rates, because of the proper inclusion of momentum dependence. In particular, we do not rely on interpolated production rates between LO χPT and perturbative QCD amplitudes, such as in [87]. The use of the actual axion distribution function, rather than a Bose-Einstein distribution with the same energy density, leads to a minor relaxation in the axion mass bound ($\sim 4\%$) with the combination of datasets above.

Our conservative bound can be either weaker or stronger than the astrophysical constraints from globular clusters [29] (see [30] for a recent more constraining update), depending on the UV structure of the axion model. It is significantly weaker than the constraints of refs. [27, 31], which are however subject to astrophysical uncertainties. In contrast, our bound only relies on standard cosmology below T_c.

While we did not include production from the non-perturbative regime in our MCMC search in the conservative spirit of our work, Fig. 2 shows that it could significantly strengthen the bound already with current datasets. Even more importantly, as shown on the right vertical axis of Fig. 2, the upcoming Simons Observatory [20] and CMB-S4 [21] will probe axion production during the QCD crossover. In particular, they could reach $m_a \sim 0.01$ eV, and possibly even below, close to the region where cold axions can be the dark matter (note that the dashed purple curves in Fig. 2 do not include production above $T = 2$ GeV nor the enhancement due to the rapid variation of $g_{*,S}$ above T_c).

Such exciting possibilities motivate a dedicated study of axion production rates by non-perturbative methods, for arbitrary axion momenta, beyond the attempts made so far only for the zero mode.

---

[1] We do not include constraints from Big Bang Nucleosynthesis because of the associated theoretical and observational uncertainties (see [16] for a discussion), nor from Lyman-α forest observations because of the lack of dedicated simulations.

Appendix A: Boltzmann equations

Here we provide more details about the numerical integration of the Boltzmann equation (1). Rather than integrating over cosmic time t, we used the dimensionless variable $x \equiv m_\pi/T$, where the conversion between time and temperature is obtained by imposing entropy conservation, i.e. $d/dt(R^3 g_{*,S} T^3) = 0$, where at weak coupling $g_{*,S}(T) = \sum_{j=\text{bosons}} (\frac{T_j}{T})^3 g_j + \frac{7}{8} \sum_{j=\text{fermions}} (\frac{T_j}{T})^3 g_j$ includes all relativistic particles, with degrees of freedom g_j and possibly with their own temperature T_j. This amounts to the replacement $\frac{d}{dt} = (3xH g_{*,S})/(3g_{*,S} + T dg_{*,S}/dT)\frac{d}{dx}$. We take $g_{*,S}$ from the numerical lattice results of [73].

We then solved the resulting equation from $T_\text{i} = 150$ MeV to $T_\text{f} = 30$ MeV, for a discrete set of 150 comoving momenta, uniformly distributed between $|\mathbf{p}|_\text{min} = T_\text{f}/5 \cdot R(t_i)$ and $|\mathbf{p}|_\text{max} = 30\, T_\text{i} \cdot R(t_i)$, at the initial time t_i, so that physical momenta at any time t are comprised between $|\mathbf{k}|_\text{min} = |\mathbf{p}|_\text{min}/R(t)$ and $|\mathbf{k}|_\text{max} = |\mathbf{p}|_\text{max}/R(t)$. We computed the energy density in relic axions at T_f using the expression for a massless particle $\rho_a = R^{-4} \int d^3\mathbf{p}/(2\pi)^3 |\mathbf{p}| f_\mathbf{p}$.

Let us now discuss the commonly used momentum-independent approach. Starting from the momentum dependent rates (2), one introduces the damping rate $\Gamma \equiv \text{Im}\Pi_a^R/E = \Gamma^>/(1 + f_\mathbf{p}^\text{eq})$, with

$$\Pi_a^R \equiv i \int d^4x\, e^{ik^\mu x_\mu} \left\langle \left[\frac{\alpha_s}{8\pi f_a} G\tilde{G}(x^\mu),\ \frac{\alpha_s}{8\pi f_a} G\tilde{G}(0) \right] \Theta(t) \right\rangle, \tag{A1}$$

being the retarded axion self energy (see e.g. [88] for a review) and $\Theta(t)$ the Heaviside function. Then the Boltzmann equation (1) takes the simpler form

$$\frac{df_\mathbf{p}}{dt} = (f_\mathbf{p}^\text{eq} - f_\mathbf{p})\,\Gamma, \tag{A2}$$

By integrating this equation in $d^3\mathbf{p}/(2\pi)^3$, assuming $f_\mathbf{p}$ to be of the form (as explained in the main text this assumption is unjustified for the axion case):

$$f_\mathbf{p} = f_\mathbf{p}^\text{eq}\, n/n^\text{eq}, \tag{A3}$$

with $n = R^{-3} \int d^3\mathbf{p}/(2\pi)^3 f_\mathbf{p}$, and using the dimensionless variable x, one obtains the momentum-independent Boltzmann equation:

$$\frac{dY}{d\log x} = (Y^\text{eq} - Y)\frac{\overline{\Gamma}}{H} \left(1 - \frac{1}{3}\frac{d\log g_{*,S}}{d\log x} \right), \tag{A4}$$

where $Y \equiv n/s$ is the yield and $s = 2\pi^2 g_{*,S} T^3/45$ the entropy density. The final relic energy density in this approach is then given by:

$$\rho_a|_{x=x_f} = \frac{\pi^{14/3}}{30} \left(\frac{s\, Y}{\zeta_3} \right)^{4/3} \Big|_{x=x_f}. \tag{A5}$$

Such an equation is correct when a given massless species is thermal at high temperature and decouples instantaneously, retaining its Bose-Einstein form after decoupling. Note that, however, this is inconsistent in all other cases (e.g. when starting with zero abundance and never reaching equilibrium), and it is inconsistent with our previous assumption eq. (A3), since the energy density computed by inserting eq. (A3) into the integral for ρ_a, does not lead to eq. (A5).

One can visualize the spectral distortion of the actual distribution function obtained by solving the full momentum-dependent equation, by comparing with a Bose-Einstein distribution. For the comparison to be meaningful, we set the temperature of the equilibrium distribution such that the resulting energy density is the same as that obtained
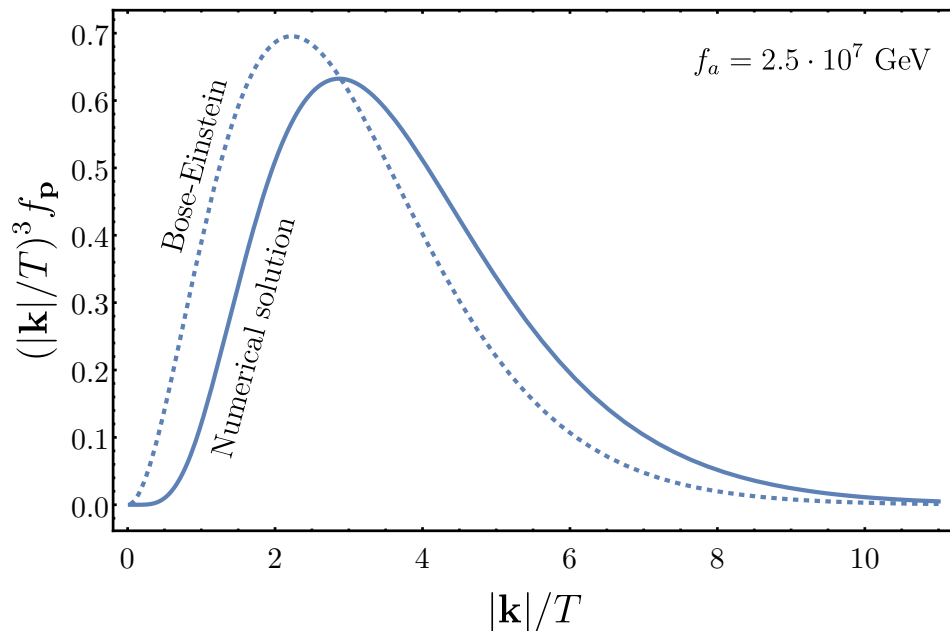
Figure 4. Distortion of the axion spectrum for an example value of f_a, as obtained by solving the momentum dependent Boltzmann equations with the rate from scatterings with pions only, below T_c (solid curve). The Bose-Einstein distribution (dashed curve) leads to the same energy density as the numerical solution lead to the same energy density, but the actual distribution is shifted to higher momenta.

with the actual distribution. The result is reported in Fig. 4 and shows that that the actual distribution is broader and peaked to higher momenta than the Bose-Einstein distribution.

In all cases, either using the exact treatment or the momentum-independent calculation, the relic axion energy density is translated into an effective number of neutrino species at $T_f = 30$ MeV, $\Delta N_{\rm eff}|_{T=T_f} \equiv \frac{8}{7}\frac{\rho_a}{\rho_\gamma}|_{T=T_f}$, then multiplied by $[g_{*,S}(T_f)/g_{*,S}(T_\nu)]^{-4/3}$, where $T_\nu = 3$ MeV corresponds to the temperature at which neutrinos start decoupling.

For the second analysis presented in this paper, i.e. including the potential production due to non-perturbative effects during the QCD crossover, we used a hybrid approach, motivated by the lack of a momentum-dependent rate during the QCD crossover. In a first step, we integrated the momentum independent equation (A4) from $T_i = 2$ GeV to $T_c = 150$ MeV, using an averaged rate $\bar{\Gamma}$ determined by dimensional analysis, as explained in the main text. We set initial conditions $Y|_{T=T_i} = Y^{\rm eq}|_{T=1~{\rm TeV}}$ and assumed $g_{*,S} = 106.75$ at $T \gtrsim$ TeV. The second step of the calculation was made using the momentum-dependent Boltzmann equation, with the corresponding rates for production due to pions only, as explained in section III. Initial conditions for the distribution function at T_c were set according to the would-be equilibrium distributions at $T_{\rm dec}$, defined as the temperature corresponding to the yield computed with the momentum-independent analysis above T_c, i.e.

$$T_{\rm dec} \equiv T_c \left(\frac{Y}{Y_{\rm eq}}\right)^{\frac{1}{3}}\Bigg|_{T=T_c}, \tag{A6}$$

and $f_{\bf p}|_{T=T_c} = f_{\bf p}^{\rm eq}|_{T=T_{\rm dec}}$.

Appendix B: Correspondence between $\pi\pi$ and $a\pi$ scattering

At leading order in $1/f_a$, the most general effective Lagrangian for the axion coupled to the lightest two quark fields $q = (u, d)$ can be written as

$$\mathcal{L} = \bar{q}\left(i\slashed{\partial} + \frac{c_0}{2f_a}\slashed{\partial}a\,\gamma_5\right)q - \bar{q}_L M_a q_R + h.c.\,, \qquad M_a \equiv \begin{pmatrix} m_u & 0 \\ 0 & m_d \end{pmatrix} e^{i\frac{a}{2f_a}(1+c_3\sigma^3)}, \tag{B1}$$

where $q_{R,L} \equiv \frac{1\pm\gamma_5}{2}q$ and c_0 and c_3 are the model dependent axion couplings to the isosinglet and isotriplet axial currents (e.g. for KSVZ models [53, 54] $c_3 = 0$, while for DFSZ models [55, 56] $c_3 = c_u - c_d$, where $c_{u,d}$ are the

couplings to the up and down quarks axial currents).

At low energies the effective theory for axion and pions reads:

$$\mathcal{L}_\pi = \frac{f_\pi^2}{4} \text{Tr} \left[\partial_\mu U \partial^\mu U^\dagger + 2B_0 (M_a U^\dagger + U M_a^\dagger) \right] + \dots \qquad U \equiv \exp(i\vec{\pi} \cdot \vec{\sigma}/f_\pi) \,, \tag{B2}$$

where $\vec{\sigma}$ are the Pauli matrices, $\vec{\pi}$ are the pion fields, the constant $B_0 = m_\pi^2/(m_u + m_d)$ at leading order and dots represent higher order terms in the derivative and quark mass expansion. Note that at leading order in the chiral expansion c_0 does not appear. Indeed since at leading order in the quark mass expansion the pion axial current is orthogonal to the isosinglet one, the contributions proportional to the coefficient c_0 will always be suppressed by at least one power of the quark masses, therefore they will not dominate at high energies.

At leading order the axion only appears in the mass term, which can be written also as

$$\frac{m_\pi^2 f_\pi^2}{2} \text{Re Tr} \left[(1 - \epsilon\sigma^3) \left(U e^{-i\frac{a}{2f_a}(1+c_3\sigma^3)} \right) \right] \,. \tag{B3}$$

In this frame there is an a-π^0 mixing that can be rotated away by an orthogonal rotation with angle $\theta_{a\pi}$:

$$\theta_{a\pi} \simeq \frac{(\epsilon - c_3) f_\pi}{2 f_a} \,, \tag{B4}$$

up to higher order terms in f_π/f_a. From this it follows that the diagonalization to mass eigenstates will introduce extra axion couplings (besides those originating directly from M_a and thus suppressed by the quark masses) from shifts of the neutral pions:

$$\pi^0 = \cos(\theta_{a\pi})\, \pi_\text{phys}^0 + \sin(\theta_{a\pi})\, a_\text{phys} \simeq \pi_\text{phys}^0 + \theta_{a\pi} a_\text{phys} \,. \tag{B5}$$

It follows that, at each order in chiral perturbation theory, the leading axion couplings at high energies (i.e. those with the maximum number of derivatives, not suppressed by light quark masses) will be proportional to the corresponding π^0 ones (the proportionality factor being given by the mixing angle itself), namely:

$$\mathcal{M}_{a\pi^i \to \pi^j \pi^k} = \theta_{a\pi} \cdot \mathcal{M}_{\pi^0 \pi^i \to \pi^j \pi^k} + \mathcal{O}\left(\frac{m_\pi^2}{s}\right) \,. \tag{B6}$$

The m_π^2/s corrections are the highest at low energy close to threshold. There the relative error can be estimated using the LO formula $|\mathcal{M}_\text{LO}|^2 = \epsilon^2(s^2 + t^2 + u^2 - 3m_\pi^4)/(4f_a^2 f_\pi^2)$ and is bounded by

$$\frac{\Delta|\mathcal{M}|^2}{|\mathcal{M}|^2} = \frac{\xi}{1 - 2\xi} \leq \frac{1}{11} \,, \qquad \xi \equiv \frac{3m_\pi^4}{2(s^2 + t^2 + u^2)} \,, \tag{B7}$$

where the bound is saturated at $s = 4m_\pi^2$, $t = u = -\frac{1}{2}m_\pi^2$. Therefore the asymptotic high energy formula is expected to hold also at low energies within 10% accuracy, better than the LO χPT formula.

Appendix C: Axion Kaon amplitudes

To estimate the importance of axion scattering with heavier hadrons below T_c we computed at leading order the scattering amplitude with single kaons ($K^\pm$, K^0 and $\overline{K}^0$), which are the next to lightest states after pions. The interactions are described by the same effective Lagrangian of pions in eq. (B2), the only difference being that the field is now defined as a 3×3 unitary matrix $U = \exp(i\vec{\lambda} \cdot \vec{\Pi})$ (with $\vec{\lambda}$ the Gell-Mann matrices and $\vec{\Pi}$ the eight meson fields including pions and kaons [50]), same for the axion dependent quark mass matrix $M_a = \text{diag}(m_u, m_d, m_s)e^{ia/(3f_a)}$, where m_s is the strange quark mass.

The total amplitude modulus squared for the processes $aK \to \pi K$ (which include $aK^\pm \to \pi^0 K^\pm$, $\pi^\pm K^0$; $aK^0 \to \pi^0 K^0$, $\pi^- K^+$; $a\bar{K}^0 \to \pi^0 \bar{K}^0$, $\pi^+ K^-$) at leading order and in the limit $m_\pi \to 0$ (consistent with the $m_\pi^2/s \to 0$ limit we used for the pions) turns out to be

$$|\mathcal{M}_{aK \to \pi K}|^2 = \frac{3\,t^2}{16 f_\pi^2 f_a^2}(1 - \epsilon^2) + \frac{(s - m_K^2)(s + t - m_K^2)}{2 f_\pi^2 f_a^2} \epsilon^2 \,, \tag{C1}$$

where in this case $s \equiv (p_a + p_K)^2$ and $t \equiv (p_a - p_\pi)^2$.

Note that the Boltzmann suppression in the scattering with kaons below T_c is partially compensated by the absence of the overall isospin suppression ϵ^2 present for the pions. Indeed the axion rate due to scatterings with kaons would become comparable to the pions one not much after T_c, at around $T = 200$ MeV (see fig. 5).
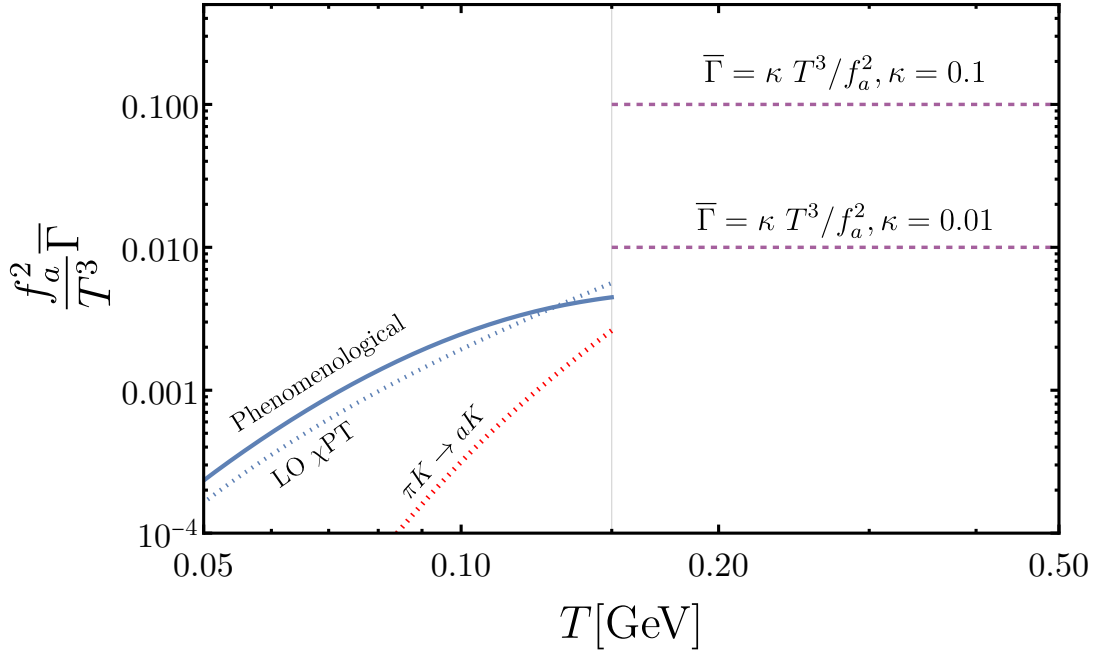
Figure 5. Tentative averaged axion production rate above T_c (see text for description), based on dimensional analysis. The phenomenological and LO rates for $a\pi$ and aK scattering below $T = 150$ MeV are also shown for comparison.

Appendix D: Computation of Scattering Rates

Here we provide more details about the numerical calculation of scattering rates. We start from the momentum-dependent rate of the destruction process $a\pi \to \pi\pi$, given by:

$$\Gamma^> = \frac{1}{2E} \int \left(\prod_{i=1}^{3} \frac{d^3\mathbf{k}_i}{(2\pi)^3 2E_i} \right) f_1^{\mathrm{eq}}(1 + f_2^{\mathrm{eq}})(1 + f_3^{\mathrm{eq}})(2\pi)^4 \delta^{(4)}(k^\mu + k_1^\mu - k_2^\mu - k_3^\mu)|\mathcal{M}|^2, \tag{D1}$$

where the subscripts $1, 2, 3$ refer respectively to the incoming pion and the two outgoing pions and $|\mathcal{M}|^2$ is the modulus squared of the scattering amplitude.

For the process of our interest, the integral above can be reduced to an integral over four variables. The steps are as follows, (we follow closely the analogous calculation for neutrino decoupling [34]): first, one uses the identity $d^3\mathbf{k}_3/(2E_3) = d^4k_3\,\delta(k_3^\mu k_{3,\mu} - m_3^2)\Theta(E_3)$ to integrate over k_3^μ. Then one introduces the scattering angles α, θ, β as:

$$\cos\alpha = \frac{\mathbf{k} \cdot \mathbf{k}_1}{|\mathbf{k}||\mathbf{k}_1|}, \qquad \cos\theta = \frac{\mathbf{k} \cdot \mathbf{k}_2}{|\mathbf{k}||\mathbf{k}_2|}, \qquad \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{|\mathbf{k}_1||\mathbf{k}_2|} = \cos\alpha\cos\theta + \sin\alpha\sin\theta\cos\beta, . \tag{D2}$$

The remaining integrations are over $d^3\mathbf{k}_1 = |\mathbf{k}_1|^2 d|\mathbf{k}_1|\,d\cos\alpha\,d\beta$, $d^3\mathbf{k}_2 = |\mathbf{k}_2|^2 d|\mathbf{k}|_2\,d\cos\theta\,d\eta$, where η is the polar angle associated to $\mathbf{k}_2$. For the process under consideration, the integration over η is trivial. Writing $k_3^2 - m_3^2 = h(\beta)$, the leftover δ function can be used to integrate over β. In order to do so, one determines the zeros of $h(\beta)$, which is in fact a linear function of $\cos\beta$. Therefore the roots have the same value of $\cos\beta$, and correspondingly two opposite values of $\sin\beta$. Since the integrand only depends on $|\sin\beta|$, the integral can be restricted from 0 to π, to select only one value of $\sin\beta$, then multiply by two. Furthermore, the condition $\cos^2\beta \le 1$ must be imposed. This leads to the following integral:

$$\Gamma^> = \frac{1}{2E}\frac{1}{(2\pi)^4} \int_0^\infty \frac{d|\mathbf{k}_1|}{2E_1} \int_0^\infty \frac{d|\mathbf{k}_2|}{2E_2} \int_{-1}^1 d\cos\alpha \int_{-1}^1 d\cos\theta\, f_1^{\mathrm{eq}}(1 + f_2^{\mathrm{eq}})(1 + f_3^{\mathrm{eq}})$$
$$\times \frac{2\Theta(E + E_1 - E_2)\Theta(1 - \cos\beta_0^2)}{|h'(\beta_0)|}|\mathbf{k}_1|^2|\mathbf{k}_2|^2|\mathcal{M}|^2, \tag{D3}$$

where f_3^{eq} is now a function of the integration variables. The modulus squared of the amplitude is a function of the Mandelstam variables $s = (k^\mu + k_1^\mu)^2$ and $t = (k^\mu - k_3^\mu)^2$ (u is eliminated using $s + t + u = 3m_\pi^2$), which can be expressed in terms of the integration variables above.

The LO χPT amplitude for $a\pi \to \pi\pi$ is given in Appendix B, above (B7). In order to obtain an amplitude from the phenomenological partial wave phases of $\pi\pi$ scattering [51, 52], the amplitude is decomposed in definite isospin $I = 0, 1, 2$ in the s-channel (see e.g. [50]):

$$T^0(s,t) = 3A(s,t,u) + A(t,u,s) + A(u,s,t), \tag{D4}$$

$$T^1(s,t) = A(t,u,s) - A(u,s,t), \tag{D5}$$

$$T^2(s,t) = A(t,u,s) + A(u,s,t), \tag{D6}$$

and then each amplitude is expanded into partial wave amplitudes t_l^I, according to:

$$T^I(s,t) = 32\pi \sum_{l=0}^{\infty} (2l+1) P_l(\cos\xi) t_l^I(s), \tag{D7}$$

where $t = (s - 4m_\pi^2)(\cos\xi - 1)/2$ and P_l is the l-th legendre polynomial. Finally, the experimental data are reported in terms of real phase shifts δ_l^I, related to the amplitudes by:

$$t_l^I(s) = \sqrt{\frac{s}{s - 4m_\pi^2}} \frac{e^{2i\delta_l^I(s)} - 1}{2i} = \sqrt{\frac{s}{s - 4m_\pi^2}} \frac{1}{\cot\delta_l^I(s) - i}. \tag{D8}$$

We took the phenomenological parametrizations for the $\delta_l^I(s)$ (specifically the δ_0^0, δ_1^1 and δ_0^2, we checked that the other waves give a negligible contribution to the rates) from Appendix A of [52], restricting ourselves to $\sqrt{s} \leq 2m_K$, where elastic waves can be used. We used two different phenomenological parametrizations for each of the two δ_0^0 and δ_0^2 shifts: one until $\sqrt{s} = 850$ MeV and a second one in the region 850 MeV $\leq \sqrt{s} \leq 2m_K$, following the analysis of [52]. The modulus squared of the total amplitude is then given by:

$$|\mathcal{M}|_{\pi\pi}^2 = \frac{1}{2} \frac{32\pi^2}{3} \left(|t_0^0|^2 + 3|3t_1^1 \cos\xi|^2 + 5|t_0^2|^2 \right). \tag{D9}$$

As discussed in the main text, the amplitude modulus squared above can be related to that of a-π scattering via rescaling by $\epsilon^2 f_\pi^2 / (4f_a^2)$.

We performed the four-dimensional integral in (D3) numerically, with the amplitudes discussed above, by using the `vegas` Monte Carlo integrator in Python, based on [89]. Specifically, we computed $\Gamma^>$ for a discrete set of values of $\log_{10}(|\mathbf{k}|/m_\pi) = [-3, 2]$ and $\log_{10}(m_\pi/T) = [-0.3, 1]$, with steps of 0.1 in both variables. We then built an interpolating function of two variables using `Mathematica` [90].

We also numerically computed the rate $\overline{\Gamma}$, according to (5) and (2):

$$\overline{\Gamma} = \frac{1}{n_{\text{eq}}} \int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{f_{\mathbf{P}}^{\text{eq}}}{1 + f_{\mathbf{P}}^{\text{eq}}} \Gamma^>. \tag{D10}$$

Note that this rate differs from the commonly used momentum-independent rate by a factor $1/(1 + f_{\mathbf{P}}^{\text{eq}})$ in the integrand. In practice, this difference is small. We followed the same procedure for the LO χPT rate for $aK \to \pi K$.

Appendix E: Rate from strong sphalerons

Here we evaluate the scale at which non-perturbative collective effects in the damping rate start being competitive with the perturbative ones in the simplified case of a pure $SU(N_c)$ theory without fermions.

A known source of such effects are "strong" sphalerons that for sufficiently weak coupling have been assessed to have a rate $\Gamma_{\text{sphal}} \simeq (N_c\alpha_s)^5 T^4$ [72]. Such a rate should be identified with what we called $\Gamma_{\text{top}}^>(k^\mu = 0)$ in eq. (2). Sphalerons are known to have a size of order $1/(N_c\alpha_s T)$, therefore we expect $\Gamma_{\text{top}}^>(E = |\mathbf{k}| < |\mathbf{k}_s|) \simeq \Gamma_{\text{sphal}}$ with $|\mathbf{k}_s| \sim N_c\alpha_s T$. Using eqs. (2) and (5) we get

$$\overline{\Gamma}_{\text{sphal}} = \frac{1}{n^{\text{eq}}} \int^{|\mathbf{k}|<|\mathbf{k}_s|} \frac{d^3\mathbf{k}}{(2\pi)^3 2E} \frac{\Gamma_{\text{sphal}}}{f_a^2} e^{-E/T} = \frac{(N_c\alpha_s)^5 T^3}{4\zeta_3 f_a^2} \left(1 - \left(1 + \frac{|\mathbf{k}_s|}{T} \right) e^{-|\mathbf{k}_s|/T} \right). \tag{E1}$$
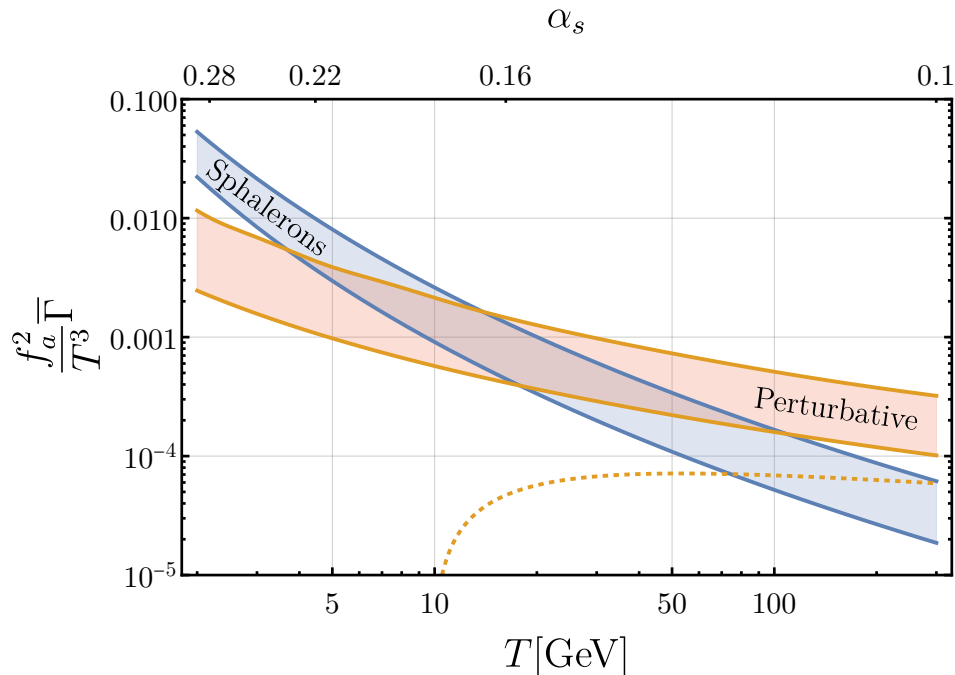
Figure 6. Comparison between non perturbative and perturbative contributions to axion rates as a function of temperature, see text for more details.

For $N_c = 3$ such a rate can be compared with the perturbative estimates performed in QCD in refs. [9, 65, 66, 68], which can be summarized as

$$\overline{\Gamma}_{\text{pert}} = \frac{\alpha_s^2 T^3}{4\pi^3 f_a^2} F_3 \tag{E2}$$

where $F_3 = g_s^2 \log(3/(2g_s))^2$ in ref. [65] and some more non-trivial function in refs. [66, 68].

In fig. 6 we compare $\overline{\Gamma}_{\text{sphal}}$ for the two choices $|\mathbf{k}_s| = 3\alpha_s T$ (lower blue line) and $|\mathbf{k}_s| = 6\alpha_s T$ (upper blue line), with $\overline{\Gamma}_{\text{pert}}$ for the three choices: F_3 from ref. [65] (orange dashed), $F_3 = g_s^2$ (lower orange curve), F_3 from ref. [68] (upper orange curve). Notice that the choice from ref. [65] breaks down already at moderate values of the coupling, due to the infrared logarithmic effects running negative. The choice $F_3 = g_s^2$ represent the expected size of the naive perturbative computation if the logarithmic factor saturates at unity. The last choice instead implement some partial resummation enhancing the rate.

In any case it appears that collective sphaleron-like effects start becoming important between $T \sim 100$ GeV and $T \sim 5$ GeV, anyway much before the breaking of QCD perturbation theory in vacuum.

The inclusion of (light) fermions in the computation of the non-perturbative effects is technically more challenging[2], but we do not expect that their presence could sensibly improve the reliability of the pertubative computation.

Appendix F: Cosmological analysis

Here we outline our implementation of hot axions into the Boltzmann solver `CLASS`, as well as provide extended results of our MCMC analysis.

We implemented the axion as an additional non-cold dark matter degree of freedom (`ncdm`) in `CLASS`, including its spectral distribution obtained by solving the momentum-dependent Boltzmann equation for the "Pions only" calculation. Specifically, we added a new distribution to the `background.c` file. We found that the integrand of the

---

[2] The effects of sphalerons are known to be damped by massless fermions at $|\mathbf{k}| = 0$, however it is easy to check that this is no longer true at $|\mathbf{k}| > 0$, in particular for the values relevant for us.

Parameter	Planck18+BAO+Pantheon
$100\omega_b$	$(2.252)\ 2.246^{+0.014}_{-0.015}$
ω_{cdm}	$(0.119)\ 0.120^{+0.001}_{-0.002}$
$100\ \theta_s$	$(1.042)\ 1.042^{+0.0003}_{-0.0003}$
$\ln 10^{10} A_s$	$(3.051)\ 3.054^{+0.015}_{-0.016}$
n_s	$(0.968)\ 0.968^{+0.004}_{-0.005}$
τ_{reio}	$(0.059)\ 0.058^{+0.007}_{-0.008}$
$\sum m_\nu$	≤ 0.14 (95% C.L.)
m_a	≤ 0.24 (95% C.L.)
$H_0[\mathrm{km/s/Mpc}]$	$(68.17)\ 67.75^{+0.51}_{-0.54}$
S_8	$(0.819)\ 0.821^{+0.011}_{-0.011}$

Table I. The mean (best-fit in parenthesis) $\pm 1\sigma$ error of the cosmological parameters obtained by fitting the $\Lambda\mathrm{CDM} + \sum m_\nu + m_a$ model to the cosmological datasets Planck18+BAO. Upper bounds are presented at 95% CL. H_0 and S_8 are derived parameters from the eight cosmological parameters.

axion energy density, i.e. $|\mathbf{k}|^3 f_{\mathbf{p}}$, is well approximated by the following functional form (in particular in the most important region $|\mathbf{k}| \gtrsim T$, see Fig. 4), in the range of axion masses of interest:

$$|\mathbf{k}|^3 f_{\mathrm{approx},\mathbf{p}} = |\mathbf{k}|^3 \frac{A}{e^{\frac{|\mathbf{k}|}{bT}+\mu} + 1}, \tag{F1}$$

where A, b and μ are parameters that vary with the axion mass. Their behavior for $m_a \leq 0.38$ eV can be simply described by polynomial fits. Notice that the actual distribution function does not grow at small $|\mathbf{k}|$, in contrast to a standard Bose-Einstein distribution. This different behavior is also visible in Fig. 4. This feature at small $|\mathbf{k}|$ is due to the fact that low momenta are not significantly produced for the values of f_a of interest, and that our initial conditions are set to be vanishing at T_c. This is the reason why the functional form (F1), with $+1$ in the denominator, fits the actual distribution function better than the analogous one with -1 in the denominator.

We have thus implemented the axion distribution function as $f_{\mathrm{approx},\mathbf{p}}$. Notice that this is a function of $q \equiv |\mathbf{k}|/T$ and m_a only. Additionally, the `ncdm` module requires to specify a temperature parameter T_{ncdm}, which for non-equilibrium distributions is just a normalization parameter. Accordingly, we have set it to be equal to the temperature of one dark species in units of the photon temperature, i.e. $(4/11)^{1/3}$. In practice, since we obtain our distribution function at $T_f = 30$ MeV rather than at the temperature of neutrino decoupling $T \approx 3$ MeV, we have set $T_{\mathrm{ncdm,\ a}} = (4/11)^{1/3}[g_{*,S}(3\ \mathrm{MeV})/g_{*,S}(30\ \mathrm{MeV})]^{1/3}$.

The output value of ΔN_{eff} obtained with our modified version of `CLASS` matches the result of Fig. 2 to within $\lesssim 1\%$ for $0.05\ \mathrm{eV} \leq m_a \leq 0.38$. For smaller axion masses the value of ΔN_{eff} is anyway negligible, see Fig. 2.

For $m_a \geq 0.38$, i.e. in the axion mass range which is anyway strongly constrained by CMB data according to our Fig. 2, we neglected the spectral distortion in `CLASS` and set the axion to have a Bose-Einstein distribution, with temperature given by:

$$T_{\mathrm{ncdm},a}(m_a) = T_\nu \left[\frac{7}{4}\Delta N_{\mathrm{eff}}(m_a)\right]^{\frac{1}{4}}, \tag{F2}$$

where $T_\nu = (4/11)^{1/3}$ eV is the temperature of one neutrino species and we used our result for "Pions only" in Fig. 2 for $\Delta N_{\mathrm{eff}}(m_a)$. We again found an analytical function that fits our result for ΔN_{eff} well and implemented it in `CLASS`.

For the neutrino sector, we followed the Planck collaboration convention of using three massive neutrinos with degenerate masses, and let their sum free to vary in our MCMC analysis. We set a prior $m_\nu \geq 0.02$ eV to take into account the lower bound on the sum of neutrino masses $\sum_\nu m_\nu \geq 0.06$ eV from neutrino oscillations [49].

We thus obtain a cosmological model with the six standard ΛCDM parameters plus the additional axion mass $m_a \geq 0$ and the sum of neutrino masses $\sum_\nu m_\nu \geq 0.06$ eV. We perform our MCMC analysis using `MontePython`, with the full Planck18 TT+TE+EE+low ℓ+lensing likelihood, the BAO likelihoods `bao_fs_boss_dr12` and `bao_smallz_2014`, as well as the `Pantheon` likelihood. We run ten chains, achieving Gelman-Rubin factors $R-1 \leq 0.005$ for all parameters. The resulting one- and two-dimensional posterior distributions for all parameters are shown in Fig. 7 (including H_0 and $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3}$. The mean $\pm 1\sigma$ error values for the ΛCDM parameters, together with the 95% C.L. bound for m_a, are reported in Table I. Additionally, we have checked that our upper bound on the axion mass is unaltered if we use the base ΛCDM model with only one massive neutrinos and $m_\nu = 0.06$ eV.
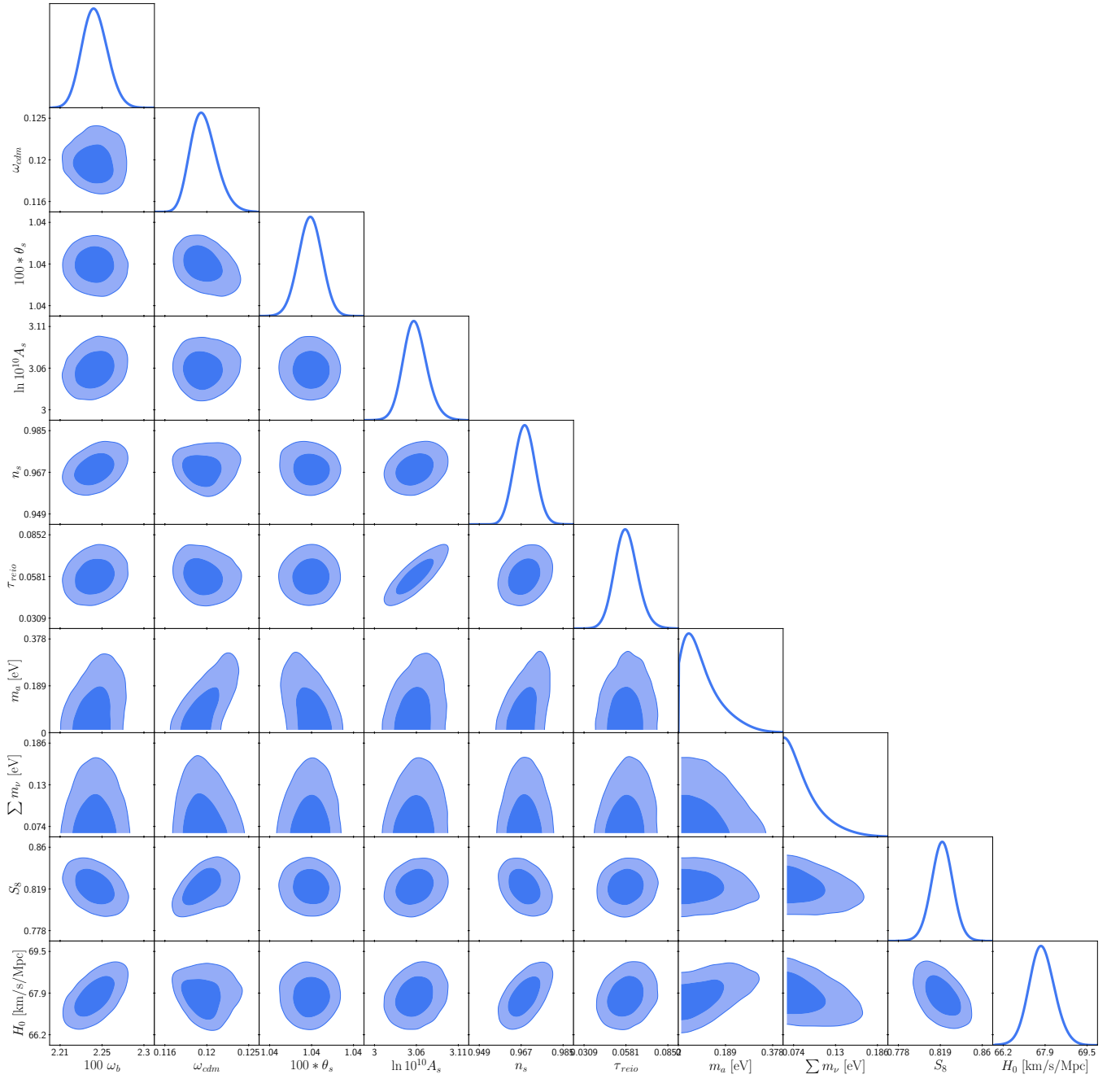
Figure 7. 1d and 2d posterior distributions for the $\Lambda\text{CDM}+\sum m_\nu+m_a$ model, extracted from the joint Planck18 TT+TE+EE+low ℓ+lensing+BAO+Pantheon likelihood.

[1] R. D. Peccei and H. R. Quinn, CP Conservation in the Presence of Instantons, Phys. Rev. Lett. **38**, 1440 (1977).
[2] R. D. Peccei and H. R. Quinn, Constraints Imposed by CP Conservation in the Presence of Instantons, Phys. Rev. D **16**, 1791 (1977).
[3] S. Weinberg, A New Light Boson?, Phys. Rev. Lett. **40**, 223 (1978).
[4] F. Wilczek, Problem of Strong P and T Invariance in the Presence of Instantons, Phys. Rev. Lett. **40**, 279 (1978).
[5] M. S. Turner, Thermal Production of Not SO Invisible Axions in the Early Universe, Phys. Rev. Lett. **59**, 2489 (1987), [Erratum: Phys.Rev.Lett. 60, 1101 (1988)].
[6] E. W. Kolb and M. S. Turner, The Early Universe, Vol. 69 (1990).

[7] Z. G. Berezhiani, A. S. Sakharov, and M. Y. Khlopov, Primordial background of cosmological axions, Sov. J. Nucl. Phys. **55**, 1063 (1992).

[8] S. Chang and K. Choi, Hadronic axion window and the big bang nucleosynthesis, Phys. Lett. B **316**, 51 (1993), arXiv:hep-ph/9306216.

[9] E. Masso, F. Rota, and G. Zsembinszki, On axion thermalization in the early universe, Phys. Rev. D **66**, 023004 (2002), arXiv:hep-ph/0203221.

[10] L. F. Abbott and P. Sikivie, A Cosmological Bound on the Invisible Axion, Phys. Lett. B **120**, 133 (1983).

[11] M. Dine and W. Fischler, The Not So Harmless Axion, Phys. Lett. B **120**, 137 (1983).

[12] J. Preskill, M. B. Wise, and F. Wilczek, Cosmology of the Invisible Axion, Phys. Lett. B **120**, 127 (1983).

[13] R. L. Davis, Cosmic Axions from Cosmic Strings, Phys. Lett. B **180**, 225 (1986).

[14] D. Harari and P. Sikivie, On the Evolution of Global Strings in the Early Universe, Phys. Lett. B **195**, 361 (1987).

[15] D. H. Lyth, Estimates of the cosmological axion density, Phys. Lett. B **275**, 279 (1992).

[16] N. Aghanim et al. (Planck), Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. **641**, A6 (2020), [Erratum: Astron.Astrophys. 652, C4 (2021)], arXiv:1807.06209 [astro-ph.CO].

[17] S. Aiola et al. (ACT), The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters, JCAP **12**, 047, arXiv:2007.07288 [astro-ph.CO].

[18] J. A. Sobrin et al. (SPT-3G), The Design and Integrated Performance of SPT-3G, Astrophys. J. Supp. **258**, 42 (2022), arXiv:2106.11202 [astro-ph.IM].

[19] A. Aghamousa et al. (DESI), The DESI Experiment Part I: Science,Targeting, and Survey Design, (2016), arXiv:1611.00036 [astro-ph.IM].

[20] P. Ade et al. (Simons Observatory), The Simons Observatory: Science goals and forecasts, JCAP **02**, 056, arXiv:1808.07445 [astro-ph.CO].

[21] K. Abazajian et al. (CMB-S4), Snowmass 2021 CMB-S4 White Paper, (2022), arXiv:2203.08024 [astro-ph.CO].

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

[23] v. Ivezić et al. (LSST), LSST: from Science Drivers to Reference Design and Anticipated Data Products, Astrophys. J. **873**, 111 (2019), arXiv:0805.2366 [astro-ph].

[24] G. G. Raffelt, Astrophysical methods to constrain axions and other novel particle phenomena, Phys. Rept. **198**, 1 (1990).

[25] J. H. Chang, R. Essig, and S. D. McDermott, Supernova 1987A Constraints on Sub-GeV Dark Sectors, Millicharged Particles, the QCD Axion, and an Axion-like Particle, JHEP **09**, 051, arXiv:1803.00993 [hep-ph].

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

[27] P. Carenza, B. Fore, M. Giannotti, A. Mirizzi, and S. Reddy, Enhanced Supernova Axion Emission and its Implications, Phys. Rev. Lett. **126**, 071102 (2021), arXiv:2010.02943 [hep-ph].

[28] 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**, 069, arXiv:1406.7712 [hep-ph].

[29] 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**, 191302 (2014), arXiv:1406.6053 [astro-ph.SR].

[30] M. J. Dolan, F. J. Hiskens, and R. R. Volkas, Advancing Globular Cluster Constraints on the Axion-Photon Coupling, (2022), arXiv:2207.03102 [hep-ph].

[31] M. Buschmann, C. Dessert, J. W. Foster, A. J. Long, and B. R. Safdi, Upper Limit on the QCD Axion Mass from Isolated Neutron Star Cooling, Phys. Rev. Lett. **128**, 091102 (2022), arXiv:2111.09892 [hep-ph].

[32] E. Armengaud et al. (IAXO), Physics potential of the International Axion Observatory (IAXO), JCAP **06**, 047, arXiv:1904.09155 [hep-ph].

[33] A. Abeln et al. (IAXO), Conceptual design of BabyIAXO, the intermediate stage towards the International Axion Observatory, JHEP **05**, 137, arXiv:2010.12076 [physics.ins-det].

[34] S. Hannestad and J. Madsen, Neutrino decoupling in the early universe, Phys. Rev. D **52**, 1764 (1995), arXiv:astro-ph/9506015.

[35] A. D. Dolgov, S. H. Hansen, and D. V. Semikoz, Nonequilibrium corrections to the spectra of massless neutrinos in the early universe, Nucl. Phys. B **503**, 426 (1997), arXiv:hep-ph/9703315.

[36] D. I. Dunsky, L. J. Hall, and K. Harigaya, Dark Radiation Constraints on Heavy QCD Axions, (2022), arXiv:2205.11540 [hep-ph].

[37] S. Hannestad, A. Mirizzi, and G. Raffelt, New cosmological mass limit on thermal relic axions, JCAP **07**, 002, arXiv:hep-ph/0504059.

[38] A. Melchiorri, O. Mena, and A. Slosar, An improved cosmological bound on the thermal axion mass, Phys. Rev. D **76**, 041303 (2007), arXiv:0705.2695 [astro-ph].

[39] S. Hannestad, A. Mirizzi, G. G. Raffelt, and Y. Y. Y. Wong, Cosmological constraints on neutrino plus axion hot dark matter, JCAP **08**, 015, arXiv:0706.4198 [astro-ph].

[40] E. Di Valentino, E. Giusarma, M. Lattanzi, O. Mena, A. Melchiorri, and J. Silk, Cosmological Axion and neutrino mass constraints from Planck 2015 temperature and polarization data, Phys. Lett. B **752**, 182 (2016), arXiv:1507.08665 [astro-ph.CO].

[41] L. Di Luzio, G. Martinelli, and G. Piazza, Breakdown of chiral perturbation theory for the axion hot dark matter bound, Phys. Rev. Lett. **126**, 241801 (2021), arXiv:2101.10330 [hep-ph].

[42] M. Gorghetto and G. Villadoro, Topological Susceptibility and QCD Axion Mass: QED and NNLO corrections, JHEP **03**,

033, arXiv:1812.01008 [hep-ph].

[43] M. L. Bellac, Thermal Field Theory, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2011).

[44] J. Froustey, C. Pitrou, and M. C. Volpe, Neutrino decoupling including flavour oscillations and primordial nucleosynthesis, JCAP 12, 015, arXiv:2008.01074 [hep-ph].

[45] J. J. Bennett, G. Buldgen, P. F. De Salas, M. Drewes, S. Gariazzo, S. Pastor, and Y. Y. Y. Wong, Towards a precision calculation of N_{eff} in the Standard Model II: Neutrino decoupling in the presence of flavour oscillations and finite-temperature QED, JCAP 04, 073, arXiv:2012.02726 [hep-ph].

[46] J. L. Goity and H. Leutwyler, On the Mean Free Path of Pions in Hot Matter, Phys. Lett. B 228, 517 (1989).

[47] H. Bebie, P. Gerber, J. L. Goity, and H. Leutwyler, The Role of the entropy in an expanding hadronic gas, Nucl. Phys. B 378, 95 (1992).

[48] A. Schenk, Pion propagation at finite temperature, Phys. Rev. D 47, 5138 (1993).

[49] R. L. Workman et al. (Particle Data Group), Review of Particle Physics, PTEP 2022, 083C01 (2022).

[50] J. Gasser and H. Leutwyler, Chiral Perturbation Theory to One Loop, Annals Phys. 158, 142 (1984).

[51] J. R. Pelaez and F. J. Yndurain, The Pion-pion scattering amplitude, Phys. Rev. D 71, 074016 (2005), arXiv:hep-ph/0411334.

[52] R. Garcia-Martin, R. Kaminski, J. R. Pelaez, J. Ruiz de Elvira, and F. J. Yndurain, The Pion-pion scattering amplitude. IV: Improved analysis with once subtracted Roy-like equations up to 1100 MeV, Phys. Rev. D 83, 074004 (2011), arXiv:1102.2183 [hep-ph].

[53] J. E. Kim, Weak Interaction Singlet and Strong CP Invariance, Phys. Rev. Lett. 43, 103 (1979).

[54] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Can Confinement Ensure Natural CP Invariance of Strong Interactions?, Nucl. Phys. B 166, 493 (1980).

[55] A. R. Zhitnitsky, On Possible Suppression of the Axion Hadron Interactions. (In Russian), Sov. J. Nucl. Phys. 31, 260 (1980).

[56] M. Dine, W. Fischler, and M. Srednicki, A Simple Solution to the Strong CP Problem with a Harmless Axion, Phys. Lett. B 104, 199 (1981).

[57] D. Baumann, D. Green, and B. Wallisch, New Target for Cosmic Axion Searches, Phys. Rev. Lett. 117, 171301 (2016), arXiv:1604.08614 [astro-ph.CO].

[58] R. Z. Ferreira and A. Notari, Observable Windows for the QCD Axion Through the Number of Relativistic Species, Phys. Rev. Lett. 120, 191301 (2018), arXiv:1801.06090 [hep-ph].

[59] F. D'Eramo, R. Z. Ferreira, A. Notari, and J. L. Bernal, Hot Axions and the H_0 tension, JCAP 11, 014, arXiv:1808.07430 [hep-ph].

[60] F. Arias-Aragón, F. D'Eramo, R. Z. Ferreira, L. Merlo, and A. Notari, Production of Thermal Axions across the ElectroWeak Phase Transition, JCAP 03, 090, arXiv:2012.04736 [hep-ph].

[61] D. Green, Y. Guo, and B. Wallisch, Cosmological implications of axion-matter couplings, JCAP 02 (02), 019, arXiv:2109.12088 [astro-ph.CO].

[62] R. Z. Ferreira, A. Notari, and F. Rompineve, Dine-Fischler-Srednicki-Zhitnitsky axion in the CMB, Phys. Rev. D 103, 063524 (2021), arXiv:2012.06566 [hep-ph].

[63] A. D. Linde, Infrared Problem in Thermodynamics of the Yang-Mills Gas, Phys. Lett. B 96, 289 (1980).

[64] D. J. Gross, R. D. Pisarski, and L. G. Yaffe, QCD and Instantons at Finite Temperature, Rev. Mod. Phys. 53, 43 (1981).

[65] P. Graf and F. D. Steffen, Thermal axion production in the primordial quark-gluon plasma, Phys. Rev. D 83, 075011 (2011), arXiv:1008.4528 [hep-ph].

[66] A. Salvio, A. Strumia, and W. Xue, Thermal axion production, JCAP 01, 011, arXiv:1310.6982 [hep-ph].

[67] V. S. Rychkov and A. Strumia, Thermal production of gravitinos, Phys. Rev. D 75, 075011 (2007), arXiv:hep-ph/0701104.

[68] F. D'Eramo, F. Hajkarim, and S. Yun, Thermal QCD Axions across Thresholds, JHEP 10, 224, arXiv:2108.05371 [hep-ph].

[69] L. D. McLerran, E. Mottola, and M. E. Shaposhnikov, Sphalerons and Axion Dynamics in High Temperature QCD, Phys. Rev. D 43, 2027 (1991).

[70] D. Bodeker, On the effective dynamics of soft nonAbelian gauge fields at finite temperature, Phys. Lett. B 426, 351 (1998), arXiv:hep-ph/9801430.

[71] P. B. Arnold, D. T. Son, and L. G. Yaffe, Effective dynamics of hot, soft nonAbelian gauge fields. Color conductivity and log(1/alpha) effects, Phys. Rev. D 59, 105020 (1999), arXiv:hep-ph/9810216.

[72] G. D. Moore and M. Tassler, The Sphaleron Rate in SU(N) Gauge Theory, JHEP 02, 105, arXiv:1011.1167 [hep-ph].

[73] S. Borsanyi et al., Calculation of the axion mass based on high-temperature lattice quantum chromodynamics, Nature 539, 69 (2016), arXiv:1606.07494 [hep-lat].

[74] L. Altenkort, A. M. Eller, O. Kaczmarek, L. Mazur, G. D. Moore, and H.-T. Shu, Sphaleron rate from Euclidean lattice correlators: An exploration, Phys. Rev. D 103, 114513 (2021), arXiv:2012.08279 [hep-lat].

[75] M. B. Mancha and G. D. Moore, The Sphaleron Rate from 4D Euclidean Lattices, (2022), arXiv:2210.05507 [hep-lat].

[76] K. V. Berghaus, P. W. Graham, D. E. Kaplan, G. D. Moore, and S. Rajendran, Dark energy radiation, Phys. Rev. D 104, 083520 (2021), arXiv:2012.10549 [hep-ph].

[77] N. Aghanim et al. (Planck), Planck 2018 results. V. CMB power spectra and likelihoods, Astron. Astrophys. 641, A5 (2020), arXiv:1907.12875 [astro-ph.CO].

[78] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant, Mon. Not. Roy. Astron. Soc. 416, 3017 (2011), arXiv:1106.3366 [astro-ph.CO].

[79] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, The clustering of the SDSS DR7 main

Galaxy sample – I. A 4 per cent distance measure at $z = 0.15$, Mon. Not. Roy. Astron. Soc. **449**, 835 (2015), arXiv:1409.3242 [astro-ph.CO].

[80] S. Alam et al. (BOSS), The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample, Mon. Not. Roy. Astron. Soc. **470**, 2617 (2017), arXiv:1607.03155 [astro-ph.CO].

[81] S. Satpathy et al. (BOSS), The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: On the measurement of growth rate using galaxy correlation functions, Mon. Not. Roy. Astron. Soc. **469**, 1369 (2017), arXiv:1607.03148 [astro-ph.CO].

[82] D. M. Scolnic et al. (Pan-STARRS1), The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample, Astrophys. J. **859**, 101 (2018), arXiv:1710.00845 [astro-ph.CO].

[83] J. Lesgourgues, The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview, (2011), arXiv:1104.2932 [astro-ph.IM].

[84] D. Blas, J. Lesgourgues, and T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes, JCAP **07**, 034, arXiv:1104.2933 [astro-ph.CO].

[85] B. Audren, J. Lesgourgues, K. Benabed, and S. Prunet, Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code, JCAP **02**, 001, arXiv:1210.7183 [astro-ph.CO].

[86] T. Brinckmann and J. Lesgourgues, MontePython 3: boosted MCMC sampler and other features, Phys. Dark Univ. **24**, 100260 (2019), arXiv:1804.07261 [astro-ph.CO].

[87] F. D'Eramo, E. Di Valentino, W. Giarè, F. Hajkarim, A. Melchiorri, O. Mena, F. Renzi, and S. Yun, Cosmological bound on the QCD axion mass, redux, JCAP **09**, 022, arXiv:2205.07849 [astro-ph.CO].

[88] M. Laine and A. Vuorinen, Basics of Thermal Field Theory, Vol. 925 (Springer, 2016) arXiv:1701.01554 [hep-ph].

[89] G. P. Lepage, Adaptive multidimensional integration: VEGAS enhanced, J. Comput. Phys. **439**, 110386 (2021), arXiv:2009.05112 [physics.comp-ph].

[90] W. R. Inc., Mathematica, Version 13.1, champaign, IL, 2022.