

New bounds on heavy QCD axions from big bang nucleosynthesis

Tae Hyun Jung^{1,*}, Takemichi Okui^{2,3,†}, Kohsaku Tobioka^{2,3,‡} and Jiabao Wang^{2,§}

¹*Particle Theory and Cosmology Group, Center for Theoretical Physics of the Universe,
Institute for Basic Science (IBS), Daejeon, 34126, Korea*

²*Department of Physics, Florida State University, Tallahassee, Florida 32306, USA*

³*High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801, Japan*



(Received 3 November 2025; accepted 12 January 2026; published 2 March 2026)

We study big bang nucleosynthesis (BBN) constraints on heavy QCD axions. BBN offers a powerful probe of new physics that modifies the neutron-to-proton ratio during the process, thanks to the precisely measured primordial Helium-4 abundance. A heavy QCD axion provides an attractive target for this probe, because not only is it a well-motivated hypothetical particle by the strong CP problem, but also it dominantly decays to hadrons if kinematically allowed. A range of its lifetime is thus excluded where the hadronic decays would significantly alter the neutron-to-proton ratio. We compute axion-induced modification of the neutron-to-proton ratio, and obtain robust upper bounds on the axion lifetimes, as low as 0.017 s for the axion mass higher than 300 MeV. Remarkably, this is stronger than projected future cosmic microwave background bounds via N_{eff} . Our bounds are largely insensitive to uncertainties in hadronic cross sections and the axion's branching fractions into various hadrons, as well as to the precise value of the initial axion abundance. We also incorporate, for the first time, several key improvements, such as scattering processes by energetic K_L and secondary hadrons, that can also be important for studying general hadronic injections during BBN, not limited to those from axion decays.

DOI: [10.1103/2m1-h1cp](https://doi.org/10.1103/2m1-h1cp)

I. INTRODUCTION

Big bang nucleosynthesis (BBN) offers a powerful probe of new physics that modifies the neutron-to-proton ratio during the process, thanks to the well-understood underlying standard model (SM) processes and precisely measured primordial ^4He abundance. The BBN probe is especially sensitive to new physics that injects extra hadrons into the plasma [1] (with a first comprehensive quantitative analysis by [2]). These extra hadrons modify the neutron-to-proton ratio via the strong interactions (e.g., $\pi^- p \rightarrow \pi^0 n$), whose cross sections are roughly 16 orders of magnitude larger than the weak interaction processes that govern the ratio in the standard BBN (e.g., $e^- p \rightarrow \nu_e n$) around the freeze-out temperature of the neutron-to-proton ratio ($T \sim 1$ MeV). (Roughly speaking, the strong cross sections are on the order of mb $\sim \text{GeV}^{-2}$, while the weak cross sections are on the

order of $G_F^2 \text{MeV}^2 \sim 10^{-16} \text{GeV}^{-2}$.) Therefore, even an exponentially suppressed amount of hadronic injection can still significantly impact the neutron-to-proton ratio. An example of such hadronic injection is a hadronically decaying new particle whose lifetime is much shorter than the freeze-out time of the neutron-to-proton ratio (~ 1 s). Indeed, as we will find in this paper, BBN can be sensitive to lifetimes of $\mathcal{O}(0.01)$ s for heavy QCD axions. BBN constraints on hadronic injection have also been studied for a variety of new physics scenarios, such as primordial black holes [3], the gravitino [4–16], Higgs-portal scalars [17], dark photons [18,19], heavy neutral leptons [20,21], dark matter annihilation [22–24], and as an attempt to solve the lithium problem [25].

A heavy QCD axion provides an especially attractive target for the BBN probe, which we study for the first time. Firstly, it is a well-motivated hypothetical particle that can solve the strong CP problem without suffering from the so-called quality problem. Secondly, it dominantly decays to hadrons if kinematically allowed. This is because a QCD axion—heavy or not—by definition dominantly couples to the SM sector via the coupling

$$\mathcal{L} \supset \frac{\alpha_s}{8\pi} \frac{a}{f_a} G\tilde{G}, \quad (1.1)$$

where a and G denote the axion field and gluon field strength, respectively, and f_a is a mass scale called the

*Contact author: thjung0720@gmail.com

†Contact author: tokui@fsu.edu

‡Contact author: ktobioka@fsu.edu

§Contact author: jwang21@fsu.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

axion decay constant. Hence, if its mass m_a is above ~ 300 MeV to kinematically allow a hadronic decay channel, the axion will decay dominantly to hadrons via the coupling Eq. (1.1). Since this coupling is what makes a solve the strong CP problem via the Peccei-Quinn (PQ) mechanism [26–33], the dominance of hadronic decays is well motivated.

For a further motivation for considering *heavy* QCD axions in particular, let us pay attention to the “quality” of $U(1)_{\text{PQ}}$, an approximate global symmetry under which a shifts. The PQ mechanism requires that the location of the minimum of the axion potential be dictated by the $U(1)_{\text{PQ}}$ breaking due to Eq. (1.1) (the condition for “good quality”). Since quantum gravity effects might completely violate global symmetries at the Planck scale ($M_{\text{Pl}} \simeq 1.2 \times 10^{19}$ GeV), the question arises as to whether the good-quality condition is well respected by $U(1)_{\text{PQ}}$ -violating operators suppressed by M_{Pl} (the quality problem). It is straightforward to see that a high-quality $U(1)_{\text{PQ}}$ prefers a high value of m_a for any given f_a . To see the idea, let Φ be the $U(1)_{\text{PQ}}$ -breaking scalar field with $\langle \Phi \rangle = f_a$. A $U(1)_{\text{PQ}}$ -violating operator of the form $\Phi^n / M_{\text{Pl}}^{n-4}$ with $n > 4$ with a generic $\mathcal{O}(1)$ complex coefficient would not observably shift the location of the minimum if $f_a^n / M_{\text{Pl}}^{n-4} \lesssim \theta_{\text{max}} f_a^2 m_a^2$, where $\theta_{\text{max}} \sim 10^{-10}$ is the current upper bound on the QCD vacuum angle. This shows that the PQ quality is maintained for a sufficiently high m_a for any given f_a .

To motivate the region of the parameter space we will explore in this work, we rewrite the above condition as

$$f_a \lesssim \theta_{\text{max}}^{\frac{1}{n-2}} M_{\text{Pl}} \left(\frac{m_a}{M_{\text{Pl}}} \right)^{\frac{2}{n-2}} \sim \begin{cases} 10^7 \text{ GeV} \left(\frac{m_a}{\text{GeV}} \right)^{\frac{1}{2}} & \text{if } n = 6, \\ 10^{11} \text{ GeV} \left(\frac{m_a}{\text{GeV}} \right)^{\frac{1}{3}} & \text{if } n = 8. \end{cases} \quad (1.2)$$

Such a region of the parameter space is not compatible with the standard relation $m_a f_a \simeq m_\pi f_\pi$ unless we take f_a to be so low that it is already excluded by existing axion searches [34]. Since the standard relation assumes that the dominant $U(1)_{\text{PQ}}$ breaking is given by the coupling of Eq. (1.1), a heavy QCD axion requires an additional source of $U(1)_{\text{PQ}}$ breaking. This must not reintroduce the quality problem, but, unlike the original quality problem, this time it is unrelated to quantum gravity and hence can be firmly addressed by model building within conventional quantum field theory. Three types of models are known depending on whether additional contributions to the axion mass originate from (i) an additional confinement gauge group unified with $SU(3)_c$ into a larger gauge group in the UV [35–39], (ii) the $SU(3)_c$ instanton contribution with somehow large strong gauge coupling [35,40–46], or (iii) a mirror QCD [47–50]. Our bounds are independent of the origin of the additional axion mass as long as the axion dominantly decays to hadrons. We

assume that the additional mass is generated at a sufficiently high energy scale that we can treat m_a as constant during the cosmological evolution relevant to our analysis.

As we will find in this paper, BBN constrains the heavy QCD axion lifetime to be $\lesssim 0.02$ s. This bound occupies a unique place in comparison with other experimental constraints when viewed as bounds on the heavy QCD axion lifetime. Collider experiments can probe prompt and displaced decays, where the longest lifetimes they can probe are set by the detector geometries to be around $\sim 10^{-9}$ s. For the mass range of our interest, $m_a \gtrsim 300$ MeV,¹ competitive bounds in this range of lifetimes are from the LHC [53–59] and B -factories [60–63]. Proton beam dump experiments and similar setups can probe much longer lifetimes up to around $\sim 10^{-6}$ s [64–68] due to much larger distances to the detector. BBN probes even longer lifetimes, ranging from about 10^{-2} s to a few minutes. The cosmic microwave background (CMB) can also probe lifetimes longer than about 0.1 s via N_{eff} [69], which is especially relevant when the hadronic decays are kinematically forbidden. As we will show, once a phase space for hadronic decays opens, our bounds from the BBN neutron-to-proton ratio become stronger than the CMB’s and constrain lifetimes of $\gtrsim 0.017$ s.

Since a well-motivated axion decays to hadrons if kinematically allowed, it is essential to include hadrons in studies of BBN bounds. Consequently, our bound is an order of magnitude stronger than the ones for the axionlike particles that assume diphoton decay to be dominant even at high masses [70–73].

II. OVERVIEW

Before discussing technical details, we would like to provide an overview of our study and results. Readers mainly interested in the results may proceed directly to Sec. VII after reading this section. In addition, we also highlight key improvements we made in the methodology of studying BBN constraints on general hadronically decaying long-lived particles, not necessarily limited to heavy QCD axions.

A. Overall framework

We begin by imagining a postinflation reheating temperature that is sufficiently high to allow axions to be in thermal and chemical equilibrium with the SM particles. As the temperature drops, but while still relativistic, the axions decouple from the rest, and their number density freezes out. This sets the initial condition for our BBN analysis, with the initial axion abundance given by $Y_a = n_a/s \sim 1/g_*^{\text{FO}}$, where g_*^{FO} is the effective number of relativistic degrees of freedom

¹For lower masses, bounds from kaon decays become competitive. See the recent review [34] for a compilation of various bounds and a list of relevant references. There are also bounds from supernova for lower masses and intermediate values of the decay constant $f_a \lesssim 10^8$ GeV [51,52].

at the axion freeze-out temperature T_{FO} . In [69], T_{FO} and Y_a are estimated for different values of m_a and f_a across the QCD deconfinement-confinement transition. We basically follow their method of estimation, which will be discussed in detail in Sec. III.

In Sec. IV the axion decay width and branching fractions will be calculated in two ways depending on whether the quark-gluon or hadronic description should be used. For $m_a < 2$ GeV, where the hadronic picture is appropriate, a data-driven method [60, 74–77] with the data in [75] will be used to capture nonperturbative physics necessary for our purpose, in particular the branching fractions of the axion decay into various hadrons. For $m_a > 2$ GeV, perturbative QCD (pQCD) will be employed to calculate the axion decay rate at next-to-next-to-leading order (NNLO) following [78] (with more technical details in Appendix A), and the branching fractions are computed by the hadronic shower programs (PYTHIA 8.306 [79] and Herwig v7.3.0 [80–82]). We find large discrepancies in the predictions of these two programs, especially when m_a is close to 2 GeV. Neither program can be trusted there, because they both fail to respect basic properties, such as the parity of the axion. While this indicates a serious need for improvement in these programs, for our purpose, we fortunately find that those order-one uncertainties in the axion branching fractions do not affect our final bounds on the axion lifetime beyond a few percent.

We aim to obtain an upper bound of $\mathcal{O}(0.01)$ s for the axion lifetime, where the relevant experimental observable being probed is the primordial ^4He abundance. In the standard BBN (see Ref. [83] for a review), the process starts with the decoupling of neutrinos at $T \sim 2$ MeV ($t \sim 0.2$ s), and the freeze-out of neutron-proton conversion around $T_n \sim 1$ MeV ($t_n \sim 0.7$ s), where the neutron fraction is defined as $X_n \equiv n_n/n_b$, with n_n and n_b being the number densities of neutrons and baryons, respectively. Later, once T falls below the so-called deuterium bottleneck temperature $T_D \sim 70$ keV ($t_D \sim 200$ s), these neutrons (after slight reduction in number due to beta decay) are all consumed to produce deuterium and then tritium, but at the end nearly all of the neutrons end up in ^4He . Thus, the primordial ^4He abundance is determined by X_n at T_D .

Conventionally, the ^4He abundance is given as its mass fraction to the total baryon energy density: $Y_p \equiv \frac{\rho(^4\text{He})}{\rho_b}$. We adopt the PDG-recommended value [84]

$$Y_p = 0.245 \pm 0.003, \quad (2.1)$$

which is based on [85–91].² As nearly all the neutrons become ^4He below T_D , it is a very good approximation to

²EMPRESS collaboration reports $Y_p = 0.2370^{+0.0034}_{-0.0033}$ [92], whose central value is about 2σ smaller than the one above, but fortunately, their uncertainties are the same. Since our constraints will be based on a fractional change $\delta Y_p/Y_p$ [or see Eq. (2.2)], this difference in the central values does not matter.

take $Y_p \simeq 2X_n(T_D)$ and just focus on calculating how the heavy QCD axion changes $X_n(T_D)$. We thus obtain a 2σ bound by

$$\left. \frac{\delta X_n}{X_n^{\text{SM}}} \right|_{T=T_D} < R_{Y_p} \equiv 2.45\%, \quad (2.2)$$

with $\delta X_n \equiv X_n - X_n^{\text{SM}}$, and we define T_D by $Y_p = 2X_n^{\text{SM}}(T_D)$, where Y_p is given in Eq. (2.1) and we calculate $X_n^{\text{SM}}(T)$ by removing the axion from our Boltzmann equations, although our constraints from Eq. (2.2) will be insensitive to the precise definition of T_D . We would also like to point out that the standard BBN predicts the central value of $Y_p = 0.247$, slightly higher than the observed central value of Eq. (2.1), so our condition Eq. (2.2) with X_n^{SM} matched to Eq. (2.1) provides a conservative bound because it turns out that $\delta X_n > 0$.

Next, we would like to summarize what hadrons and reactions are relevant to $n \leftrightarrow p$ conversion. As we will explain shortly, the lifetime of a hadron dictates whether it should be included in the analysis. In Table I, we list all hadrons and reactions included in our analysis. Relevant hadrons are limited to $\pi^\pm$, $K^\pm$, K_L , nucleons $N = p, n$, and antinucleons $\bar{N} = \bar{p}, \bar{n}$. Other hadrons, such as hyperons, decay too rapidly to participate in $n \leftrightarrow p$ conversion, even though their cross sections may be larger.

To identify the hadrons important for $n \leftrightarrow p$ conversion, consider the disappearance terms in the Boltzmann equation for an injected hadron h ,

$$\dot{n}_h \supset -(3H + \Gamma_h + \langle \sigma v \rangle_{\text{dis}} n_b) n_h, \quad (2.3)$$

where H is the Hubble expansion rate, Γ_h is the hadron's decay width, and $\langle \sigma v \rangle_{\text{dis}}$ is the sum of the thermally averaged cross sections for all baryon- h scattering processes that consume the h , which in particular include $n \leftrightarrow p$ conversion processes. First, consider an h other than N or $\bar{N}$. On the right-hand side, since H is $\mathcal{O}(1) \text{ s}^{-1}$, the $3H$ term is always negligible in comparison with the two other terms. Therefore, in order for the h to participate in $n \leftrightarrow p$ conversion before it decays, the Γ_h term must be smaller than or comparable to the scattering rate,

$$\Gamma_h \lesssim \langle \sigma v \rangle_{\text{dis}} n_b(T) \sim 10^{-17} \text{ GeV} \frac{\langle \sigma v \rangle_{\text{dis}}}{10 \text{ mb}} \left(\frac{T}{\text{MeV}} \right)^3. \quad (2.4)$$

Among mesons, this condition may be satisfied only by the following three:

$$\begin{aligned} \pi^\pm : \tau_{\pi^\pm} &= 2.603 \times 10^{-8} \text{ s}, & \Gamma_{\pi^\pm} &= 2.53 \times 10^{-17} \text{ GeV}, \\ K^\pm : \tau_{K^\pm} &= 1.238 \times 10^{-8} \text{ s}, & \Gamma_{K^\pm} &= 5.32 \times 10^{-17} \text{ GeV}, \\ K_L : \tau_{K_L} &= 5.116 \times 10^{-8} \text{ s}, & \Gamma_{K_L} &= 1.29 \times 10^{-17} \text{ GeV}. \end{aligned}$$

TABLE I. Summary of all $n \leftrightarrow p$ conversion processes considered in our analysis. Y denotes a hyperon: $\Sigma^\pm$, Σ^0 , or Λ . Q is the total mass of the initial state minus that of the final state. Processes with negative Q are kinematically forbidden unless the initial state has sufficient kinetic energy. We keep track of the time evolution of n , p , and all hadrons listed in the leftmost column. The $\ddagger$ symbol on a secondary hadron indicates that its contribution is neglected in the Boltzmann equation for that hadron, for the reasons explained in Sec. II B.

$n \rightarrow p$	$p \rightarrow n$	Q/MeV	Secondary	σ
π^-	$p\pi^- \rightarrow n\pi^0$	3.30		Fig. 21
	$p\pi^- \rightarrow n\gamma$	138		Fig. 23
π^+	$n\pi^+ \rightarrow p\pi^0$	5.89		Fig. 22
	$n\pi^+ \rightarrow p\gamma$	141		Fig. 24
K^-	$pK^- \rightarrow \pi Y (Y \rightarrow n)$	$\gtrsim 100$	$\pi^\pm$	Fig. 25
	$pK^- \rightarrow n\bar{K}^0$	-5.23	$K_S^\ddagger, K_L^\ddagger$	Fig. 29
	$nK^- \rightarrow \pi Y (Y \rightarrow p)$	$\gtrsim 100$	π^-	Fig. 28
K^+	$nK^+ \rightarrow pK^0$	-2.64	$K_S^\ddagger, K_L^\ddagger$	Fig. 30
K_L	$pK_L \rightarrow \pi Y (Y \rightarrow n)$	$\gtrsim 100$	$\pi^\pm$	Figs. 27, 34
	$pK_L \rightarrow nK^+$	2.64	K^+	Figs. 30, 34
	$nK_L \rightarrow \pi Y (Y \rightarrow p)$	$\gtrsim 100$	$\pi^\pm$	Figs. 26, 33
	$nK_L \rightarrow pK^-$	5.23	K^-	Figs. 29, 33
	$pK_L \rightarrow pK_S$	0	$K_S(\rightarrow\pi^\pm)$	Fig. 31
	$nK_L \rightarrow nK_S$	0	$K_S(\rightarrow\pi^\pm)$	Fig. 31
$\bar{n}$	$p\bar{n} \rightarrow \text{mesons}$	1600	mesons ‡	Fig. 35
$\bar{p}$	$n\bar{p} \rightarrow \text{mesons}$	1600	mesons ‡	Fig. 35

Among baryons, all unflavored heavy baryons (e.g., Δ) are clearly too short-lived to satisfy the above condition. Hyperons, though much longer-lived, also fail to satisfy the condition.³ Finally, needless to say, nucleons and antinucleons (p , n , $\bar{p}$, and $\bar{n}$) are important in $n \leftrightarrow p$ conversion. Here, once produced from axion decays, the antinucleons will immediately annihilate with existing nucleons, e.g., $n\bar{p} \rightarrow \pi^-\pi^0 + \dots$ and $p\bar{n} \rightarrow \pi^+\pi^0 + \dots$.

Let us comment on the impact of the axions on the background cosmology. The energy density in the axions alters the evolution of the Hubble rate. For axion masses greater than a few GeV, there is even a short period of a matter-dominated era by nonrelativistic axions, which ends as the axions decay into the radiation component, restarting the radiation-dominated era. This alters the evolution of the Hubble rate, which in turn shifts the freeze-out temperature of the neutron-to-proton ratio. Our Boltzmann equations account for this effect.

Another effect on the background cosmology is that axion decays after neutrino decoupling effectively dilute the neutrino population relative to the plasma, as the latter heats up by the axion decay products. As discussed in [69], this changes the relative evolution of the neutrino temperature and results in a value of N_{eff} smaller than the SM

³The decay widths of hyperons are on the order of 10^{-15} GeV, while their scattering cross sections are at most ~ 100 mb, thereby failing to meet the condition.

value. We find that our BBN bounds are stronger than projected future CMB bounds via N_{eff} .

Heating up the plasma also dilutes the baryon-to-photon ratio, $\eta_b \equiv n_b/n_\gamma$, whose value is tightly constrained by CMB in the late Universe. Therefore, we fix the value of η_b after the dilution to be $\eta_b = 6.115 \times 10^{-10}$, as measured by Planck [93]. Once the background cosmology and n_b evolution are obtained, we solve Boltzmann equations for the hadrons and examine if δX_n is compatible with the ^4He measurement.

In Fig. 1, we schematically depict the thermal history of the Universe as well as the steps taken in our analysis described in this subsection.

B. Improvements in the treatment of hadrons

We also introduce important improvements in the quantitative treatment of the reactions listed in Table I. The following is a brief summary of the improvements:

- (i) *Revisit and update hadronic cross sections.* We improve the treatment of hadronic cross sections pioneered by [1,25]. In particular, cross sections involving $K^\pm$ and K_L are significantly improved. Within a partial wave analysis, we take a complete set of isospin-related 2-to-2 processes and, if necessary, also include p -waves and/or momentum-dependent scattering lengths. As elaborated in the next bullet, cross sections at high momenta are important for

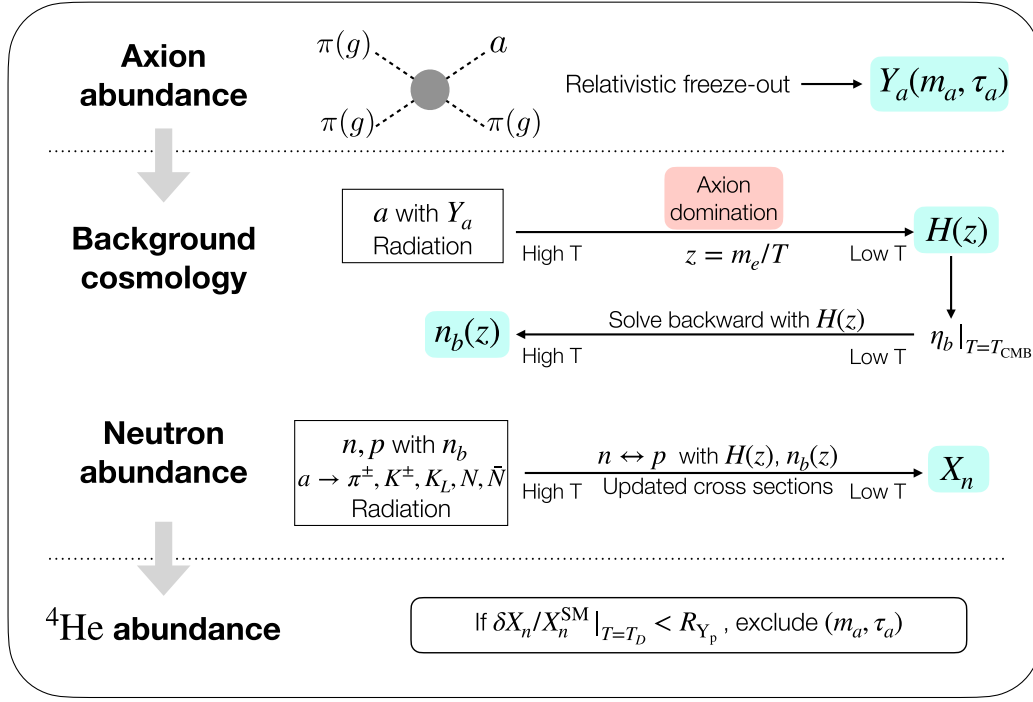


FIG. 1. Schematic depiction of the thermal history of the Universe as well as steps taken in our analysis, as summarized in Sec. II A. We adopt the 2σ bound of ^4He abundance $R_{Y_p} = 2.45\%$.

scattering of K_L . See Sec. VI for a more concrete summary and Appendix C for detailed derivations.

- (ii) *Include energetic K_L properly.* As pointed out in Refs. [1,2], being electrically neutral, the K_L s do not thermalize with the plasma. Therefore, for K_L , we must evaluate the momentum distribution to calculate cross sections and time-dilated decay lifetimes. Our fitting of hadronic cross sections finds elastic $K_L N$ scattering to be sizable, especially at high momenta, so we examine how the K_L momentum spectrum inherited from axion decays is modified by the elastic scattering (see Figs. 12, 13 and Appendixes C 3, D for more detail). These were not done in previous studies. For instance, Refs. [17,20,25] used K_L cross sections at the threshold, but as we will show (see Fig. 14), the average cross sections are typically about half of the threshold values.
- (iii) *Include secondary hadrons from scattering.* In the 5th column in Table I, we list hadrons produced from scattering, including those from decays of final-state particles of scattering. We call them secondary hadrons, as opposed to primary hadrons produced from axion decays. These secondary hadrons from scattering were never included in the literature. We find that taking into account the secondary hadrons is essential for estimating δX_n correctly. For example, $nK^- \rightarrow \pi^- \Lambda (\rightarrow p\pi^-)$ seems to convert n to p , but the two secondary π^- s do the reverse conversion twice, and therefore the net conversion is actually

$p \rightarrow n$. Not included in our analysis are the secondary hadrons marked by $\ddagger$, where the $\ddagger$ on a hadron means that its contribution is ignored in the Boltzmann equation for that hadron. The “ $K_S^\ddagger$ ” and “ $K_L^\ddagger$ ” are neglected because the initial nucleons and $K^\pm$ do not have enough kinetic energy at $T \sim T_n$ to overcome the negative Q values. The “mesons ‡ ” from antibaryons are ignored because this channel is subdominant, as we can see in the right plot of Fig. 15.

- (iv) *Include temperature dependence of reaction rates.* Since the temperature dependence of cross sections can be significant, using the updated cross sections, we fully incorporate their temperature dependence in the Boltzmann equations.

With these improvements, we will find that the contributions of all the mesons are relevant. This suggests that the community should like to revisit existing studies of hadronically decaying long-lived particles during BBN.

C. Qualitative picture behind our bounds

We can briefly discuss how hadronic injections with $\tau_a \sim \mathcal{O}(10^{-2})$ s change X_n and how robust our bounds are. In the standard BBN, X_n freezes out when the $n \leftrightarrow p$ conversion processes,

$$n + \nu_e \leftrightarrow p + e^- \quad \text{and} \quad n + e^+ \leftrightarrow p + \bar{\nu}_e, \quad (2.5)$$

become slower than the Hubble expansion rate H . Being mediated by the weak interactions, these reaction rates are small: $\Gamma_{n \leftrightarrow p}^{\text{weak}} \sim \langle \sigma v \rangle n_{e,\nu} \sim 10^{-24} \text{ GeV}(T/\text{MeV})^5$. This smallness implies that even an exponentially suppressed amount of hadronic injection can still contribute to $n \leftrightarrow p$ conversion. The conversion rate by an injected hadron h is given by

$$\Gamma_{n \leftrightarrow p}^h = \langle \sigma v \rangle_{n \leftrightarrow p} n_h \sim \langle \sigma v \rangle_{n \leftrightarrow p} \frac{N_{a \rightarrow h} \Gamma_a n_a}{\langle \sigma v \rangle_{\text{dis}} n_b + \Gamma_h}, \quad (2.6)$$

where $N_{a \rightarrow h}$ is the effective number of h 's produced per axion decay, and $\langle \sigma v \rangle_{n \leftrightarrow p}$ is the sum of the average cross sections for the $n \rightarrow p$ and $p \rightarrow n$ conversions due to scattering with h . The estimate of n_h in the rightmost expression in Eq. (2.6) comes from balancing the appearance and disappearance rates of h .

To estimate $\Gamma_{n \leftrightarrow p}^h$ in Eq. (2.6), recall that, as we already discussed below Eq. (2.3), the Γ_h term in the denominator of Eq. (2.6) is smaller than or comparable to the $\langle \sigma v \rangle_{\text{dis}}$ term. So, we can drop Γ_h and then cancel $\langle \sigma v \rangle_{\text{dis}}$ with $\langle \sigma v \rangle_{n \leftrightarrow p}$ as they are on the same order of magnitude. In addition, due to relativistic decoupling, the axion number density is quite large, $n_a \sim [n_\gamma / g_*^{\text{FO}}] e^{-t/\tau_a}$ with $g_*^{\text{FO}} \sim 100$, in comparison to $n_b \sim 6 \times 10^{-10} n_\gamma$. All combined, we obtain

$$\Gamma_{n \leftrightarrow p}^h \sim 10^{-15} \text{ GeV} \cdot e^{-t/\tau_a} N_{a \rightarrow h} \frac{10^{-2} \text{ s}}{\tau_a}, \quad (2.7)$$

which is vastly larger than $\Gamma_{n \leftrightarrow p}^{\text{weak}}$. Therefore, at these early times, hadronic injections from axion decays completely dominate over the SM reactions.

At later times, the e^{-t/τ_a} suppression of n_a eventually makes $\Gamma_{n \leftrightarrow p}^h$ go below $\Gamma_{n \leftrightarrow p}^{\text{weak}}$, after which the standard weak interactions take over $n \leftrightarrow p$ conversion. We can estimate this transition time, $t = t_{\text{wf}}$, by solving $\Gamma_{n \leftrightarrow p}^{\text{weak}} = \Gamma_{n \leftrightarrow p}^h$, which gives

$$t_{\text{wf}} \sim \tau_a \cdot \left[18 + \log \left(N_{a \rightarrow h} \frac{10^{-2} \text{ s}}{\tau_a} \right) \right]. \quad (2.8)$$

Here, “wf” stands for “waterfall” since X_n falls rapidly and drastically from a high value due to a large $\Gamma_{n \leftrightarrow p}^h$ down to nearly the standard evolution of X_n due to $\Gamma_{n \leftrightarrow p}^{\text{weak}}$. Since the $n \leftrightarrow p$ conversion rate is exponentially larger than the standard rate until the time becomes very close to t_{wf} , we see that any t_{wf} after the standard neutron decoupling time, $t_{\text{wf}} > t_n \sim 0.7 \text{ s}$ ($\tau_a \gtrsim 0.04 \text{ s}$), is robustly ruled out. This concludes a qualitative explanation of how our bounds work, where we have seen how BBN places a robust bound around $\tau_a \sim 10^{-2} \text{ s}$ for heavy QCD axions.

Once the Boltzmann equations, as presented in Sec. V, are solved to obtain a more accurate bound, we find that the above rough estimate is fairly accurate, only off by a factor

of a few. The accurate bound with all parameters included will be derived in Sec. VII B and presented in Eq. (7.13). Here, we present a simplified version with only $N_{a \rightarrow h}$ dependence shown,

$$\tau_a \lesssim \frac{0.02 \text{ s}}{1 + \frac{1}{18} \log[N_{a \rightarrow h}]}. \quad (2.9)$$

This shows that the lifetime bound is quite insensitive to the modification of the axion decay patterns, indicating that our upper bound on the lifetime around 0.02 s is nearly model-independent. Similarly, the result is also only logarithmically dependent on other parameters, such as hadronic cross sections, the initial abundance, and R_{Y_p} in Eq. (2.2).

III. THERMAL PRODUCTION AND ABUNDANCE OF HEAVY QCD AXIONS

In this work, we estimate the axion abundance by following [69]. While their focus is the study of N_{eff} bounds on heavy QCD axions, we share the same axion production mechanism from the $aG\tilde{G}$ coupling. We assume that the postinflation reheating occurs at a sufficiently high temperature that the axions first thermalize with the SM particles and then chemically decouple from the SM sector later, freezing out their number density. This sets the initial axion abundance for our BBN analysis before the axions begin to decay. We simply assume an instantaneous freeze-out at $T = T_{\text{FO}}$ and adopt a thermal Bose-Einstein distribution (with zero chemical potential) for the axion number density. As we will see, our bounds are insensitive to the precise value of the axion abundance so this simplification is not a limitation of our analysis. Then, below the freeze-out temperature, the Boltzmann equation for the axion number density n_a reads

$$\dot{n}_a + 3Hn_a = -\Gamma_{\text{dis}}(n_a - n_a^{(\text{eq})}), \quad (3.1)$$

where Γ_{dis} is the rate of axion disappearance (neglecting processes that reduce the number of axions by more than 1), and $n_a^{(\text{eq})}$ is the equilibrium axion number density determined by the temperature and axion mass, where we ignore thermal contributions to m_a and treat it as constant.

Since we are looking at lifetimes of roughly 10^{-2} s or longer, we can completely ignore axion decays in Γ_{dis} . Then, before the deconfinement-confinement crossover, Γ_{dis} is dominated by $ag \rightarrow gg$. Its reaction rate is given by

$$\Gamma_{ag \rightarrow gg}(T) = \frac{16}{\pi} \left(\frac{g_3^2(T)}{32\pi^2} \right)^2 \frac{T^3}{f_a^2} \mathcal{F}_g(T), \quad (3.2)$$

where we take $\mathcal{F}_g(T)$ from Refs. [94–96], and evaluate the QCD coupling g_3 by taking the $\overline{\text{MS}}$ scale $\mu = T$. $g_3(\mu)$ diverges as it approaches to $\mu = \Lambda_{\text{QCD}} \simeq 300 \text{ MeV}$, so the perturbative expansion becomes eventually invalid.

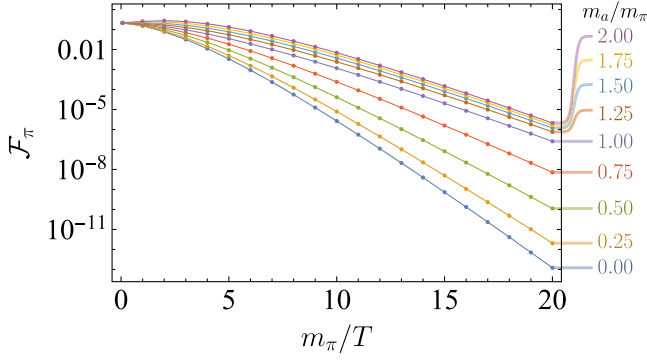


FIG. 2. $\mathcal{F}_\pi(m_a, T)$ used in Eq. (3.3) for different values of m_a . The dots indicate our numerical estimation, and the lines are interpolations of the dots.

Therefore, to ensure the validity of Eq. (3.2) we need to restrict T to be sufficiently high, and we choose $T > T_{\text{QCD}} \equiv 2 \text{ GeV}$.

After the deconfinement-confinement crossover, axion interactions are described in terms of hadrons rather than quarks/gluons. For Γ_{dis} we only include the pion-induced process $a\pi \leftrightarrow \pi\pi$, whose rate is given by

$$\Gamma_{a\pi \rightarrow \pi\pi}(T) = \frac{T^5}{f_a^2 f_\pi^2} \frac{A^2}{(1-r^2)^2} \mathcal{F}_\pi(m_a, T), \quad (3.3)$$

where $r = m_a/m_\pi$ and $A = \frac{1}{3}(m_d - m_u)/(m_d + m_u)$, and $\mathcal{F}_\pi(m_a, T)$ is evaluated by following [69]. Our evaluation of $\mathcal{F}_\pi(m_a, T)$ is plotted in Fig. 2, where different colors correspond to various choices of m_a , as indicated to the right of the plot.⁴

In contrast to pQCD, the approximation $\Gamma_{\text{dis}} = \Gamma_{a\pi \rightarrow \pi\pi}$ is good for sufficiently low values of T , and we choose $T < T_\pi \equiv 0.1 \text{ GeV}$. The validity of this approximation requires the following conditions: (i) Primakoff processes should be suppressed, (ii) the decay and inverse decay rates must be sufficiently small, and (iii) the freeze-out temperature must be sufficiently low, e.g., $T_{\text{FO}} < T_\pi$, since we have ignored processes involving the other mesons. The condition (i) is satisfied with $f_a > 10^5 \text{ GeV}$, while the remaining conditions are well satisfied in the parameter space of interest, that is, $m_a \gtrsim 300 \text{ MeV}$ and $\tau_a \lesssim 0.1 \text{ s}$.

Combining the two cases, we take the axion disappearance rate to be given by

$$\Gamma_{\text{dis}}(T) = \begin{cases} \Gamma_{ag \rightarrow gg}(T) & \text{for } T > T_{\text{QCD}}, \\ \Gamma_{a\pi \rightarrow \pi\pi}(T) & \text{for } T < T_\pi, \end{cases} \quad (3.4)$$

⁴Figure 16 of [69] disagrees with our result due to an error confirmed by the authors of [69], although the error does not affect their other analysis and results.

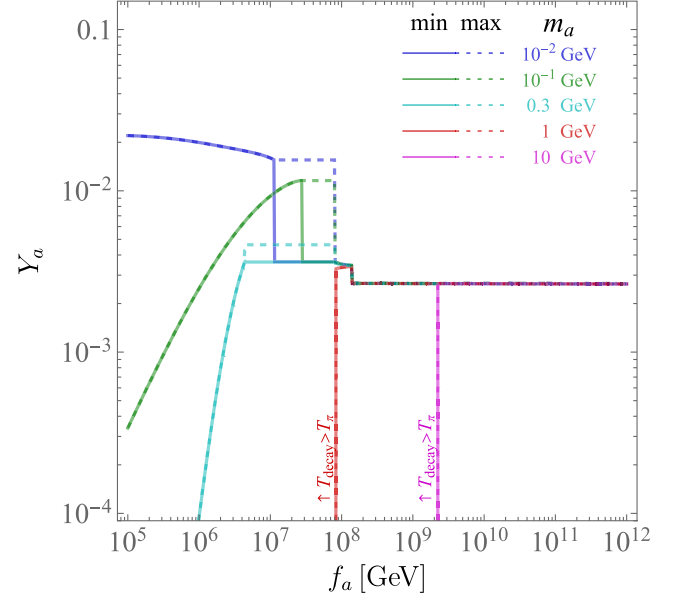


FIG. 3. $Y_a^{(\text{min})}$ (solid lines) and $Y_a^{(\text{max})}$ (dashed lines) as functions of f_a for various choices of m_a represented by different colors. The sharp drops at $f_a \sim 10^8$ and 10^9 GeV for the $m_a = 1$ and 10 GeV cases, respectively, are our truncations by hand because their respective temperature T_{decay} becomes above T_π in the region indicated by “ $\leftarrow$ ”.

but we still do not know Γ_{dis} for $T_\pi < T < T_{\text{QCD}}$. To deal with this lack of information, we adopt the prescription of [69]:

- (1) First, if $\Gamma_{a\pi \rightarrow \pi\pi}(T_\pi) > 3H(T_\pi)$, the freeze-out temperature is determined by $\Gamma_{a\pi \rightarrow \pi\pi}(T_{\text{FO}}) = 3H(T_{\text{FO}})$.
- (2) Otherwise, i.e., if $\Gamma_{a\pi \rightarrow \pi\pi}(T_\pi) < 3H(T_\pi)$, we check the freeze-out with $\Gamma_{ag \rightarrow gg}$. If $\Gamma_{ag \rightarrow gg}(T_{\text{QCD}}) < 3H(T_{\text{QCD}})$, we adopt the freeze-out temperature from $\Gamma_{ag \rightarrow gg}(T_{\text{FO}}) = 3H(T_{\text{FO}})$.
- (3) The situation is uncertain if $\Gamma_{a\pi \rightarrow \pi\pi}(T_\pi) < 3H(T_\pi)$ and $\Gamma_{ag \rightarrow gg}(T_{\text{QCD}}) > 3H(T_{\text{QCD}})$, which implies that T_{FO} is somewhere between T_π and T_{QCD} . To estimate the range of uncertainties due to not knowing T_{FO} , we scan T_{FO} between T_π and T_{QCD} and take the smallest and largest values of $Y_a = n_a^{(\text{eq})}(T_{\text{FO}})/s(T_{\text{FO}})$, denoted as $Y_a^{(\text{min})}$ and $Y_a^{(\text{max})}$, respectively, and will study how results depend on $Y_a^{(\text{max})}$ versus $Y_a^{(\text{min})}$.

Although our bound is insensitive to the precise value of Y_a , as emphasized in Sec. II C, and our scheme above is sufficient for our purpose, there are subtleties in a precise estimation of the axion production rate across the QCD deconfinement-confinement crossover. This has been studied for conventional axion models in terms of axion-hadron interactions after the crossover [97–103] as well as in terms of axion-gluon [94,104,105] and axion-quark [106,107] interactions before the crossover. Estimations incorporating all these interactions are performed in [108] and with a smooth interpolation in [95,96]. In addition, Ref. [109]

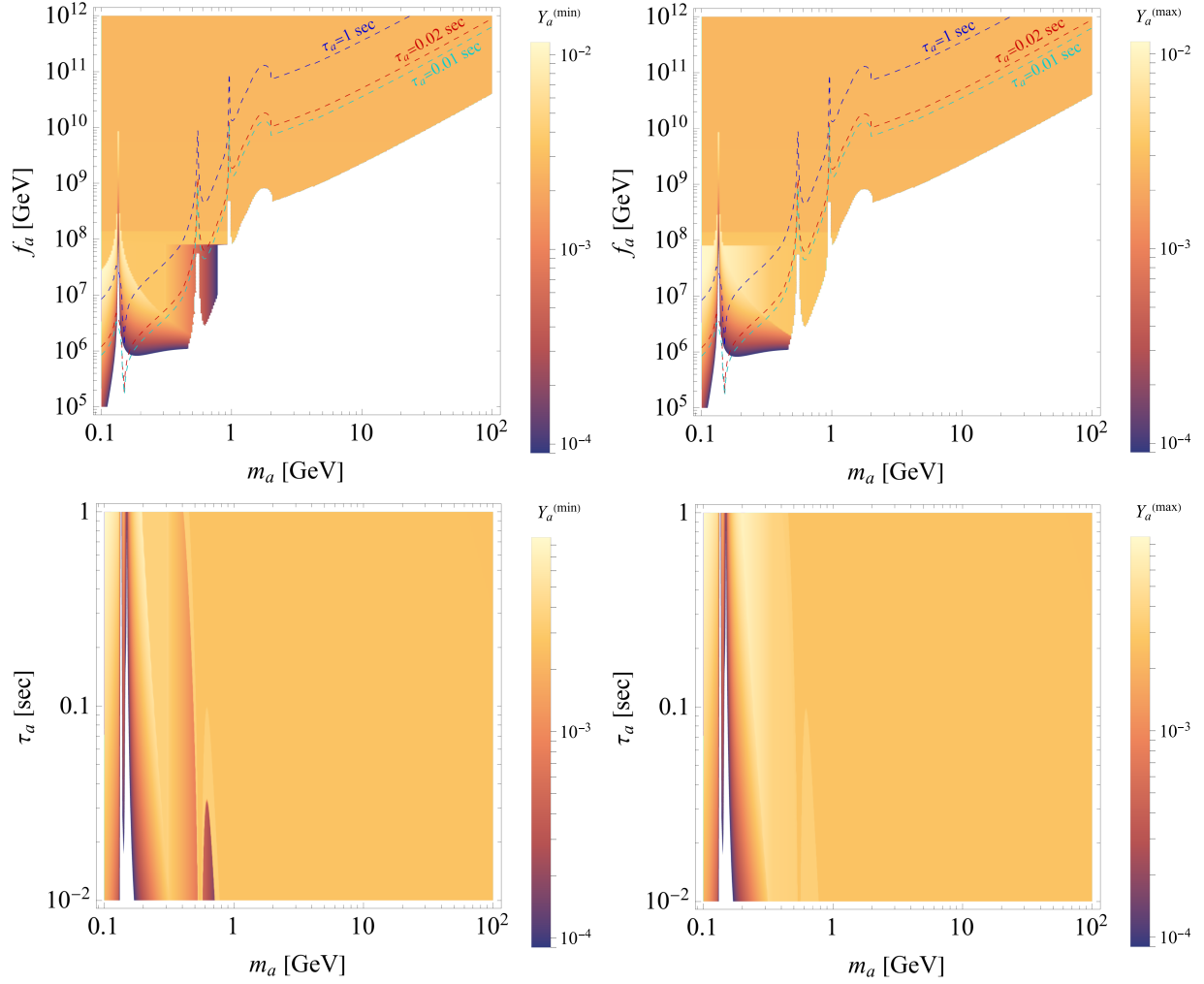


FIG. 4. The left (right) plots show $Y_a^{(\min)}$ ($Y_a^{(\max)}$) in the (m_a, f_a) space (top) and in the (m_a, τ_a) space (bottom). The blue, red, and cyan dashed lines in the upper plots correspond to $\tau_a = 1$ s, 0.02 s, and 0.01 s, respectively.

points out the importance of strong sphalerons, and [110] refines it with the DESI result.

Our estimation of Y_a is shown in Fig. 3. For each choice of m_a , $Y_a^{(\min)}$ ($Y_a^{(\max)}$) is given a function of f_a by a solid (dashed) line. Different values of m_a are represented by different colors. For $m_a \gtrsim 1$ GeV, there is no difference between $Y_a^{(\min)}$ and $Y_a^{(\max)}$. The sharp drops at $f_a \sim 10^8$ and 10^9 GeV for the $m_a = 1$ and 10 GeV cases, respectively, are our truncations, reflecting the fact that, below those values of f_a , the respective decay temperature T_{decay} defined via $\Gamma_a = 3H|_{T=T_{\text{decay}}}$ becomes above T_π , thereby modifying nothing in BBN.

In Fig. 4, top left (top right), the values of $Y_a^{(\min)}$ ($Y_a^{(\max)}$) are shown in the (m_a, f_a) space, while a few values of the axion lifetime are indicated by the dashed lines for $\tau_a = 1$ s (blue), 0.02 s (red), and 0.01 s (cyan). (See Sec. IV for how the lifetime is evaluated.) In the blank region, BBN is not modified because either $T_{\text{decay}} > T_\pi$ at $m_a \gtrsim 1$ GeV or Y_a is exponentially suppressed due to $T_{\text{FO}} \ll m_a$ at $m_a \lesssim 1$ GeV.

The bottom plots of Fig. 4 again show $Y_a^{(\min)}$ and $Y_a^{(\max)}$ but this time in the (m_a, τ_a) space. Since the axions freeze out while still relativistic for $m_a \gtrsim 1$ GeV, Y_a is large (about $3 \times 10^{-3} \gg \eta_b$) in most of the space ($m_a > 0.3$ GeV and $0.01 \text{ s} < \tau_a < 1 \text{ s}$). Observe that a noticeable difference between $Y_a^{(\min)}$ and $Y_a^{(\max)}$ only appears in a narrow region of $0.2 \text{ GeV} \lesssim m_a \lesssim 0.7 \text{ GeV}$.

Due to the large axion yield, its energy density can temporarily dominate the Universe as matter until the axions begin to decay into radiation. The Universe is axion matter dominated, while the temperature T falls in the range

$$6 \text{ MeV} \left(\frac{m_a}{3 \text{ GeV}} \right) \left(\frac{Y_a}{3 \times 10^{-3}} \right) \gtrsim T \gtrsim T_{\text{decay}}, \quad (3.5)$$

where the beginning temperature comes from comparing $\rho_a = m_a Y_a s(T)$ with $\rho(T) = \frac{4}{3} T s(T)$. This shows that as the axion mass is increased, the axion-dominated period becomes longer, and the standard radiation-dominated

background cosmology becomes increasingly inappropriate. Our numerical codes properly handle the presence of axion domination by solving the Boltzmann equation for the axion number density (after freeze-out) and appropriately modifying the Hubble parameter by the axion energy density.

Additionally, since the baryon asymmetry is diluted during the axion-dominated epoch, we begin with a larger initial baryon asymmetry such that the dilution reduces it to the correct value of η_b measured by CMB. We will give more detail in Sec. V B.

IV. HADRONIC AXION DECAYS

We have two frameworks depending on m_a . For $m_a > 2$ GeV, we use pQCD and consider two hadronization models: the string fragmentation model used in PYTHIA and the cluster model used in Herwig. For $m_a < 2$ GeV, both pQCD and these hadronization algorithms fail, so we instead adopt a data-driven method proposed and developed in [60]. This framework is based on the chiral perturbation theory and vector meson dominance, but extends the domain of validity to higher energies than $\sim 4\pi f_\pi$ with the help of experimental data. While this is not a controlled approximation, it gives sensible results for the limited purpose of estimating the decay width and various branching fractions of the axion, as validated in [60].

We will present the following quantities relevant for our BBN analysis: (i) the total decay width, (ii) the average number of each hadron per axion decay, and (iii) the energy spectrum of K_L . As we discussed in Sec. II A, the hadrons that can affect the neutron-to-proton ratio are limited to the $\pi^\pm$, $K^\pm$, K_L , (anti-)proton, and (anti-)neutron. These are sufficiently long-lived to have time to interact with the nucleons in the plasma. Therefore, for (ii), we only count the average number of each of these hadrons per axion decay. Other hadrons, e.g., K_S , hyperons, etc., decay before they have a chance to scatter with the nucleons.

A. $m_a > 2$ GeV

For $m_a > 2$ GeV, we obtain the total decay width by using the NNLO calculation of pseudoscalar decaying to gluons in Ref. [78]. We obtain

$$\Gamma(a \rightarrow gg) = \frac{2}{\pi} \left(\frac{\alpha_s(\mu)}{8\pi} \right)^2 \frac{m_a^3}{f_a^2} \times [1 + A(\mu)\alpha_s(\mu) + B(\mu)\alpha_s^2(\mu)], \quad (4.1)$$

where μ is the renormalization scale. For the coefficient functions $A(\mu)$ and $B(\mu)$, see Eq. (A1) in Appendix A.

We evaluate uncertainties in the total width by varying the renormalization scale from $\mu = m_a/2$ to $\mu = 2m_a$, which is shown by the blue band in Fig. 5. Note that what BBN directly constrains is the axion lifetime, so our bound on the lifetime is robust and independent of this uncertainty.

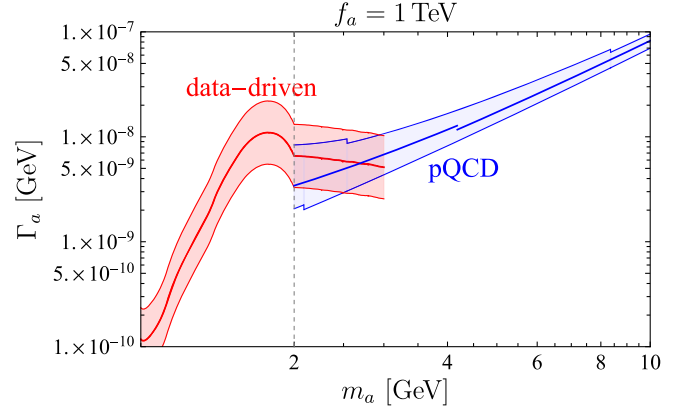


FIG. 5. The total decay width Γ_a as a function of m_a . Here, f_a is fixed to 1 TeV, but one can rescale the rate by $(\text{TeV}/f_a)^2$ for a different value f_a . Estimations based on pQCD and the data-driven method are shown in blue and red, respectively. The bands represent the uncertainties of the estimations, where the blue band is determined by varying the renormalization scale from $\mu = m_a/2$ to $\mu = 2m_a$, while we assign a factor of 2 uncertainty to the data-driven estimation for the red band.

However, the bound on f_a is affected by the uncertainty when it is translated from the bound on the lifetime through the above relation.

For the branching fractions, we use the two most widely used parton shower and hadronization programs, PYTHIA 8.306 [79] and Herwig v7.3.0 [80–82], to obtain the average numbers of $\pi^\pm$, $K^\pm$, K_L , N , and $\bar{N}$ from the axion decay. We generate 10^4 parton-showering and hadronization events of $a \rightarrow gg$ by PYTHIA and Herwig at each value of the axion mass. Since the predictions from the two programs are sometimes different, we retain the result using both of them.

In Fig. 6, the average numbers of hadrons per axion decay are presented, where the estimations from PYTHIA and Herwig are shown in red and blue, respectively. The (nearly invisible) error bars represent the statistical uncertainties coming from the finite number of simulated events. Note that indirect contributions from the decays of unstable hadrons such as $\rho \rightarrow \pi\pi$ are also included in the counts. We find that the average number of pions per axion decay in the plotted axion mass range is ~ 1 –10, while the average numbers of charged kaons and (anti-)nucleons are ~ 0.2 –1.2 and ~ 0 –0.7, respectively. Since our framework of $K_L N$ scattering is within only $2 \rightarrow 2$ scattering processes, we restrict the K_L momentum to be less than 2 GeV in the axion rest frame, which will be the rest frame of the background nucleons during BBN. For comparison, we also include the average K_L numbers without this restriction. While the two programs agree fairly well for the mesons at large values of m_a , $\mathcal{O}(1)$ discrepancies are present at all m_a for the baryons.

In Figs. 7 and 8, the predicted energy distributions of K_L by PYTHIA and Herwig, respectively, are presented for various values of m_a . As emphasized in Sec. II B, this

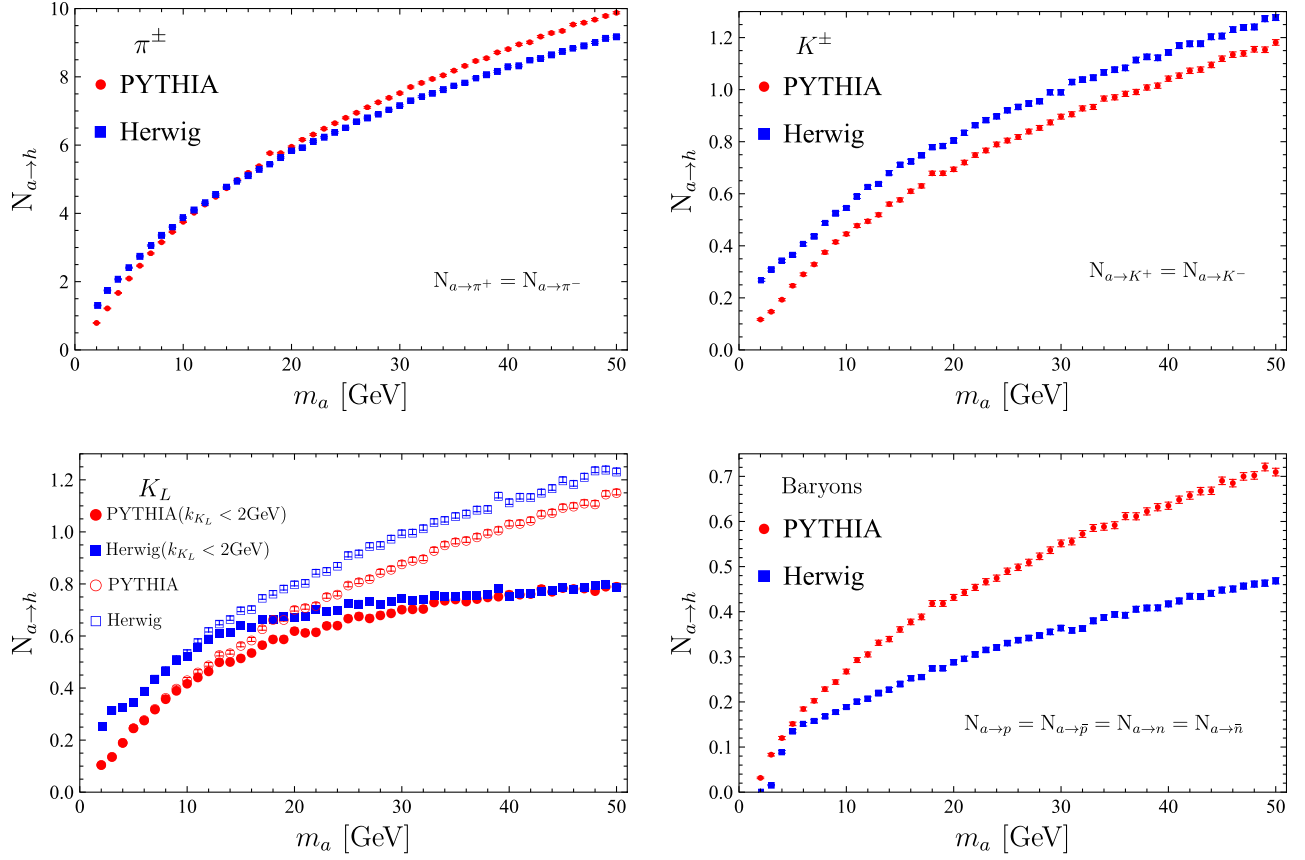


FIG. 6. Average numbers of mesons and (anti-)baryons as functions of the axion mass. The error bars are due to statistical uncertainties of finite simulation samples. For K_L , the momentum range is restricted as $k_{K_L} < 2$ GeV in our analysis, but the original numbers without this cut are shown for comparison. The tabulated samples are given in Ref. [111].

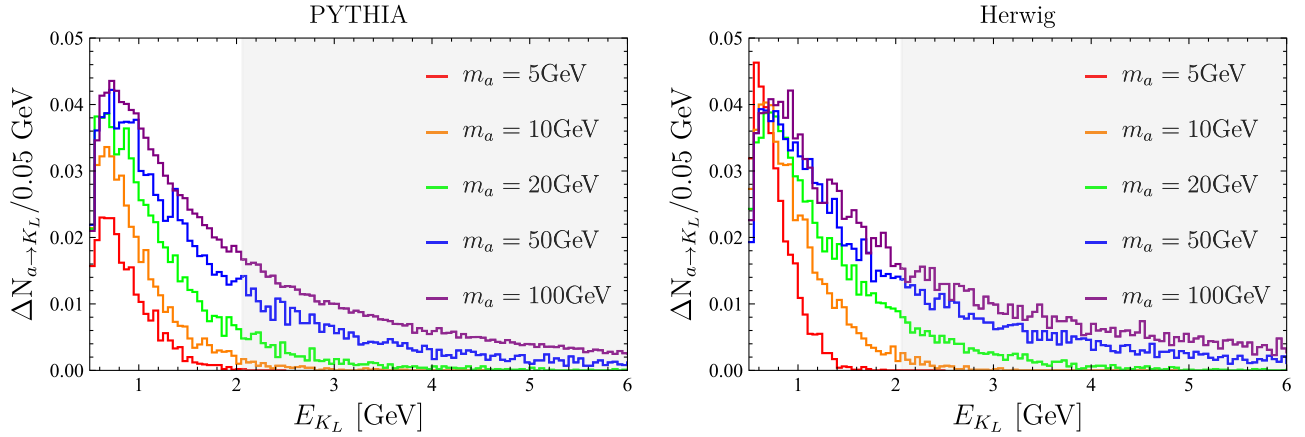


FIG. 7. Energy distribution of K_L from the axion decay for $m_a = 5, 10, 20, 50$, and 100 GeV, obtained from PYTHIA (left) and Herwig (right).

K_L 's spectrum is distorted by elastic scattering of K_L with background baryons. The reshaped spectra are shown in Figs. 12 and 13.

We have also found fundamental problems with hadronization by both programs. For example, neither program respects the fundamental property of the axion that it is a

pseudoscalar. PYTHIA allows the axion to decay to two pions. While this particular forbidden decay happens to be absent in Herwig, that is not because it respects parity but because of an accidental outcome of kinematical restrictions imposed in its cluster hadronization model. We add some relevant discussion in Appendix B.

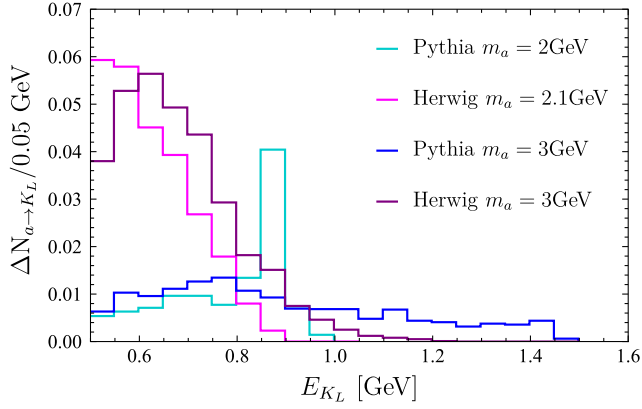


FIG. 8. Energy distribution of K_L from the axion decay for low axion masses, where large discrepancies between PYTHIA and Herwig predictions are present.

Fortunately, none of these subtleties impact our final result. The $\mathcal{O}(1)$ discrepancy in the baryon counts is inconsequential because our final constraint is dominantly determined by pions and kaons. The lifetime upper bound is only logarithmically sensitive to changes in the amounts of hadrons, as one has seen in the $N_{a \rightarrow h}$ dependence in Eq. (2.9). The K_L spectrum uncertainties affect the averaged cross sections by at most 20% (see Fig. 14), but they also only logarithmically impact the lifetime bound, as seen in Eq. (7.13). The absences of impacts of these subtleties are seen in Fig. 20, left. Also see Sec. VII B for further discussions.

B. $m_a < 2$ GeV

For $m_a < 2$ GeV, not only do perturbative QCD calculations become increasingly inaccurate but also the shower/hadronization by PYTHIA and Herwig fail. Instead, we adopt a data-driven method based on the chiral perturbation theory and vector meson dominance theory, which was first

developed in Ref. [60] and later refined in Ref. [74]. We take branching fractions necessary for our purpose from Ref. [75]. See also Refs. [76,77] for recent calculations. Again, recall that our bound on the axion lifetime is not sensitive to the precise values of the branching fractions, as can be seen in Eq. (2.9).

Adding up all the decay widths of individual exclusive decay modes from the data-driven method, we obtain the total decay width of the axion, shown in red in Fig. 5. Since there is no simple way to estimate uncertainties in the data-driven method, we assign a factor of 2 uncertainty for the total decay width. As can be seen in Fig. 5, the total decay width predictions from the pQCD and data-driven methods overlap quite well when the latter is extended to $m_a = 3$ GeV.

The axion decay modes included in our analysis are $a \rightarrow \pi\pi\gamma, 3\pi, \eta\pi\pi, \eta'\pi\pi, KK\pi, \rho\rho, \omega\omega$, and $\gamma\gamma$, and their partial widths with $f_a = 1$ TeV are shown in Fig. 9, left. Unimportant modes like $a \rightarrow 3\pi^0$ are not shown there, but they are included in the total width. To obtain the average numbers of hadrons per axion decay relevant for BBN, we further decay short-lived particles such as ρ down to $\pi^\pm$, $K^\pm$, and K_L . Figure 9, right, shows how individual modes contribute to the effective number of $\pi^\pm$. The effective numbers of $K^\pm$, K_L , and $\pi^\pm$ are shown in Fig. 10.

The K_L energy distribution is obtained from $a \rightarrow KK\pi$ decay, where we assume a uniform amplitude in the Dalitz plot for simplicity. For $\pi^\pm$ and $K^\pm$, we only need their average number per axion decay, presented in Fig. 10.

V. NEUTRON FREEZE-OUT

In the standard BBN, neutrinos are the first ones to freeze out, occurring around $T \sim 2$ MeV. As the Universe continues to cool down across $T \sim m_e$, electrons and positrons annihilate into photons, injecting energy and entropy into the plasma. At around a similar temperature

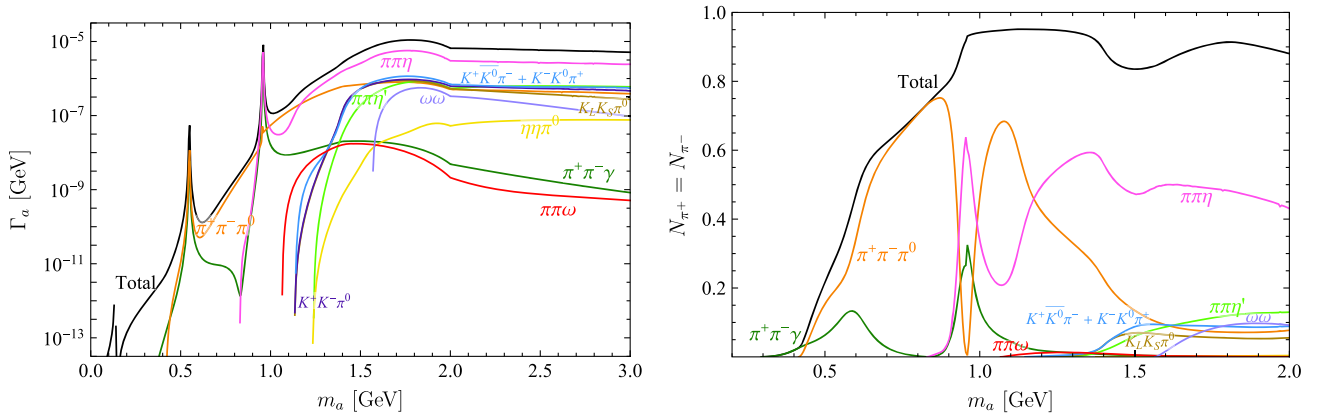


FIG. 9. Left: the axion decay widths from the data-driven method for $f_a = 1$ TeV as a function of m_a , taken from Ref. [75]. Right: effective number of π^+ (which is also equal to that of π^-) from each decay channel. The $a \rightarrow 3\pi^0$ decay is not shown since the π^0 's decay rapidly to photons before interacting with the background baryons.

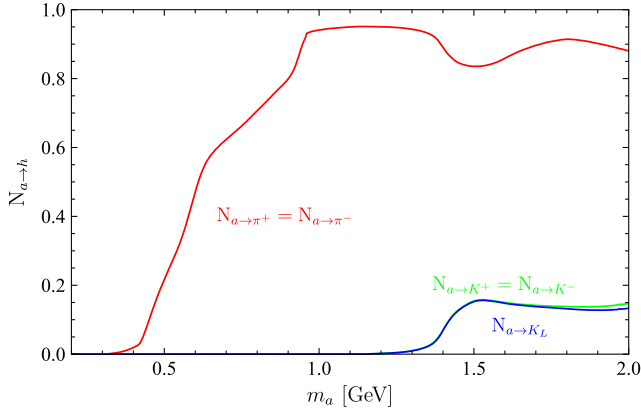


FIG. 10. Effective number of $\pi^\pm$, $K^\pm$, and K_L per axion decay as a function of m_a .

($T = T_n \sim 1$ MeV), the weak interactions interconverting neutrons and protons become slower than the expansion rate and the neutron fraction X_n freezes out. After this neutron freeze-out, X_n gradually decreases due to neutron β -decay until the temperature reaches the so-called deuterium bottleneck temperature T_D , above which the deuterium photodissociation rate due to the background photons is too high to initiate deuterium synthesis.

Once T drops below T_D and a sufficient amount of deuterium is synthesized, a chain of nuclear processes happens, but the net result of this is that nearly all the neutrons end up in ${}^4\text{He}$, with all other elements having exponentially suppressed abundances. Thus, to an excellent approximation, the number of ${}^4\text{He}$ is given by half the number of neutrons at T_D . The latter can be calculated by solving the Boltzmann equation for X_n without including its fusion processes. The largest source of uncertainty in our calculation of Y_p is the value of T_D because $X_n \propto \exp(-t_D/\tau_n) \simeq 1 - t_D/\tau_n$, where $t_D \sim 200$ s is the time corresponding to T_D and τ_n is the neutron lifetime (878.4 s). Nevertheless, as we discussed in Sec. II C, our axion lifetime bound comes from $\delta X_n/X_n^{\text{SM}}$, which is insensitive to the precise value of T_D , as δX_n and X_n^{SM} have approximately the same T_D dependence. To maximally implement this cancellation of T_D sensitivities, we define T_D via $Y_p(\text{observed}) = 2X_n^{\text{SM}}(T_D)$ with $X_n^{\text{SM}}(T_D)$ calculated by the *same* Boltzmann equations as what we use to calculate X_n with the axion removed. This insensitivity will be shown explicitly in Fig. 15, right, in Sec. VII A.

A. Boltzmann equations and modified background cosmology

The Boltzmann equation for X_n depends on the Hubble expansion rate and the $n \leftrightarrow p$ conversion rates of the standard weak interactions and hadronic interactions with the injected hadrons from axion decays. We write evolution equations for (i) the photon temperature, neutrino energy

density, and axion number density, (ii) total baryon number density, and (iii) number densities of hadrons in Table I, and X_n . As illustrated in Fig. 1, we solve the systems of equations in (i)–(iii) in this order, where the temperature ranges are (i) $T_{\text{FO}} > T > T_f = 64$ keV, (ii) $T_i = 30$ MeV $> T > T_f$, and (iii) $T_i > T > T_f$. In our Boltzmann equations, T always means the photon temperature.

The evolution of T below T_i is determined by

$$\begin{aligned} \dot{\rho}_{\ell\gamma} + 3H(\rho_{\ell\gamma} + p_{\ell\gamma}) \\ = \Gamma_a n_a m_a + C_{\nu_e}(\rho_{\nu_e}^2 - \rho_{\nu_e,\text{eq}}^2) + 2C_{\nu_\mu}(\rho_{\nu_\mu}^2 - \rho_{\nu_\mu,\text{eq}}^2), \end{aligned} \quad (5.1)$$

where $\rho_{\ell\gamma}$ ($p_{\ell\gamma}$) is the sum of the energy densities (pressures) of $e^\pm$, $\mu^\pm$, and photons, ρ_{ν_i} ($i = e, \mu$) is the energy density of neutrino ν_i , $C_{\nu_e} = 0.68G_F^2 T$ ($C_{\nu_\mu} = 0.3G_F^2 T$) is the total rate for both charged and neutral current (only neutral current) processes [72], and $\rho_{\nu_i,\text{eq}} = 2 \cdot \frac{7}{8} \frac{\pi^2}{30} T^4$. Since $\rho_{\ell\gamma}$ and $p_{\ell\gamma}$ are simply functions of T , Eq. (5.1) can be used to convert between T and t , as we will see in more detail in the next subsection. The factor of 2 in front of the C_{ν_μ} term of Eq. (5.1) is to take into account ν_τ .

The Boltzmann equations of the neutrino sector can be written as

$$\begin{aligned} \dot{\rho}_{\nu_e} + 4H\rho_{\nu_e} &= -C_{\nu_e}(\rho_{\nu_e}^2 - \rho_{\nu_e,\text{eq}}^2), \\ \dot{\rho}_{\nu_\mu} + 4H\rho_{\nu_\mu} &= -C_{\nu_\mu}(\rho_{\nu_\mu}^2 - \rho_{\nu_\mu,\text{eq}}^2). \end{aligned} \quad (5.2)$$

We neglect neutrinos that are produced from axion decay chains, e.g., $a \rightarrow \pi^+ + \dots \rightarrow \mu^+ \nu_\mu + \dots \rightarrow e^+ \nu_e \nu_\mu + \dots$, since this contribution is negligible for the lifetimes of our interest, $\tau_a = \mathcal{O}(0.01)$ s. A bulk of the axion energy becomes the kinetic energies of primary daughter particles, all of which except K_L quickly lose their energies into the plasma via electromagnetic interactions before they decay. Energy injection from axion decays into the neutrino sector could potentially be relevant to Eq. (5.2) only after neutrino decoupling, but the axion energy density then would be very small, as most of the axions have decayed by then. Given the axion energy injection into $\rho_{\ell\gamma}$ is already as small as $\mathcal{O}(1)\%$, energy injection into the neutrino sector from the axion decay chains is even smaller.

The Hubble expansion rate is affected by the axion energy density,

$$H = \frac{1}{M_{\text{Pl}}} \sqrt{\frac{8\pi}{3}(\rho_{\ell\gamma} + \rho_\nu + m_a n_a)}, \quad (5.3)$$

with the Planck mass $M_{\text{Pl}} = 1.221 \times 10^{19}$ GeV. We neglect small contributions to the energy density coming from the baryon asymmetry and thermal hadrons.

The evolutions of the baryon and axion number densities are given by their conservation in the comoving box while including the decay rate of the axion,

$$\dot{n}_b + 3Hn_b = 0, \quad (5.4)$$

$$\dot{n}_a + 3Hn_a = -\Gamma_a n_a. \quad (5.5)$$

The Boltzmann equations for other hadron number densities denoted by n_h can be written as

$$\begin{aligned} \dot{n}_h = & N_{a \rightarrow h} \Gamma_a n_a - \{3H + \tilde{\Gamma}_h + \langle \sigma v(n + h \rightarrow \dots) \rangle n_n \\ & + \langle \sigma v(p + h \rightarrow \dots) \rangle n_p\} n_h \\ & + \sum_{h' \neq h} \{N_{h' \rightarrow h} \tilde{\Gamma}_{h'} + \langle \sigma v(n + h' \rightarrow h \dots) \rangle n_n \\ & + \langle \sigma v(p + h' \rightarrow h \dots) \rangle n_p\} n_{h'}, \end{aligned} \quad (5.6)$$

where $h(h') = \pi^+, \pi^-, K^+, K^-, K_L, \bar{p}$, and $\bar{n}$. $\tilde{\Gamma}_{h(h')}$ is the effective decay width of $h(h')$ including the averaged inverse boost factor for K_L , as we will discuss in Sec VI. We ignore the boost of the other hadrons because they immediately become nonrelativistic through thermalization. $N_{a \rightarrow h}$ is the averaged number of h per axion decay (i.e., the sum of branching ratios multiplied by the multiplicity of h), see Figs. 6 and 10. The produced hadrons either decay or scatter against the background nucleon, including both $n \leftrightarrow p$ conversion and nonconversion processes.

The second line in Eq. (5.6) accounts for the secondary production of h through the decay or scattering of another hadron h' , where we omit the multiplicity factors to keep the simplicity of the expression. For instance, we multiply a factor of 2 in the process of $nK^- \rightarrow Y\pi^- \rightarrow p\pi^-\pi^-$ for the case of $h = \pi^-$ and $h' = K^-$. If K_S is produced, its decay product into $\pi^\pm$ is accounted for, assuming the decay is instantaneous.

Secondary hadrons are necessary to correctly evaluate the net conversion effects, especially for channels initiated by a kaon, as discussed in the third bullet point of Sec. II B and shown in Table I. However, we omit the secondary “ $K_S^\pm$ ” and “ $K_L^\pm$ ” listed in Table I because the processes that produce them are only relevant at temperatures higher than T_n . Secondary mesons from antinucleon-nucleon annihilation are also omitted, as this channel is subdominant. While they may become important at higher temperatures, we have checked that the conversion due to these secondary mesons remains subdominant around T_n .

Finally, the Boltzmann equation of X_n is given by

$$\dot{X}_n = \frac{\dot{n}_n n_b - \dot{n}_b n_n}{n_b^2} = -\Gamma_{n \rightarrow p} X_n + \Gamma_{p \rightarrow n} (1 - X_n). \quad (5.7)$$

The total conversion rates of $n \rightarrow p$ and $p \rightarrow n$, including both the SM- and axion-induced processes, are encoded in $\Gamma_{n \rightarrow p}$ and $\Gamma_{p \rightarrow n}$, to be described in detail in Eqs. (5.19) and (5.20).

Note that the above equations remain valid regardless of the axion energy density. They hold even during an axion-

dominated era, Eq. (3.5). However, our approximation $\rho_a = m_a n_a$ breaks down for very light axions, which are outside the scope of this work.

B. Steps of numerical calculation

For interested readers who want to reproduce our results, here we present specific procedures to numerically solve the evolution.

In order to solve the Boltzmann equations, we redefine various quantities to be dimensionless (we divide or multiply a power of the temperature and denote it with a hat, e.g., $\hat{\rho}_{\nu_i} = T^{-4} \rho_{\nu_i}$, $\hat{n}_b = T^{-3} n_b$, etc.). The overall routine is outlined in Fig. 1. For the photon temperature T , we define a dimensionless variable

$$z \equiv m_e/T, \quad (5.8)$$

which is commonly used in the BBN studies, e.g., see Refs. [112–114]. By the chain rule, we have $\dot{\rho}_{\ell\gamma} = \frac{\partial \rho_{\ell\gamma}}{\partial z} \dot{z}$, and therefore, from Eq. (5.1),

$$\begin{aligned} \dot{z} = & \left(\frac{\partial \rho_{\ell\gamma}}{\partial z} \right)^{-1} \{ -3H(\rho_{\ell\gamma} + p_{\ell\gamma}) + \Gamma_a n_a m_a \\ & + C_{\nu_e}(\rho_{\nu_e}^2 - \rho_{\nu_{i,\text{eq}}}^2) + 2C_{\nu_\mu}(\rho_{\nu_\mu}^2 - \rho_{\nu_{i,\text{eq}}}^2) \}. \end{aligned} \quad (5.9)$$

Assuming that the visible sector is always in thermal equilibrium, we take

$$\rho_\gamma(z) = 3p_\gamma = 2 \cdot \frac{\pi^2}{30} \left(\frac{m_e}{z} \right)^4, \quad (5.10)$$

$$\rho_\ell(z) = 4 \left(\frac{m_e}{z} \right)^4 \int_0^\infty \frac{d\hat{p}}{2\pi^2} \frac{\hat{p}^2 \sqrt{\hat{p}^2 + \left(\frac{m_\ell}{m_e} \right)^2 z^2}}{e^{\sqrt{\hat{p}^2 + \left(\frac{m_\ell}{m_e} \right)^2 z^2}} + 1}, \quad (5.11)$$

$$p_\ell(z) = 4 \left(\frac{m_e}{z} \right)^4 \int_0^\infty \frac{d\hat{p}}{2\pi^2} \frac{\hat{p}^4}{3\sqrt{\hat{p}^2 + \left(\frac{m_\ell}{m_e} \right)^2 z^2}} \frac{1}{e^{\sqrt{\hat{p}^2 + \left(\frac{m_\ell}{m_e} \right)^2 z^2}} + 1}, \quad (5.12)$$

with $\ell = e$ and μ .

Firstly, we solve the Boltzmann equations for the background cosmology with

$$\hat{\rho}'_{\nu_i} = \left(\frac{4}{z} - \frac{4H}{\dot{z}} \right) \hat{\rho}_{\nu_i} - \frac{\Gamma_{\nu_i}}{\dot{z}} (\hat{\rho}_{\nu_i}^2 - \hat{\rho}_{\nu_{i,\text{eq}}}^2) \quad \text{for } i = e, \mu, \quad (5.13)$$

$$\hat{n}'_a = \left(\frac{3}{z} - \frac{3H}{\dot{z}} \right) \hat{n}_a - \frac{\Gamma_a}{\dot{z}} \hat{n}_a, \quad (5.14)$$

where $'$ denotes $\frac{\partial}{\partial z}$. From the axion freeze-out down to T_i , since the SM particles are in thermal equilibrium, we evaluate only Eq. (5.14), using $Y_a(T_{\text{FO}})$ that is obtained in

Sec. III for each (m_a, f_a) and H accounting for all relevant degrees of freedom.⁵ After $\hat{n}_a$ at T_i is obtained, we solve the Boltzmann equations from T_i to T_f ($z_f = 8$) to extract the necessary information on the modified cosmology, such as $\dot{z}$ and $H(z)$.

Once $\hat{\rho}_\nu = \hat{\rho}_{\nu_e} + 2\hat{\rho}_{\nu_\mu}$ is obtained, we can also estimate N_{eff} by

$$N_{\text{eff}} = \left(2 \cdot \frac{7}{8} \frac{\pi^2}{30} \left(\frac{4}{11}\right)^{4/3}\right)^{-1} \hat{\rho}_\nu(z_f), \quad (5.15)$$

for a large $z_f \gg 1$. As a check, we obtain $N_{\text{eff}} \simeq 3.040$ for the standard BBN, which agrees well with the PDG estimation $N_{\text{eff}} = 3.045$ [84].

In the presence of axions, N_{eff} is reduced by additional radiation, and we obtain the N_{eff} bound following the method of Ref. [69]. From the Planck results [93], we adopt $N_{\text{eff}} > 2.43$ ($\Delta N_{\text{eff}} > -0.61$), while Ref. [69] uses $N_{\text{eff}} > 2.62$ assuming a fixed Y_p value. More specifically, we obtain the 2σ bound, $N_{\text{eff}} > 2.43$, from Fig. 41 of Ref. [93] by projecting the fitting contours without Y_p measurements (blue contours) onto the N_{eff} axis. We also do not use other stronger bounds in [93] that assume underlying physics incompatible with hadronic injections during BBN (see Ref. [116] for a related discussion).

In the next step, we solve for the total baryon number density,

$$\hat{n}'_b = \left(\frac{3}{z} - \frac{3H}{\dot{z}}\right) \hat{n}_b. \quad (5.16)$$

The boundary condition of $\hat{n}_b$ must be fixed at a low temperature by the CMB data as

$$\hat{n}_b(z_f) = T_f^{-3} \eta_b n_\gamma(z_f), \quad (5.17)$$

where $\eta_b(T_{\text{CMB}}) = 6.115 \times 10^{-11}$ is provided in Sec. 24.4 of Ref. [84] based on the Planck measurement. To a good approximation, $\eta_b(T_f) = \eta_b(T_{\text{CMB}})$ holds.

Finally, given that $\dot{z}$, H , and $\hat{n}_b$ are obtained over the relevant temperature, we can solve the evolution for the neutron fraction, Eq. (5.7), which is differentiated with respect to z ,

$$X'_n = \frac{1}{z} [-\Gamma_{n \rightarrow p} X_n + \Gamma_{p \rightarrow n} (1 - X_n)]. \quad (5.18)$$

The total conversion rates of $n \rightarrow p$ and $p \rightarrow n$ are given by SM reactions and axion-related ones ($\Delta\Gamma$),

⁵Since it is before the neutrino decoupling temperature, the SM plasma energy density ρ_{SM} simply evolves with $\dot{\rho}_{\text{SM}} + 3H(\rho_{\text{SM}} + p_{\text{SM}}) = \Gamma_a n_a m_a$, $\rho_{\text{SM}} = \frac{\pi^2}{30} g_*(T) T^4$, $p_{\text{SM}} = w \rho_{\text{SM}}$, and $M_{\text{Pl}}^2 H^2 = \frac{8\pi}{3} (m_a n_a + \rho_{\text{SM}})$. We take g_* and w from Ref. [115].

$$\Gamma_{n \rightarrow p} = \Gamma_{n\nu_e \rightarrow p e^-} + \Gamma_{ne^+ \rightarrow p \bar{\nu}_e} + \Gamma_{n \text{ decay}} + \Delta\Gamma_{n \rightarrow p}, \quad (5.19)$$

$$\Gamma_{p \rightarrow n} = \Gamma_{p \bar{\nu}_e \rightarrow n e^+} + \Gamma_{pe^- \rightarrow n \nu_e} + \Delta\Gamma_{p \rightarrow n}, \quad (5.20)$$

respectively. Here, we take the beta decay rate $\Gamma_{n \text{ decay}} = (878.4 \text{ s})^{-1}$ [84]. For the weak interaction rates, we adopt the equations developed in Ref. [117],⁶

$$\Gamma_{n\nu_e \rightarrow p e^-} \simeq \frac{1 + 3g_A^2}{2\pi^3} G_F^2 Q^5 J_1^\infty(T, T_{\nu_e}), \quad (5.21)$$

$$\Gamma_{ne^+ \rightarrow p \bar{\nu}_e} \simeq \frac{1 + 3g_A^2}{2\pi^3} G_F^2 Q^5 J_\infty^{-m_e/Q}(T, T_{\nu_e}), \quad (5.22)$$

$$\Gamma_{p \bar{\nu}_e \rightarrow n e^+} \simeq \frac{1 + 3g_A^2}{2\pi^3} G_F^2 Q^5 K_\infty^{-m_e/Q}(T, T_{\nu_e}), \quad (5.23)$$

$$\Gamma_{pe^- \rightarrow n \nu_e} \simeq \frac{1 + 3g_A^2}{2\pi^3} G_F^2 Q^5 K_1^\infty(T, T_{\nu_e}), \quad (5.24)$$

where $g_A \simeq 1.27$, and $Q = m_n - m_p - m_e \simeq 1.293 \text{ MeV}$. The J_a^b and K_a^b functions are defined by

$$J_a^b(T, T_{\nu_e}) = \int_a^b \sqrt{1 - \frac{(m_e/Q)^2}{q^2}} \frac{q^2(q-1)^2 dq}{(1 + e^{-\frac{Q}{T}q})(1 + e^{-\frac{Q}{T_{\nu_e}}(q-1)})}, \quad (5.25)$$

$$K_a^b(T, T_{\nu_e}) = \int_a^b \sqrt{1 - \frac{(m_e/Q)^2}{q^2}} \frac{q^2(q-1)^2 dq}{(1 + e^{\frac{Q}{T}q})(1 + e^{-\frac{Q}{T_{\nu_e}}(q-1)})}. \quad (5.26)$$

We evaluate the neutrino temperature T_{ν_e} by solving $\rho_{\nu_e} = 2 \frac{7}{8} \frac{\pi^2}{30} T_{\nu_e}^4$, assuming that neutrinos follow the Fermi-Dirac distribution.

The axion effect through the hadron injection is contained in $\Delta\Gamma_{n \rightarrow p}$ and $\Delta\Gamma_{p \rightarrow n}$. They can be expressed as

$$\Delta\Gamma_{n \rightarrow p} = \sum_h \langle \sigma v (n + h \rightarrow p \cdots) \rangle X_h \left(\frac{m_e}{z}\right)^3 \hat{n}_b, \quad (5.27)$$

$$\Delta\Gamma_{p \rightarrow n} = \sum_h \langle \sigma v (p + h \rightarrow n \cdots) \rangle X_h \left(\frac{m_e}{z}\right)^3 \hat{n}_b, \quad (5.28)$$

where the sum runs over $h = \pi^+, \pi^-, K^+, K^-, K_L, \bar{p}$, and $\bar{n}$, and we define $X_h \equiv n_h/n_b$. Note that, unlike in Eq. (5.6), only the processes that convert $n \leftrightarrow p$ are relevant here, as indicated by p or n appearing in the final state. The cross sections are discussed in Sec. VI.

⁶Note that we do not assume $\Gamma_{p \rightarrow n} = e^{-Q/T} \Gamma_{n \rightarrow p}$, which is not satisfied after the neutrino decoupling where $T \neq T_{\nu_e}$.

For X_h in the presence of axion hadronic decays, we solve the following Boltzmann equation:

$$\begin{aligned}
 X'_h = \frac{1}{z} \Big[& N_{a \rightarrow h} \Gamma_a X_a - \left\{ \tilde{\Gamma}_h + (\langle \sigma v(n+h \rightarrow p \cdots) \rangle X_n \right. \\
 & + \langle \sigma v(p+h \rightarrow n \cdots) \rangle (1-X_n)) \left(\frac{m_e}{z} \right)^3 \hat{n}_b \Big\} X_h \\
 & + \sum_{h'} \left\{ N_{h' \rightarrow h} \tilde{\Gamma}_{h'} + (\langle \sigma v(n+h' \rightarrow h \cdots) \rangle X_n \right. \\
 & + \langle \sigma v(p+h' \rightarrow h \cdots) \rangle (1-X_n)) \left(\frac{m_e}{z} \right)^3 \hat{n}_b \Big\} X_{h'} \Big], \quad (5.29)
 \end{aligned}$$

which is modified from Eq. (5.6) to adopt the new parametrization.

The boundary conditions of Eqs. (5.6) and (5.29) specified at $T = T_i$ are set by the quasistable solutions where $X'_n = 0$ and $X'_h = 0$. This choice is merely for numerical stability, and the final results are insensitive to the precise details of the boundary conditions. Both X_n and X_h rapidly settle down to the quasistable values after the evolution begins, erasing any dependence on the initial conditions.

In the standard BBN framework, we find $T_D = 73.7$ keV acts as the effective bottleneck temperature at which we reproduce $Y_p = 0.245$ [84], under the assumption of instantaneous neutron capture into deuterium. As discussed in Sec. II, we evaluate a more robust observable: the fractional deviation from standard BBN predictions due to new physics, and we impose $\delta X_n / X_n^{\text{SM}}|_{T=T_D} < R_{Y_p}$.

VI. UPDATES ON HADRONIC CROSS SECTIONS

In this section, we briefly discuss the methodology to obtain cross sections used in Eqs. (5.27), (5.28), and (5.29) and address the major updates. The $n \leftrightarrow p$ conversion processes are summarized in Table I. More details are explained in Appendix C.

As discussed in Sec. II, we need to know the corresponding cross sections often near the threshold. The long-lived hadrons, except for K_L , immediately slow down due to the electromagnetic interaction and become highly nonrelativistic. We use the partial wave analysis, with s - and p -waves, and fit the measured data, and then the threshold cross sections are extrapolated. The higher momentum contribution is important as well to perform thermal averaging.

However, challenges exist for many processes, often involving n or K_L , because there are no corresponding measurements due to the experimental difficulties. Sometimes, the data of the inverse process exist, for example, we can use $n\gamma \rightarrow p\pi^-$ data to obtain $p\pi^- \rightarrow n\gamma$ by time reversal, while the phase space factor has to be

corrected because the mass difference is crucial near the threshold. In addition, we use the isospin symmetry. For example, in the isospin limit, we can obtain the $n\pi^+ \rightarrow p\pi^0$ cross section by $p\pi^- \rightarrow n\pi^0$ measurements, but in reality we need to factor out the Coulomb attraction in addition to correcting the phase space, as it matters more at low momentum.

For processes initiated by a kaon, we essentially perform the same prescriptions: relating by isospin, manipulating the Coulomb effect, and correcting the phase space factor. Nevertheless, the situation is more complicated because the data are limited to the scattering processes with initial state of $pK^\pm$ or pK_L . Denoting the $(\bar{K}^0, K^-)$ isospin doublet by $\bar{K}$ and hyperons by Y , the amplitudes of processes $N\bar{K} \rightarrow N'\bar{K}'$ and $N\bar{K} \rightarrow Y\pi$ can be categorized by the total isospin, $I = 0$ and $I = 1$, following the formalism suggested in Ref. [118] (see Appendix C 3 for the details). Additionally, K_L scattering processes require additional amplitudes of $NK \rightarrow N'K'$, denoting $K = (K^+, K^0)$, such as the elastic scattering ($pK^+ \rightarrow pK^+$), the charge exchange ($pK_L \rightarrow nK^+$), and regeneration processes ($pK_L \rightarrow pK_S$, $nK_L \rightarrow nK_S$). We include all available experimental data, to the best of our knowledge, and find the relevant scattering amplitudes by simultaneous fitting. We limit the datasets to $k_{\text{lab}} < 2$ GeV, where k_{lab} is the kaon momentum in the target rest frame because more processes beyond our framework are relevant at high momenta. The restriction of the K_L spectrum in Figs. 6 and 7 is aligned with this limitation.

Our fitting, as shown in Fig. 11, gives various $2 \rightarrow 2$ scattering cross sections of pK_L (left) and nK_L (right). The known data for pK_L scattering are overlaid, which show a good agreement with our fitting. More results are given in Appendix C 3. We find p -wave contributions are significant, which makes the elastic scattering dominant for $k_{\text{lab}} \gtrsim 1$ GeV.

For our BBN analysis, the elastic scattering cross sections are important for obtaining the correct K_L spectrum. Although the elastic scattering is not directly involved in $n \leftrightarrow p$ conversion, it modifies the energy distribution of the K_L 's from axion decays. Accounting for the elastic scattering, we obtain the effective K_L distribution from the one given by the axion decay, as shown in Figs. 12 and 13. The reshape scheme is given in Appendix D.

Using the reshaped distributions, the averaged cross sections of NK_L as a function of m_a are evaluated, as in Fig. 14. Also, the effective K_L decay width is calculated by averaging this distribution as $\tilde{\Gamma}_{K_L} = \langle \gamma^{-1} \rangle \Gamma_{K_L}$, where $\gamma^{-1} = m_{K_L} / E_{K_L}$, which is about 0.6–0.8. For example, using PYTHIA (Herwig), $\langle \gamma^{-1} \rangle = 0.70$ (0.80), 0.69 (0.69), 0.63 (0.63) at $m_a = 3, 10$, and 50 GeV, respectively.

In the literature, evaluation of the K_L cross sections has been missing, which is partially why the K_L process was omitted or oversimplified. The threshold cross section was

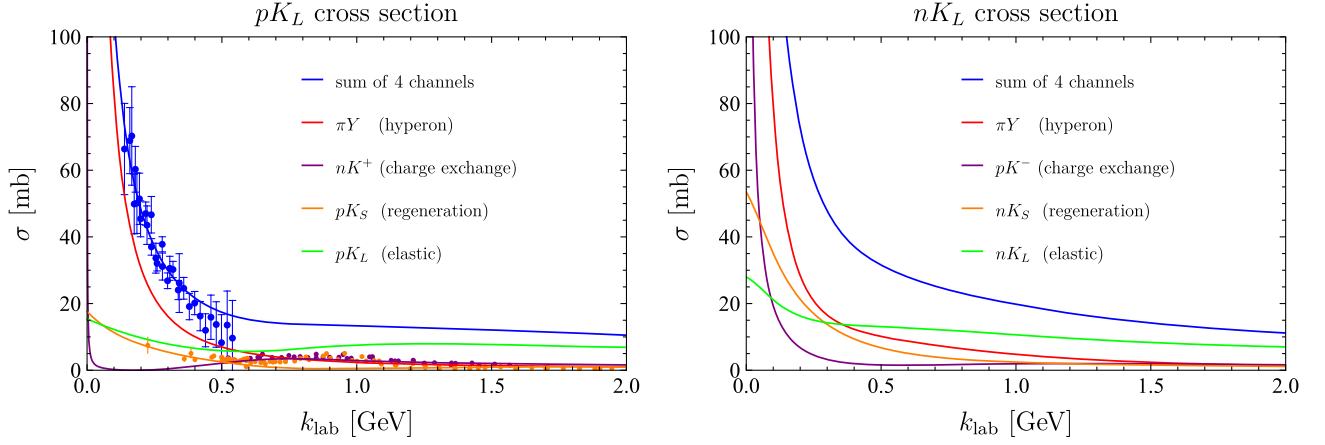


FIG. 11. Left: the cross sections of pK_L into four different final states obtained by the method described in Appendix C 3 and their sum: hyperon channel πY (red), charge exchange nK^+ (purple), regeneration pK_S (orange), elastic scattering pK_L (green), and the sum of four channels (blue). We show relevant datasets as dots with uncertainty bars: the total cross section from Fig. 4 of Ref. [119] (which includes data from Ref. [120]), the charge exchange cross section from Ref. [121], and the regeneration cross section from Fig. 27 of Ref. [122] (Ref. [123]). See Fig. 32 for the individual fitted processes. Right: nK_L scattering cross sections in the corresponding four different final states and their sum. These are the predictions of our fitting.

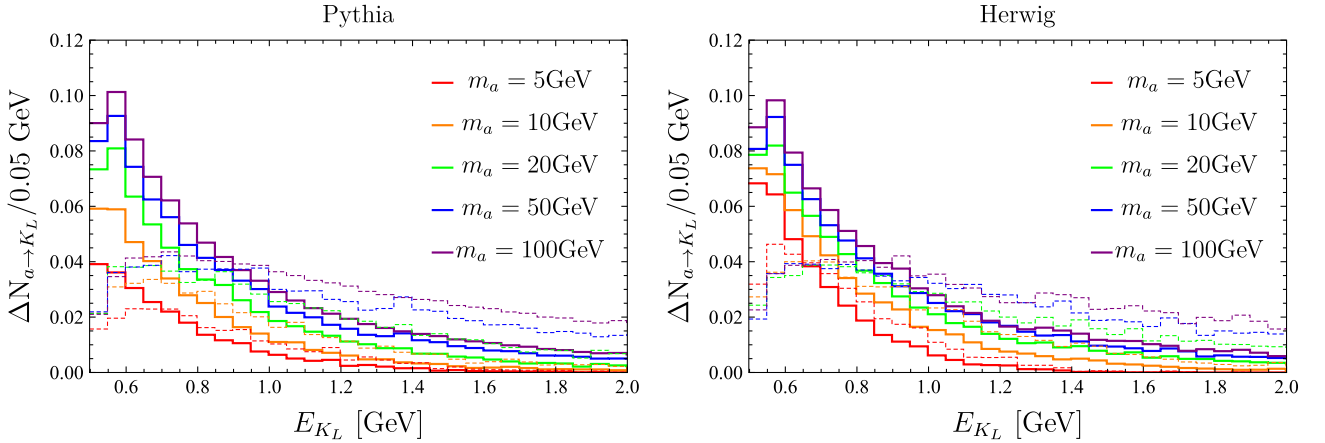


FIG. 12. Reshaped K_L energy spectrum due to elastic scattering. Energy distribution of K_L from the axion decay for $m_a = 5, 10, 20, 50$, and 100 GeV, obtained from PYTHIA (left) and Herwig (right). These spectra are still different from thermal distributions.

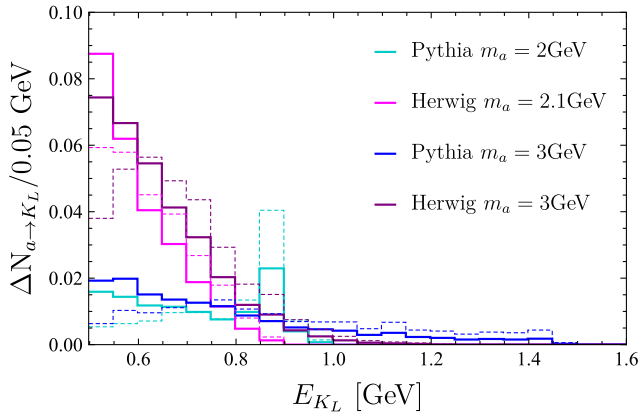


FIG. 13. Reshaped spectrum due to elastic scattering. Energy distribution of K_L from the axion decay for low axion masses, where large discrepancies between PYTHIA and Herwig predictions are present.

often used for nK_L scattering, but, as we noted, the high-momentum region is important. In fact, cross sections weighted with the K_L momentum profile per axion mass, which are shown in Fig. 14 (stars, crosses, or circles), are quite different from the threshold cross sections (solid lines).

Finally, let us comment on the annihilation of injected antibaryons with background baryons. We take the annihilation cross sections given in Ref. [124] and give details in Appendix C 4.

VII. RESULTS

A. Evolutions

Our numerical results on $\delta X_n / X_n^{\text{SM}}$ at T_D are obtained as described in Sec. V. We solve Eq. (5.18) with $\Gamma_{n \rightarrow p}$ and $\Gamma_{p \rightarrow n}$ given in Eqs. (5.19) and (5.20). The heavy axion

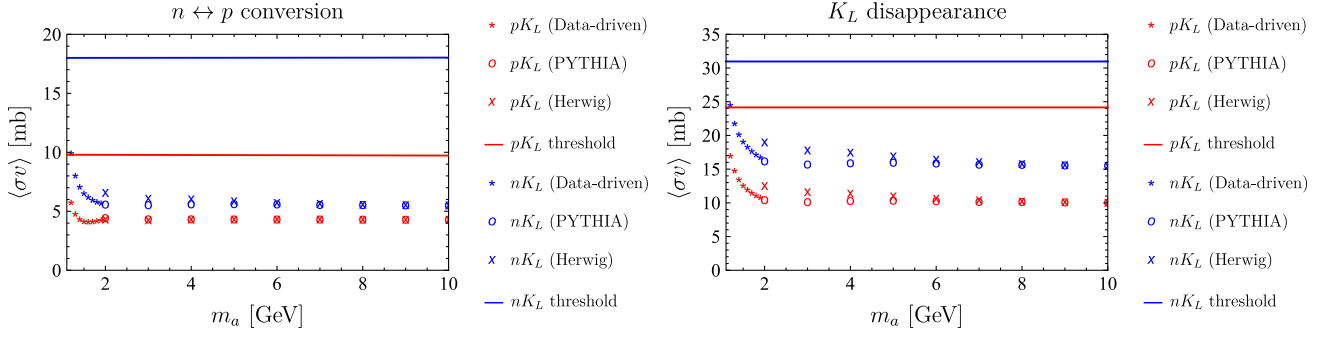


FIG. 14. Left: scattering cross sections (σv) with incident pK_L (red) or nK_L (blue) weighted by the reshaped K_L distribution from axion decay (Figs. 12 and 13). Only $n \leftrightarrow p$ conversion processes are considered. For $m_a \leq 2$ GeV, the data-driven method is used (star), and otherwise, PYTHIA (circle) or Herwig (cross) is used. The horizontal lines show the cross sections at the threshold for comparison. Right: the averaged cross sections for the K_L disappearance processes, namely, both conversion and nonconversion processes. The plot scheme is the same as in the left panel. The breakdown of the processes is found in Figs. 33 and 34.

contributions $\Delta\Gamma_{n \rightarrow p}$ and $\Delta\Gamma_{p \rightarrow n}$ follow Eqs. (5.27) and (5.28) with X_h obtained by solving Eq. (5.29).

Figure 15 shows how the evolution of X_n is modified in the presence of heavy axions (blue solid line) compared to the standard BBN (black solid line, denoted by X_n^{SM}). We take $m_a = 10$ GeV and $\tau_a = 0.0175$ s as a benchmark [note that $Y_a^{(\min)} = Y_a^{(\max)}$ in this case]. The left and right panels depict the evolutions of X_n and $\delta X_n/X_n^{\text{SM}}$, respectively. Various colored lines other than the blue one correspond to X_n obtained by turning on only a subset of hadronic components from the axion decay, while the secondary hadrons are always included. The orange, cyan, light green, dark green, and pink lines represent $\pi^\pm$, $K^\pm$, K_L , all the kaons ($K^\pm, K_L$), and baryons.

At $T \gtrsim 2$ MeV, in the gray-shaded region of Fig. 15, left, our curves are not completely correct because we have not

included some reactions that can be important at high temperatures. However, those missing reactions have positive threshold energies greater than several MeV, so they are suppressed at low temperatures. Thus, all errors from missing those reactions are washed out below $T = 2$ MeV.

At high temperatures, injected hadrons are so abundant that the reaction rates of $n \rightarrow p$ and $p \rightarrow n$ are governed by the hadronic cross sections. The solution of X_n in this regime corresponds to the one with $\dot{X}_n \simeq 0$ in (5.7). We call it the *quasistable* solution, where “quasi” indicates that this stable point of X_n changes in time, as the hadronic cross sections are temperature dependent.

The quasistable behavior is maintained until hadronic injections get exponentially suppressed, as most of the axions have decayed. Afterward, a rapid relaxation occurs, which we call *waterfall* (see around $T \simeq 1.5$ MeV in the

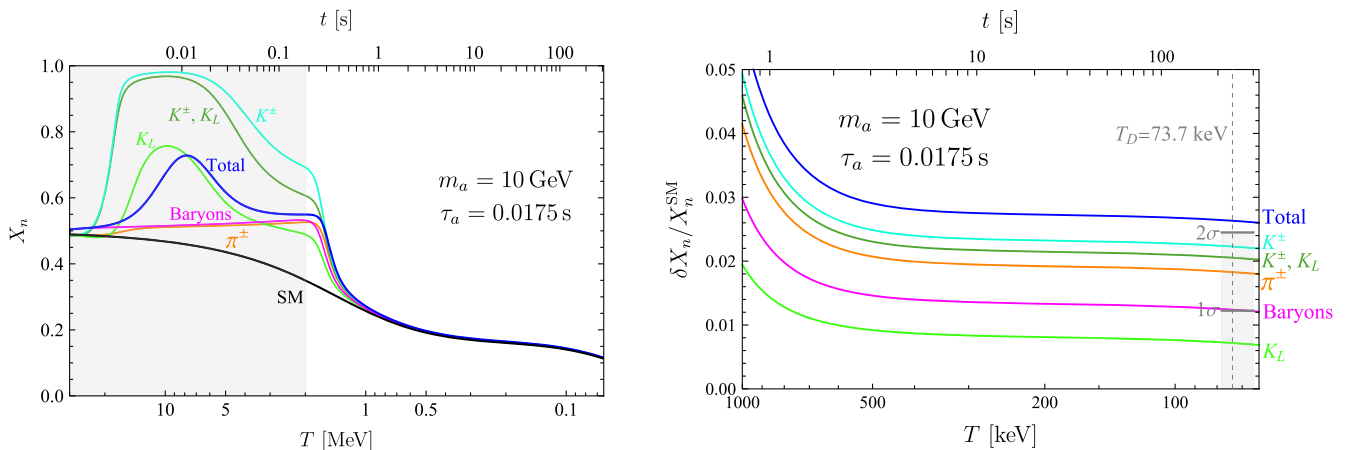


FIG. 15. Modification of the X_n evolution in the presence of a heavy axion (blue solid line) is compared to the SM scenario (black solid line). Here, we take $m_a = 10$ GeV and $\tau_a = 0.0175$ s as a benchmark, while $Y_a^{(\min)} = Y_a^{(\max)}$. Various colored lines other than blue correspond to modified X_n with turning on only a subset of hadronic components from the axion decay. Orange, cyan, light green, dark green, and pink lines represent cases of $\pi^\pm$, $K^\pm$, K_L , all the kaons ($K^\pm, K_L$), and baryons. The left and right panels depict the evolutions of X_n and $\delta X_n/X_n^{\text{SM}}$, respectively. For convenience, the corresponding time is shown on the top axes, assuming SM radiation dominance.

left panel of Fig. 15). Analytical and detailed estimations of the quasistable and waterfall behaviors will be provided in the next subsection. In a nutshell, our final constraint is determined dominantly by how much the waterfall drives the relaxation from the quasistable value. As shown in the right panel, the benchmark results in $\delta X_n/X_n^{\text{SM}}|_{T=T_D}$, slightly greater than 2σ of the Y_p measurement (1σ and 2σ bounds are indicated by the gray bars on the right panel at $T_D \simeq 73.7$ keV), so this parameter is already ruled out. Note that $\delta X_n/X_n^{\text{SM}}$ is quite insensitive to T_D , so we ignore the uncertainty in the value of T_D .

From Fig. 15, right, we can also see the importance of individual channels. As we chose $m_a = 10$ GeV, all the hadronic channels are open. We observe that the pions and kaons are equally important. In terms of $N_{a \rightarrow h}$, pions have an $\mathcal{O}(10)$ greater number than kaons, but their cross sections are $\mathcal{O}(0.1)$ smaller, so they are compensated. If the axion mass is lighter than about 1.1 GeV, which forbids $a \rightarrow KK\pi$ decay kinematically, then only the pion channel remains.

Pions and baryons have their quasistable values close to $1/2$ because their approximately isospin symmetric cross sections make the rates of $n \rightarrow p$ and $p \rightarrow n$ roughly the same. For instance, we have $\sigma(p\pi^- \rightarrow n\pi^0, n\gamma) \simeq \sigma(n\pi^+ \rightarrow p\pi^0, p\gamma)$, $\sigma_{\text{ann}}(p\bar{n}) = \sigma_{\text{ann}}(n\bar{p})$, and $\sigma_{\text{ann}}(p\bar{p}) \simeq \sigma_{\text{ann}}(n\bar{n})$. On the other hand, kaons severely violate such symmetry between $n \rightarrow p$ and $p \rightarrow n$, which leads to random quasistable values. This is because (i) the rapid decay of K_S makes the isospin doublet formation of K^- (or K^+) incomplete, and (ii) the K^- and K^+ are asymmetric in the $n \leftrightarrow p$ conversion, e.g., K^- produces hyperons while K^+ cannot due to strangeness conservation.

If the axion lifetime τ_a is taken larger, the situation changes, as in Fig. 16. Here, we take $\tau_a = 0.1$ s (top), 1 s (middle), and 10 s (bottom), while we reduce Y_a for the last case by multiplying a factor of 10^{-7} to make the curves distinguishable. All these parameters are ruled out. At $\tau_a = 0.1$ s, the waterfall time t_{wf} shifts later proportionally to τ_a , and then neutrons freeze out before a sufficient relaxation to the SM line occurs. In the other cases, τ_a is simply too large to have relaxation.

It is noteworthy to comment on the case of $\tau_a = 10$ s (lower panel), where the baryon channel dominates. There, the mesons produced by the axion decay quickly because the constant decay terms are bigger than the conversion rates, which decrease by T^3 . On the other hand, the baryon channel remains since the produced antinucleon $\bar{N}$ is stable and annihilates with an existing neutron or proton.

B. Parametric dependences and uncertainty estimation

Before presenting our final results, let us provide an analytic understanding of the evolution. This allows us to assess the uncertainty of our final constraint in the τ_a space. As will be shown in this section, an $\mathcal{O}(1)$ uncertainty in

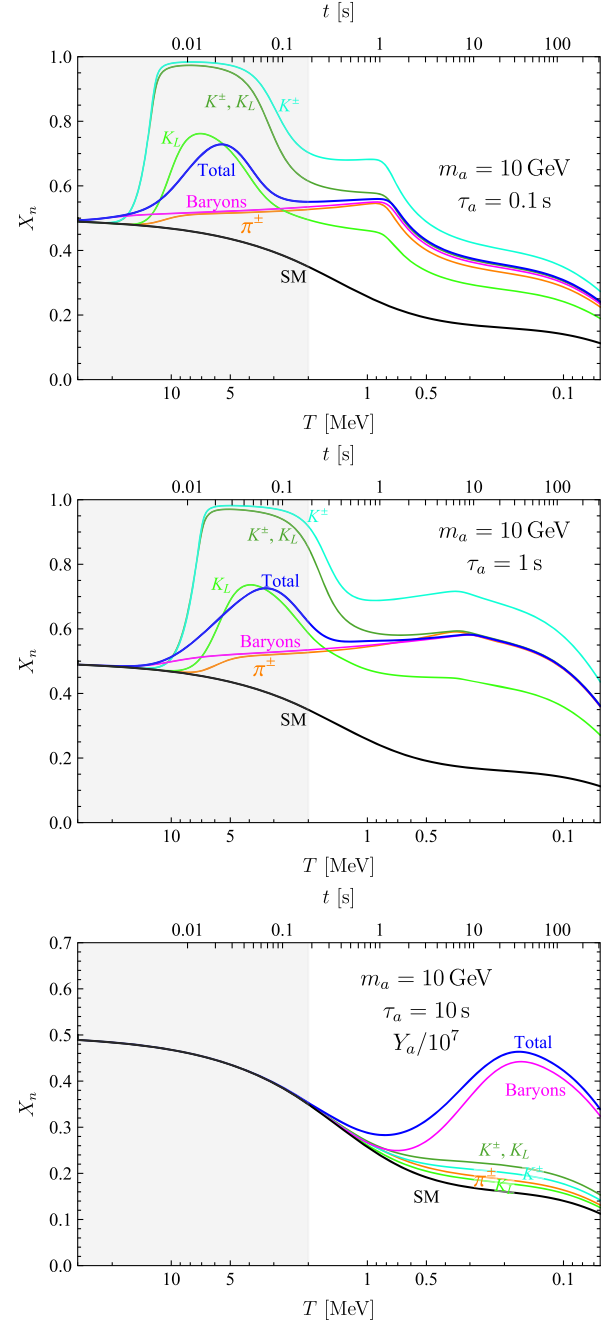


FIG. 16. The evolutions of X_n are depicted for different $\tau_a = 0.1$ s (upper), 1 s (middle), and 10 s (lower). The color scheme remains the same as that of Fig. 15.

$N_{a \rightarrow h}$ and the hadronic cross sections result in only a few % level uncertainty in the upper bound of τ_a . This ensures the reliability and robustness of our final result. Moreover, one can use our equations presented in this section to quickly estimate constraints on lifetimes of other long-lived particles that have different $N_{a \rightarrow h}$.

In the following discussion, we assume that the final result is primarily determined by a specific set of hadrons, e.g., K_L , $\pi^\pm$, or baryons, so we focus on a single

species. We do not include the effect of the secondary production of other hadrons via scattering/decay for simplicity. Furthermore, we introduce notations that shorten many expressions for various cross sections,

$$\langle\langle nh \rangle\rangle \equiv \langle\sigma v(n + h \rightarrow \dots)\rangle, \quad (7.1)$$

$$\langle\langle nh \rangle\rangle_c \equiv \langle\sigma v(n + h \rightarrow p + \dots)\rangle, \quad (7.2)$$

where $\langle\langle nh \rangle\rangle$ represents the averaged total cross section of n 's consuming h while $\langle\langle nh \rangle\rangle_c$ includes only processes that convert n to p . We also define $\langle\langle ph \rangle\rangle$ and $\langle\langle ph \rangle\rangle_c$ in the same way by interchanging n and p .

The hadron number density from the axion decay can be obtained by requiring $\dot{n}_h \simeq 0$ in Eq. (5.6), yielding

$$n_h \simeq \frac{N_{a \rightarrow h} \Gamma_a n_a}{\tilde{\Gamma}_h + \langle\langle nh \rangle\rangle n_n + \langle\langle ph \rangle\rangle n_p} \quad (7.3)$$

with $X_p = n_p/n_b = 1 - X_n$. Here, we ignore the Hubble constant since $\tilde{\Gamma}_h + \langle\langle nh \rangle\rangle n_n + \langle\langle ph \rangle\rangle n_p \gg H$.

The neutron ratio in the quasistable regime is given by the ratio of conversion rates, requiring $\dot{X}_n \simeq 0$ in Eq. (5.7),

$$X_n^{\text{qs}} = \frac{\Gamma_{p \rightarrow n}}{\Gamma_{n \rightarrow p} + \Gamma_{p \rightarrow n}} \quad (7.4)$$

$$= \frac{\Gamma_{p \rightarrow n}^{\text{weak}} + \langle\langle ph \rangle\rangle_c n_h}{\Gamma_{p \rightarrow n}^{\text{weak}} + \Gamma_{n \rightarrow p}^{\text{weak}} + (\langle\langle ph \rangle\rangle_c + \langle\langle nh \rangle\rangle_c) n_h}, \quad (7.5)$$

where $\Gamma_{n \rightarrow p}^{\text{weak}} \equiv \Gamma_{n \nu_e \rightarrow p e^-} + \Gamma_{n e^+ \rightarrow p \bar{\nu}_e} + \Gamma_{n \text{ decay}}$ and $\Gamma_{p \rightarrow n}^{\text{weak}} \equiv \Gamma_{p \bar{\nu}_e \rightarrow n e^+} + \Gamma_{p e^- \rightarrow n \nu_e}$. The quasistable value, X_n^{qs} , is determined by the cross section ratio independently of n_h ,

$$X_n^{\text{qs}} \simeq \frac{\langle\langle ph \rangle\rangle_c}{\langle\langle ph \rangle\rangle_c + \langle\langle nh \rangle\rangle_c}, \quad (7.6)$$

when the axion-induced process dominates.⁷ This explains why $X_n^{\text{qs}} \simeq 0.5$ for $h = \pi^\pm$ or baryons (see Figs. 15 and 16) because their cross sections are nearly symmetric between n and p due to isospin symmetry. On the other hand, in the kaon cases, since the isospin symmetry is not respected, X_n^{qs} can take any values.

The solution in Eq. (7.6) should be interfered at the waterfall time t_{wf} when the weak interaction rate catches up with the axion-induced conversion rate as the hadron number density gets exponentially small. We estimate t_{wf} by solving $\Gamma_{n \rightarrow p}^{\text{weak}} + \Gamma_{p \rightarrow n}^{\text{weak}} = (\langle\langle nh \rangle\rangle_c + \langle\langle ph \rangle\rangle_c) n_h$ with

⁷In the presence of multiple hadrons, the quasistable value becomes

$$X_n^{\text{qs}} \simeq \frac{\sum_h \langle\langle ph \rangle\rangle_c n_h}{\sum_h (\langle\langle ph \rangle\rangle_c + \langle\langle nh \rangle\rangle_c) n_h}.$$

As seen from this formula, contributions from multiple hadrons could compensate rather than adding up. For example, X_n^{qs} from both K_L and $K^\pm$ is always located between the one from K_L and the one from $K^\pm$.

Eq. (7.3) and $n_a \simeq Y_a s e^{-t/\tau_a}$, and obtain

$$t_{\text{wf}} = \tau_a \ln \left[\frac{Y_a \Gamma_a s}{(\Gamma_{n \rightarrow p}^{\text{weak}} + \Gamma_{p \rightarrow n}^{\text{weak}})_{t=t_{\text{wf}}}} \frac{N_{a \rightarrow h} (\langle\langle nh \rangle\rangle_c + \langle\langle ph \rangle\rangle_c)}{(\tilde{\Gamma}_h + \langle\langle nh \rangle\rangle n_n + \langle\langle ph \rangle\rangle n_p)} \right]. \quad (7.7)$$

A rough evaluation with typical parameters, such as $Y_a \sim 10^{-3}$, $n_b/s \simeq 8.7 \times 10^{-11}$, and $N_{a \rightarrow h} \sim 1$, gives the logarithm from 15 to 18.⁸

After the exit of the quasistable regime, the weak interaction dominates, and the axion-induced effect can be treated as a perturbation from the SM scenario. The analytic solution of $\delta X_n \equiv X_n - X_n^{\text{SM}}$ can be written as

$$\begin{aligned} \delta X_n(t) &= (X_n(t_{\text{wf}}) - X_n^{\text{SM}}(t_{\text{wf}})) G(t) \\ &+ G(t) \int_{t_{\text{wf}}}^t (\langle\langle ph \rangle\rangle_c X_p^{\text{SM}}(t') - \langle\langle nh \rangle\rangle_c X_n^{\text{SM}}(t')) \\ &\times n_h(t') G(t') dt', \end{aligned} \quad (7.8)$$

where

$$G(t) = \exp \left[- \int_{t_{\text{wf}}}^t (\Gamma_{p \rightarrow n} + \Gamma_{n \rightarrow p}) dt' \right]. \quad (7.9)$$

As we are interested in the solution at $t > t_{\text{wf}}$, we neglect the second line in Eq. (7.8) since $n_h \propto e^{-t/\tau_a}$ is rapidly decreasing with $t/\tau_a \gtrsim 15$. By taking a junction from the quasistable solution in Eq. (7.6) to this at t_{wf} , i.e., $X_n(t_{\text{wf}}) \simeq X_n^{\text{qs}}(t_{\text{wf}})$, we obtain

$$\delta X_n(t) \simeq (X_n^{\text{qs}}(t_{\text{wf}}) - X_n^{\text{SM}}(t_{\text{wf}})) G(t). \quad (7.10)$$

Thus, the waterfall solution is nothing but the relaxation from a deviation set by the axion-induced effect.

δX_n gets suppressed until the relaxation rate becomes smaller than the Hubble rate, and then it becomes frozen. Since the relaxation rate is approximately given by the sum of the weak interaction rates, we evaluate δX_n at $t = t_n \simeq 0.73 \text{ s}$ ($T = T_n = 1.0 \text{ MeV}$), at which $\Gamma_{p \rightarrow n}^{\text{weak}} + \Gamma_{n \rightarrow p}^{\text{weak}} = 3H$, as its freeze-out value. More precisely speaking, we evaluate the ratio $\delta X_n/X_n^{\text{SM}}$ since δX_n does not get completely frozen because of the neutron decay.

With $\Gamma_{p \rightarrow n}^{\text{weak}} + \Gamma_{n \rightarrow p}^{\text{weak}} \propto T^5$, the exponent of $G(t_n)$ can be approximated as

⁸Unlike X_n^{qs} , the multiple hadron contributions constructively add up for t_{wf} as

$$t_{\text{wf}} = \tau_a \ln \left[\frac{Y_a \Gamma_a s}{(\Gamma_{n \rightarrow p}^{\text{weak}} + \Gamma_{p \rightarrow n}^{\text{weak}})_{t=t_{\text{wf}}}} \sum_h \frac{N_{a \rightarrow h} (\langle\langle nh \rangle\rangle_c + \langle\langle ph \rangle\rangle_c)}{(\tilde{\Gamma}_h + \langle\langle nh \rangle\rangle n_n + \langle\langle ph \rangle\rangle n_p)} \right].$$

$$-\int_{t_{\text{wf}}}^{t_n} (\Gamma_{p \rightarrow n}^{\text{weak}} + \Gamma_{n \rightarrow p}^{\text{weak}}) dt' \simeq -\left(\left(\frac{t_n}{t_{\text{wf}}}\right)^{\frac{3}{2}} - 1\right), \quad (7.11)$$

where we take $\Gamma^{\text{weak}}(T) \simeq \Gamma^{\text{weak}}(T_n)(T/T_n)^5$, $H(T) \simeq H(T_n)(T/T_n)^2$, and $t \simeq 1/2H$. The bound we adopt, $\delta X_n/X_n^{\text{SM}}|_{T=T_D} \simeq \delta X_n/X_n^{\text{SM}}|_{T=T_n} < 2.45\% \equiv R_{Y_p}$, can now be rearranged as

$$t_{\text{wf}} \lesssim t_n \left(1 + \ln \left[\frac{1}{R_{Y_p}} \frac{\delta X_n(t_{\text{wf}})}{X_n^{\text{SM}}(t_n)} \right]\right)^{-\frac{2}{3}}, \quad (7.12)$$

where the logarithm is roughly 3.7 as $\delta X_n(t_{\text{wf}})/X_n^{\text{SM}} \sim \mathcal{O}(1)$.

Finally, replacing t_{wf} on the left-hand side by Eq. (7.7) leads to the bound on the axion lifetime,

$$\tau_a \lesssim \frac{t_n}{\left(1 + \ln \left[\frac{1}{R_{Y_p}} \frac{\delta X_n(t_{\text{wf}})}{X_n^{\text{SM}}(t_n)} \right]\right)^{\frac{2}{3}} \ln \left[\frac{Y_a \Gamma_a S}{\Gamma_{n \rightarrow p}^{\text{weak}} + \Gamma_{p \rightarrow n}^{\text{weak}}} \frac{N_{a \rightarrow h} \langle \langle nh \rangle \rangle_c + \langle \langle ph \rangle \rangle_c}{\Gamma_h + \langle \langle nh \rangle \rangle_n + \langle \langle ph \rangle \rangle_p} \right]}. \quad (7.13)$$

This formula works quite well. Putting the typical numerical values mentioned below Eq. (7.7), we obtain $\tau_a \lesssim 0.014\text{--}0.018$ s from Eq. (7.13), which agrees well with our full numerical result. Note that $\delta X_n(t_{\text{wf}})$ in the first logarithmic factor depends on cross sections as they appear in Eq. (7.6) and $\delta X_n(t_{\text{wf}}) \simeq X_n^{\text{qs}}(t_{\text{wf}}) - X_n^{\text{SM}}(t_{\text{wf}})$.

Now, we can estimate the uncertainties of our final constraint. In our calculation, Y_a , $N_{a \rightarrow h}$, and hadronic cross sections have the largest uncertainties, so let us estimate their impact on our final constraint in terms of τ_a by using Eq. (7.13).

Suppose the hadron yield from the axion decay is modified by the change of $Y_a N_{a \rightarrow h} \rightarrow Y'_a N'_{a \rightarrow h}$. The shift of t_{wf} from this change can be obtained from Eq. (7.7) as

$$t'_{\text{wf}} - t_{\text{wf}} = \tau_a \ln \frac{Y'_a N'_{a \rightarrow h}}{Y_a N_{a \rightarrow h}}, \quad (7.14)$$

and consequently, the neutron fraction is modified as

$$\frac{\delta X'_n}{\delta X_n} \Big|_{t=t_D} \simeq \exp \left[- \int_{t'_{\text{wf}}}^{t_{\text{wf}}} (\Gamma_{p \rightarrow n}^{\text{weak}} + \Gamma_{n \rightarrow p}^{\text{weak}}) dt \right] \quad (7.15)$$

$$\sim \left(\frac{Y'_a N'_{a \rightarrow h}}{Y_a N_{a \rightarrow h}} \right)^{(\Gamma_{p \rightarrow n}^{\text{weak}}(t_{\text{wf}}) + \Gamma_{n \rightarrow p}^{\text{weak}}(t_{\text{wf}})) \tau_a}, \quad (7.16)$$

where the exponent is roughly

$$(\Gamma_{p \rightarrow n}^{\text{weak}}(t_{\text{wf}}) + \Gamma_{n \rightarrow p}^{\text{weak}}(t_{\text{wf}})) \tau_a \sim 0.3 \left(\frac{\tau_a}{0.0175 \text{ s}} \right)^{-\frac{3}{2}}. \quad (7.17)$$

Clearly, it is not linearly scaling, unlike the naive expectation. We checked and confirmed this scaling behavior with our full numerical code.

Conservatively, we estimate the uncertainty of $N_{a \rightarrow h}$ as $\sim 100\%$ since PYTHIA and Herwig can have a difference of at most 100% depending on m_a and channels (except for the baryonic channel in the region very close to 2 GeV). If we multiply a factor of 2 to every $N_{a \rightarrow h}$, this only changes $\delta X_n|_{t=t_D}$ by 23%, as indicated in Eqs. (7.16) and (7.17). Then, the upper bound of τ_a is only affected by 4%, as Eq. (7.13) shows.

Our numerical check of this feature is depicted in the left panel of Fig. 17, where the blue line is the evolution of $\delta X_n/X_n^{\text{SM}}$ including all the hadrons, while the red line is obtained by multiplying a factor of 2 to $N_{a \rightarrow h}$. As expected, the red line is increased by about 26% from the blue line. Then, δX_n with roughly 4% smaller τ_a is depicted by the green line, which comes close to the blue line.

On the other hand, the difference between $Y_a^{(\min)}$ and $Y_a^{(\max)}$ is negligible in most of the parameter space, but it can be large in a particular region, $m_a \lesssim 0.7$ GeV. The difference can be as large as a factor of about 10 around $m_a \sim 0.7$ GeV. However,

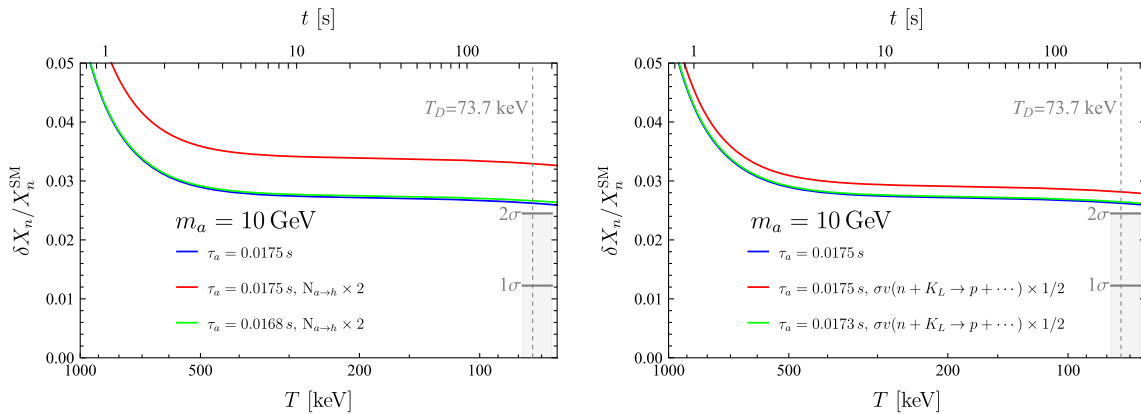


FIG. 17. $\delta X_n/X_n^{\text{SM}}$ including all the hadrons with $m_a = 10$ GeV and $\tau_a = 0.0175$ s is depicted by the blue line, while the red lines correspond to that with artificial changes of $N_{a \rightarrow h}$ (left) and $\sigma v(n + K_L \rightarrow p + \dots)$ (right) by a factor of 2 to see parametric dependence. As the green lines show, a few percent modification of τ_a makes the evolution back to the blue line.

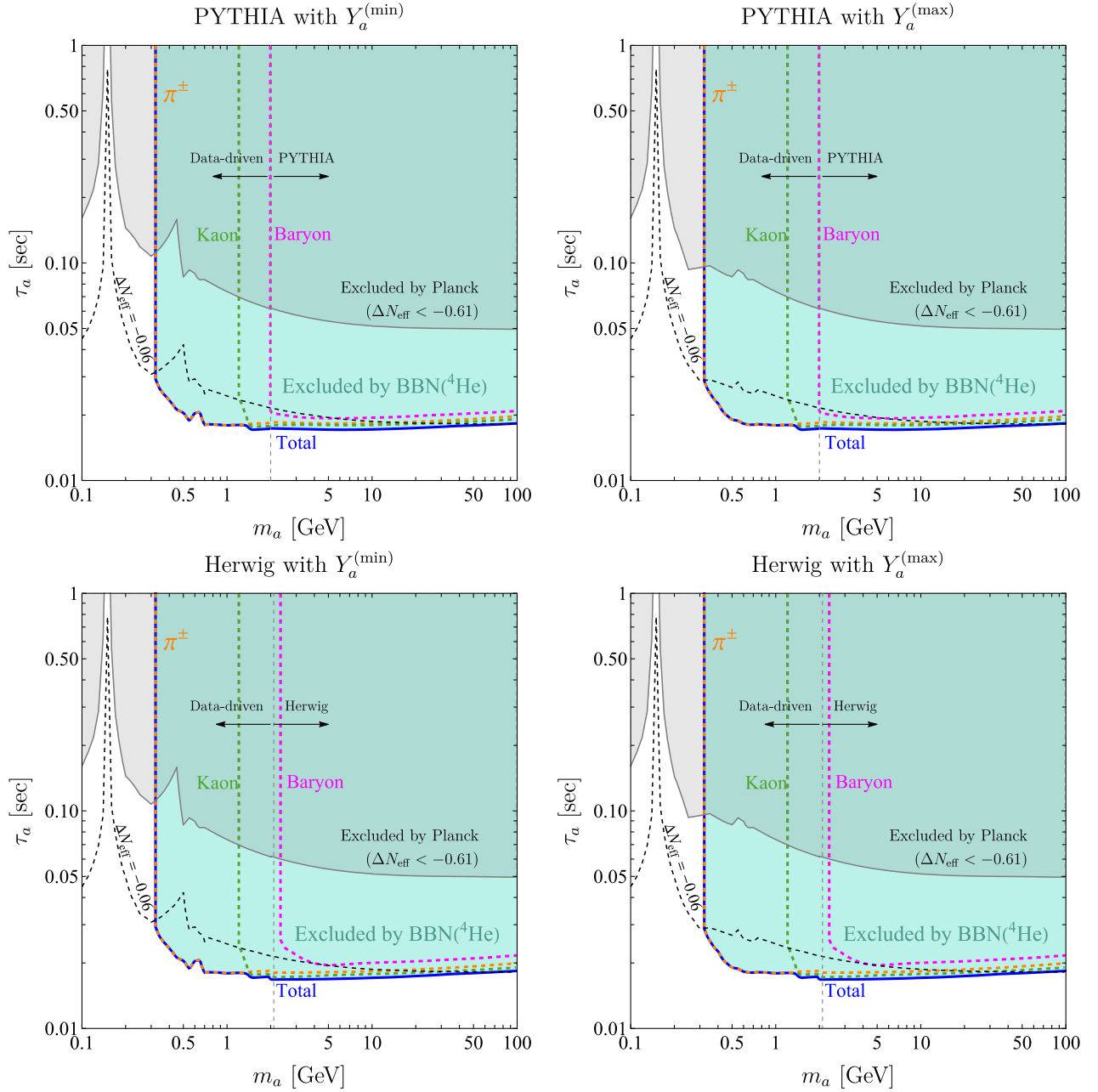


FIG. 18. The exclusion by our BBN calculation is given by the cyan-shaded region surrounded by the blue contour in the (m_a, τ_a) space. Left and right panels take $Y_a^{(\min)}$ and $Y_a^{(\max)}$ and upper and lower panels use PYTHIA and Herwig, respectively. The orange, dark green, and pink dashed contours represent constraints estimated from individual channels of $\pi^\pm$, kaons, and baryons. The gray vertical dashed line at $m_a \simeq 2$ GeV indicates the boundary of changing our scheme of estimating the axion decay rate and branching ratios. We use the data-driven method for $m_a < 2$ GeV and PYTHIA or Herwig for $m_a > 2$ GeV.

as the Y_a dependence is still in the logarithm in Eq. (7.13), the upper bound of τ_a changes only about 10%. This change will be shown in the next subsection.

Analyzing the impact of uncertainties in hadronic cross sections is slightly more complicated, as they contribute to both logarithms in Eq. (7.13). As implied by Eqs. (7.6) and (7.7), a multiplication of an overall factor to all hadronic cross sections has almost zero impact, ignoring the decay width $\tilde{\Gamma}_h$, which we confirmed numerically. What matters

in the end is the ratio of $p \rightarrow n$ and $n \rightarrow p$ cross sections. Once the ratio is changed, it modifies both the quasistable value and the waterfall timing.

For the kaon cross sections, uncertainties are dominated by those from K_L , and we consider, conservatively, a factor of 2 uncertainty on the K_L cross sections. For example, in the right panel of Fig. 17, we show the evolution of $\delta X_n/X_n^{\text{SM}}$ with (red) and without (blue) multiplying the cross sections of K_L -induced $n \rightarrow p$ conversion by a factor

of $1/2$. This changes $\delta X_n/X_n^{\text{SM}}$ by 7%. We can bring the modified curve back if τ_a is increased by 1%.

Therefore, we conclude that uncertainty in our upper bound on τ_a is on the order of a few percent.

C. Bounds on the axion lifetime and decay constant

In our calculation, we have two possibilities of the axion abundance, $Y_a^{(\min)}$ and $Y_a^{(\max)}$, and two programs for the

parton shower and hadronization (PYTHIA and Herwig). Moreover, we have two ways of presenting our constraint: τ_a vs f_a . Therefore, there are eight combinations in presenting our results (see Figs. 18 and 19).

Figure 18 shows our results in (m_a, τ_a) space, where left and right panels take $Y_a^{(\min)}$ and $Y_a^{(\max)}$ and upper and lower panels use PYTHIA and Herwig, respectively. The cyan-shaded region surrounded by the blue contour is the region excluded

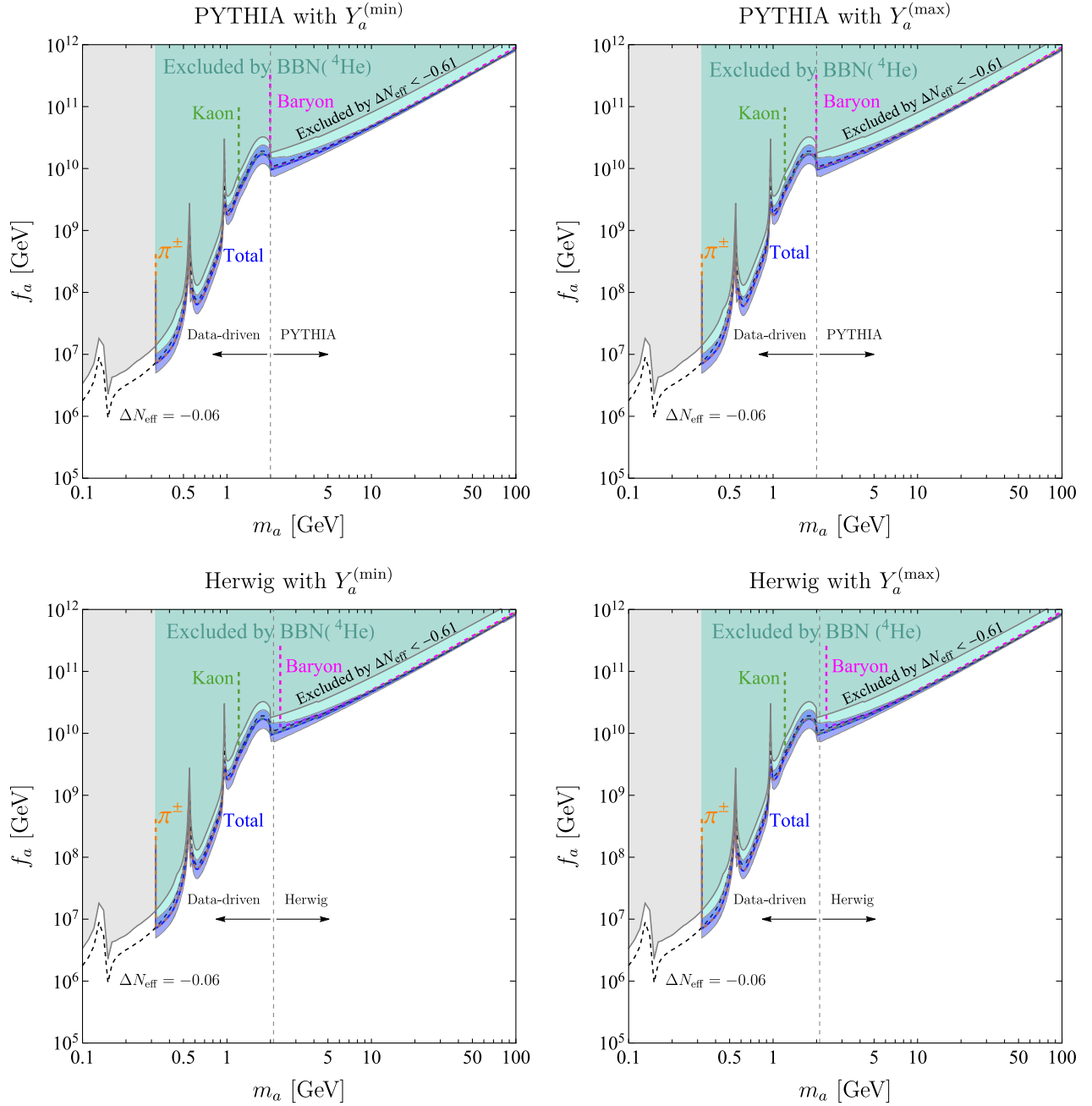


FIG. 19. The exclusion by our BBN calculation is now given in the (m_a, f_a) space, where the color scheme is unchanged from Fig. 18. Here, we also demonstrate the large uncertainty coming from translating τ_a into f_a by the blue band (we omit the uncertainty bands for the CMB constraint contours to avoid overcrowding).

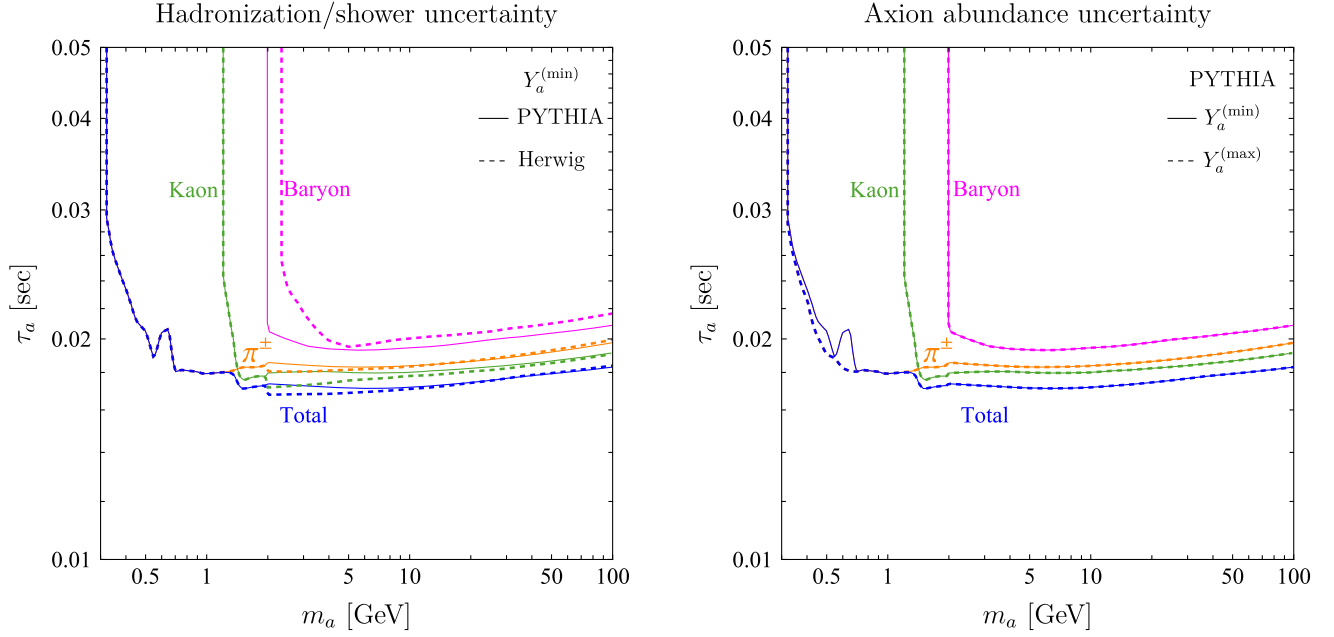


FIG. 20. Comparison of our final constraints among different sets of Y_a and parton-showering/hadronization programs. In the left panel, we fix the axion yield by $Y_a^{(\min)}$, and compare PYTHIA (solid lines) vs Herwig (dashed lines). In the right panel, we use PYTHIA, and compare $Y_a^{(\min)}$ (solid) vs $Y_a^{(\max)}$ (dashed). The color scheme is unchanged from Fig. 18.

by our BBN constraint (^4He), while the orange, dark green, and pink dashed contours represent constraints estimated from individual channels of $\pi^\pm$, kaons, and baryons (with secondary hadrons turned on). The gray vertical dashed line at $m_a \simeq 2$ GeV indicates the boundary of changing our scheme to estimate the axion decay rate and branching ratios. We use the data-driven method for $m_a < 2$ GeV and PYTHIA or Herwig for $m_a > 2$ GeV.

Our upper bound on τ_a is almost flat, although $N_{a \rightarrow h}$ increases as m_a increases (e.g., see Fig. 6). The bound becomes even weaker at higher mass, although this dependence is tiny. This counterintuitive feature is because a heavier axion actually induces a longer axion-dominated period, decreasing N_{eff} and the Hubble rate. It consequently delays the neutron freeze-out, leading to a longer relaxation time, and decreases X_n . This effect competes with the enhancement of X_n due to larger $N_{a \rightarrow h}$.

The gray-shaded region is excluded by the N_{eff} bound from the CMB fitting at Planck collaboration [93]. Our N_{eff} bound is slightly different from that of Ref. [69] because (i) we use a weaker criterion $N_{\text{eff}} > 2.43$, as discussed in Sec. VB, and (ii) we take the different total decay width, as described in Sec. IV,⁹ that causes modification of Y_a in the τ_a space (the axion disappearance rate relevant for Y_a is evaluated in terms

of f_a , not τ_a). The black dashed line ($\Delta N_{\text{eff}} = -0.06$) represents the potential sensitivity of the N_{eff} bound expected at the CMB-S4 experiment with about 1% precision [125]. As shown in the plot, our BBN constraint, if present, is stronger than the projected CMB bound.

The N_{eff} bound is essential for constraining the lifetimes in the low-mass region where the hadronic decay is forbidden. Potentially, there would be other BBN constraints, such as the one from the primordial deuterium measurement, which is sensitive to the late-time photo-dissociation. However, we do not consider those bounds because the N_{eff} bound is substantially stronger.

These constraints are then transferred to the (m_a, f_a) space in Fig. 19—again, left and right panels show the results with $Y_a^{(\min)}$ and $Y_a^{(\max)}$ and upper and lower panels are obtained by using PYTHIA and Herwig, respectively. As we discussed in Sec. IV with Fig. 5, there is a large uncertainty in the axion total decay width estimation in terms of m_a and f_a . We depict this uncertainty by the blue band for our BBN constraint, while we do not present similar uncertainty bands for the N_{eff} constraint and sensitivity contour, as they would make the plot unreadable.

In the left panel of Fig. 20, we depict the results with PYTHIA (solid lines) and Herwig (dashed line) with $Y_a^{(\min)}$ while the color scheme remains unchanged. As we discussed in the earlier subsection, their impact on τ_a is small except for the baryon channel around $m_a \sim 2$ GeV, where a large discrepancy appears, as the thresholds for the baryonic channel do not match due to their different

⁹We take the result of Ref. [75] for $m_a < 2$ GeV and the NNLO expression [78] for $m_a > 2$ GeV, while Ref. [69] used the result of Ref. [60] for $m_a < 2$ GeV and the NLO expression for $m_a > 2$ GeV.

hadronization algorithms, as shown in Fig. 6. Although the discrepancy in the baryon channel is somewhat large, our final result is dominated by the charged pions and kaons, which are much more stable.

In the right panel of Fig. 20, we compare the results using $Y_a^{(\min)}$ (solid) or $Y_a^{(\max)}$ (dashed), where PYTHIA is used. For $\tau_a \sim 0.02$ s, the difference in Y_a only appears at $m_a \lesssim 0.7$ GeV (see the bottom panels in Fig. 4). The largest discrepancy appears around $0.5 \text{ GeV} \lesssim m_a \lesssim 0.7 \text{ GeV}$, where $Y_a^{(\max)}/Y_a^{(\min)}$ can be as large as 10. As we argued in the previous subsection, however, this leads to only 20% discrepancy in τ_a .

VIII. CONCLUSION

In this work, we have estimated the BBN constraint on the heavy QCD axion that decays hadronically. The constraint is derived by computing the axion-induced modification to the neutron-to-proton ratio, which directly determines the primordial ^4He abundance. The axion yield is evaluated by its freeze-out value, assuming a large reheating temperature, and its hadronic branching ratios are obtained by using the data-driven method at $m_a < 2 \text{ GeV}$ and PYTHIA or Herwig at $m_a > 2 \text{ GeV}$. With these input quantities, we solve the Boltzmann equation for X_n , and obtain the constraint in the (m_a, τ_a) space. Our constraint is also depicted in the (m_a, f_a) space, although there is an $\mathcal{O}(1)$ uncertainty in the conversion of τ_a and f_a .

Our analysis incorporates several key updates on hadronic injection scenarios during BBN that can be applied to other models. We especially include K_L contributions using isospin relations to obtain their cross sections and account for their momentum distributions, as K_L does not get thermalized kinetically. We also trace secondary hadrons produced from decays and scatterings, ensuring a more consistent treatment. Our methodology can also be applied to a broader class of long-lived particles that decay into hadrons.

Based on these improvements, we have derived a robust upper bound on the axion lifetime, $\tau_a \lesssim 0.02$ s, across a wide range of axion masses above 0.3 GeV . We find that our BBN constraint is more stringent than existing and even projected CMB constraints via N_{eff} . This highlights the importance of the BBN analysis for hadronically decaying long-lived particles.

Our bound is shown to be quite insensitive to the branching fractions, hadronic cross sections, and the initial axion yield, as the dependence of the τ_a bound on these parameters appears in a logarithm with a small coefficient. As discussed thoroughly in Sec. VII B, even 100% modification of the initial axion yield or the branching fractions due to different models would lead to only a 4% shift in the lifetime bound. Hence, our constraint is robust and nearly model independent.

ACKNOWLEDGMENTS

We thank Sabyasachi Chakraborty, Sean Dobbs, David Dunskey, Keisuke Harigaya, Kazunori Kohri, Takeo Moroi, Yotam Soreq, Sheng-Quan Wang, and Seokhoon Yun for useful discussions. We thank Sabyasachi Chakraborty for correspondence regarding [75] and David Dunskey for correspondence regarding [69]. This work was supported in part by IBS under the project code IBS-R018-D1, the US Department of Energy grant DE-SC0010102, and the FSU Bridge Funding 047302.

DATA AVAILABILITY

The data that support the findings of this article are openly available [111], embargo periods may apply.

APPENDIX A: AXION DECAY RATE AT NNLO

The QCD corrections to the pseudoscalar decay to gluons are calculated in Ref. [78]. Adopting the relevant corrections ($\tilde{C}_1^2$ part),

$$\begin{aligned} \Gamma(a \rightarrow gg) = & \frac{2}{\pi} \left(\frac{\alpha_s(\mu)}{8\pi} \right)^2 \frac{m_a^3}{f_a^2} \left[1 + \frac{\alpha_s(\mu)}{\pi} \left(\frac{97}{4} - \frac{7n_f}{6} - \left(\frac{11}{2} - \frac{n_f}{3} \right) \log \frac{m_a^2}{\mu^2} \right) \right. \\ & + \left(\frac{\alpha_s(\mu)}{\pi} \right)^2 \left(\frac{51959}{96} - \frac{363}{8} \zeta(2) - \frac{495}{8} \zeta(3) + \left(-\frac{469}{8} + \frac{11}{2} \zeta(2) + \frac{5}{4} \zeta(3) \right) n_f + \left(\frac{251}{216} - \frac{1}{6} \zeta(2) \right) n_f^2 \right. \\ & \left. \left. + \left(\frac{3405}{16} - \frac{73}{3} n_f + \frac{7}{12} n_f^2 \right) \log \frac{\mu^2}{m_a^2} + \left(\frac{363}{16} - \frac{11}{4} n_f + \frac{1}{12} n_f^2 \right) \log^2 \frac{\mu^2}{m_a^2} \right) \right], \end{aligned} \quad (\text{A1})$$

where $\alpha_s(\mu)$ is obtained by solving the renormalization group equation and matching condition, Eqs. (9.3) and (9.4) of [84], and $\alpha_s(m_Z) = 0.1177$. n_f is the number of quark flavors lighter than μ . We vary the renormalization scale by a factor of 2 to estimate the uncertainty.

APPENDIX B: HADRON YIELDS FROM PYTHIA AND Herwig

We use the simulation programs PYTHIA and Herwig to obtain the effective hadron yields from axion decays. These programs automatically simulate the decays of unstable

particles, such as vector mesons and hyperons. For each axion mass, we generate 10,000 events and compute the average hadron yields. We observe that both programs give consistent results for charge-conjugate pairs, such as π^+/π^- and K^+/K^- , as expected, so we take the average of them for each hadron type. The tabulated data files are available in [111].

We specifically record the information of the K_L energy spectrum, as shown in Figs. 7 and 8. Although K_L does not slow down by the electromagnetic interaction, the K_L spectrum is modified by the NK_L elastic scattering (Appendix C 3), and the reshaped spectrum is presented in Figs. 12 and 13, following the procedure described in Appendix D. We do not retain similar information for other hadrons because they are kinetically thermalized rapidly.

Let us comment on the instabilities of the two generators, which lead to discrepancies in the intermediate steps, such as the effective number of hadrons and the cross section weighted with the K_L spectrum. However, the final result, the bound on the axion lifetime, is almost insensitive to these issues, see Fig. 20, left.

- (i) *Peak from $a \rightarrow \bar{K}^0 K^*, K^0 \bar{K}^*$ for $m_a \lesssim 2.5$ GeV* The K_L distributions obtained by the two generators are unstable for the low mass $m_a \lesssim 3$ GeV. There is an abnormal peak of the K_L spectrum in PYTHIA for $m_a \lesssim 2.5$ GeV (e.g., see Fig. 8). The peak comes from direct two-body decay events of $a \rightarrow K^0 K^*$. We observe that PYTHIA does not generate three-body decay of $a \rightarrow KK\pi$ at $m_a = 2$ GeV even though it is kinematically allowed. However, this feature fades away as the axion mass increases to 3 GeV. This kind of behavior leads to some discrepancies in the average NK_L cross sections in Fig. 14.
- (ii) *Parity violation* Simulation programs do not handle polarization information properly. Therefore, sometimes it would give unphysical processes that violate parity. For example, we find up to 30% of events are invalid for $m_a = 2$ GeV. Although both PYTHIA and Herwig have such flaws, they are still the representative shower programs. We use both to solve the system, and the difference between their results can be taken as unknown systematics.
- (iii) *Baryon yield discrepancy* The two programs predict the number of nucleons, but their predictions even differ by about a factor of 2, which gets worse as m_a increases. However, this baryon decay channel is subdominant in affecting the neutron freeze-out dynamics, so this discrepancy is negligible for the final result.

APPENDIX C: UPDATES ON HADRONIC CROSS SECTIONS

In this appendix, we present the schemes to obtain the hadronic cross sections, which are necessary for the

Boltzmann equations addressed in Sec. V. We present general treatments in Sec. C 1: how we treat the average of cross sections, the Coulomb correction, and conversion of the phase space factors involved in the time reversal or isospin transformations.

Then, we apply these techniques to the cross sections involving injected pions in Sec. C 2 and kaons in Sec. C 3. We use Ref. [124] for the baryon annihilation cross sections, which is given in Sec. C 4 for completeness.

1. General treatments

Kinematically averaged cross sections. As shown in Ref. [2], hadrons injected from the axion decay, except for K_L , get quickly thermalized kinetically via electromagnetic interactions with the background photons. The timescale of the kinetic thermalization is much shorter than that of number-changing processes and, therefore, we take their kinetic distributions as the thermal distribution determined by the photon temperature, while their number densities are solved via the Boltzmann equation.

Because the hadron masses are much greater than the BBN temperature range, we can take the nonrelativistic limit where the momentum distribution is given by the Maxwell-Boltzmann distribution $f_{\text{MB}}(p, m) \propto \exp[-\frac{1}{T}(m + \frac{p^2}{2m})]$. In this approximation, we can rewrite the phase space integration as

$$d^3p_1 d^3p_2 f_{\text{MB}}(p_1; m_1) f_{\text{MB}}(p_2; m_2) = d^3P d^3p_{\text{cm}} f_{\text{MB}}(P; m_1 + m_2) f_{\text{MB}}(p_{\text{cm}}; \bar{\mu}(m_1, m_2)), \quad (\text{C1})$$

where $\vec{P} = \vec{p}_1 + \vec{p}_2$, $\vec{p}_{\text{cm}} = \bar{\mu}(\vec{p}_1/m_1 - \vec{p}_2/m_2)$ with $\bar{\mu}(m_1, m_2) = (m_1^{-1} + m_2^{-1})^{-1}$. Note that $\vec{p}_{\text{cm}}$ ($-\vec{p}_{\text{cm}}$) is the momentum of the particle 1 (2) in the center-of-momentum frame. Therefore, the averaged cross sections can be obtained by

$$\langle \sigma v(NX \rightarrow N'X') \rangle = \frac{1}{C} \int \frac{d^3p_{\text{cm}}}{(2\pi)^3} f_{\text{MB}}(p_{\text{cm}}; \bar{\mu}(m_N, m_X)) \sigma v(NX \rightarrow N'X'), \quad (\text{C2})$$

with $N(N') = p, n, X(X') = p, \bar{p}, n, \bar{n}, \pi^\pm, K^\pm$, and $C = \int \frac{d^3p}{(2\pi)^3} f_{\text{MB}}(p; \mu)$.

The relative velocity v should be the Møller velocity [126]. It is defined as

$$v \equiv \frac{|\vec{p}_{\text{cm}}| E_{\text{cm}}}{E_1 E_2} = \frac{1}{E_1 E_2} \sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2} = \frac{1}{E_1 E_2} \sqrt{(E_1 E_2 - |\vec{p}_1| |\vec{p}_2| \cos \theta)^2 - m_1^2 m_2^2}, \quad (\text{C3})$$

with the center-of-mass energy E_{cm} . This velocity becomes a familiar form, $v \simeq |\vec{p}_{\text{cm}}|/\bar{\mu} \simeq |\vec{v}_1 - \vec{v}_2|$, in the

nonrelativistic limit where $E_1 \simeq m_1 + \frac{|\vec{p}_1|^2}{2m_1}$ and $E_2 \simeq m_2 + \frac{|\vec{p}_2|^2}{2m_2}$, although we do not employ this approximation in the numerical integral.

As the thermalization of K_L is highly suppressed [2], we take the energy distribution determined by the axion decay and subsequent elastic scattering (see Sec. VI for its distribution for different m_a and Appendix D for relevant discussion) and denote it $f_{K_L}(p_{K_L}; m_a)$. Therefore, averaged cross sections for K_L -involved processes are estimated by

$$\langle \sigma v(NK_L \rightarrow N'X) \rangle = \frac{1}{C'} \int \frac{d^3 p_N}{(2\pi)^3} \frac{d^3 p_{K_L}}{(2\pi)^3} f_{\text{MB}}(p_N; m_N) \times f_{K_L}(p_{K_L}; m_a) \sigma v(NK_L \rightarrow N'X), \quad (\text{C4})$$

where $C' = \int \frac{d^3 p_N}{(2\pi)^3} f_{\text{MB}}(p_N; m_N) \cdot \int \frac{d^3 p_{K_L}}{(2\pi)^3} f_{K_L}(p_{K_L}; m_a)$.

Coulomb correction. We take into account the Coulomb correction by multiplying the Sommerfeld enhancement/suppression factor

$$F(Z, v) = \frac{|\psi(0)|^2}{|\psi_0(0)|^2}, \quad (\text{C5})$$

where $\psi(x)$ and $\psi_0(x)$ are the wave functions with and without the Coulomb potential. The general form of the correction can be written as

$$F(Z, v) = 9 \cdot 2(1+S)(2pR)^{-2(1-S)} \frac{|\Gamma(S+i\eta)|^2}{\Gamma(2S+2)} e^{\pi\eta}, \quad (\text{C6})$$

which is known as the Fermi function [127], where $S = (1 - \alpha^2 Z^2)^{1/2}$, p is the momentum of relative motion, and $R \simeq 1$ fm is the proton radius that provides the UV cutoff. $\eta = Z\alpha/v$ is the Sommerfeld parameter with the relative velocity $v = p/E$ [which approximates the Møller velocity (C3)]. As $S \simeq 1 + \mathcal{O}(\alpha^2 Z^2)$, we can approximate $|\Gamma(S+i\eta)|^2 \simeq |\Gamma(1+i\eta)|^2 = \pi\eta/\sinh \pi\eta$, and obtain

$$F(Z, v) \simeq \frac{2\pi\eta}{1 - e^{-2\pi\eta}} + \mathcal{O}(\alpha^2 Z^2). \quad (\text{C7})$$

Note that the R contribution appears in the $Z^2 \alpha^2$ order with logarithmic dependence as $\sim Z^2 \alpha^2 \log pR$, so we ignore it. We use Eq. (C7) since it is more stable numerically than using the full expression of Eq. (C6).

Phase space factors. There are many processes whose experimental measurements or fitted functions do not exist. Those cross sections can be inferred either by taking data from the reverse process or from a combination of isospin transformations. We relate scattering amplitudes by time reversal or isospin, ignoring the mass differences, while we

still need to correct the phase space difference, which is crucial near the threshold.

More explicitly, we approximate the cross section to contain the simple phase space factor as $\sigma(12 \rightarrow 34) \propto \frac{1}{E_{\text{cm}}^2} \frac{k_{\text{cm}}(3,4)}{k_{\text{cm}}(1,2)}$ and obtain the relation

$$\sigma(12 \rightarrow 34) \simeq \sigma(1'2' \rightarrow 3'4') \left[g \cdot \frac{\text{PS}(12 \rightarrow 34)}{\text{PS}(1'2' \rightarrow 3'4')} \right]_X, \quad (\text{C8})$$

where g accounts for the Coulomb correction as well as the change of internal degrees of freedom. The subindex X emphasizes that we need to fix a kinematic variable X depending on cases—it could be $k_{\text{cm}}(1,2) = k_{\text{cm}}(1',2')$ or $E_{\text{cm}}(1,2) = E_{\text{cm}}(1',2')$. We take $X = E_{\text{cm}}$ when we utilize the reverse process since we are using the invariance of the amplitude under the time reversal. On the other hand, when we use the relations under isospin transformations for nonrelativistic scattering processes, we take $X = k_{\text{cm}}$ since the nonrelativistic scattering amplitude should not care about the total mass.

The ratio of the phase space factors is given by

$$\left[\frac{\text{PS}(12 \rightarrow 34)}{\text{PS}(1'2' \rightarrow 3'4')} \right]_X = \left(\frac{k_{\text{cm}}(3,4)}{k_{\text{cm}}(1,2)} \right)_X \left(\frac{k_{\text{cm}}(1',2')}{k_{\text{cm}}(3',4')} \right)_X \left(\frac{E_{\text{cm}}(1',2')^2}{E_{\text{cm}}(1,2)^2} \right)_X. \quad (\text{C9})$$

For two cases of X , we use

$$X = E_{\text{cm}}:$$

$$k_{\text{cm}}(1,2) = \frac{E_{\text{cm}}}{2} \sqrt{\left(1 - \frac{(m_1 + m_2)^2}{E_{\text{cm}}^2}\right) \left(1 - \frac{(m_1 - m_2)^2}{E_{\text{cm}}^2}\right)}, \quad (\text{C10})$$

$$X = k_{\text{cm}}:$$

$$E_{\text{cm}}(1,2) = \sqrt{m_1^2 + k_{\text{cm}}(1,2)^2} + \sqrt{m_2^2 + k_{\text{cm}}(1,2)^2}. \quad (\text{C11})$$

Experimental data are often given in T_{lab} , the kinetic energy of an injected particle in the lab frame where the target is fixed. Therefore, it is useful to write down explicit formulas for E_{cm} and k_{cm} in terms of T_{lab} . Denoting the particle 2 as the beam particle in the lab frame for the $12 \rightarrow 34$ process, we obtain

$$E_{\text{cm}} = \sqrt{(m_1 + m_2)^2 + 2m_1 T_{\text{lab}}}, \quad (\text{C12})$$

$$k_{\text{cm}}(1,2) = \frac{m_1 \sqrt{T_{\text{lab}}(2m_2 + T_{\text{lab}})}}{\sqrt{(m_1 + m_2)^2 + 2m_1 T_{\text{lab}}}}. \quad (\text{C13})$$

2. Injected pions

$p\pi^- \rightarrow n\pi^0$ (Fig. 21): We take the partial wave analysis (PWA) [128] presented in the George Washington

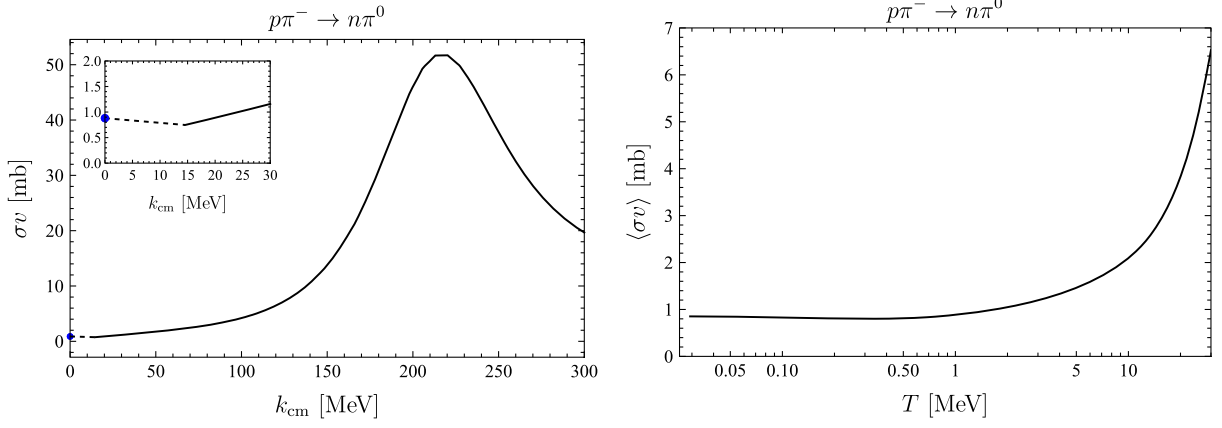


FIG. 21. Cross sections for $p\pi^- \rightarrow n\pi^0$. In the left panel, k_{cm} is momentum in the initial system ($p\pi^-$). The small panel on the left is to magnify the threshold region. The threshold value (blue point) is inferred from analysis of the $1S$ bound state of $\pi^- p$. The dashed line is a linear interpolation from SAID program data to the threshold value. The right panel shows the thermally averaged cross section in the relevant temperature range.

University SAID program [129,130]. The fitted function is provided in the format of a data table up to $E_{\pi^-} = 1$ MeV, which is still higher than what we need to know. Therefore, for the threshold cross section at $E_{\pi^-} = 0$, we take the *inferred* value of Ref. [131] from the analysis of the $1S$ bound state of $\pi^- p$,

$$\sigma v(p\pi^- \rightarrow n\pi^0) = 0.88 \text{ mb}, \quad (\text{C14})$$

at the threshold. We then take a linear interpolation, as shown in Fig. 21. The cross section we use in k_{cm} is shown on the left of Fig. 21, and the thermally averaged cross section is shown on the right.

$n\pi^+ \rightarrow p\pi^0$ (Fig. 22): We do not have experimental data for this process. Therefore, we take an isospin rotation of $p\pi^- \rightarrow n\pi^0$. In this case, as all the external particles are nonrelativistic, we fix the initial center-of-mass momentum $k_{\text{cm}}^{\text{in}}$,

$$\begin{aligned} \sigma(n\pi^+ \rightarrow p\pi^0)(k_{\text{cm}}^{\text{in}}) \\ = \sigma(p\pi^- \rightarrow n\pi^0)(k_{\text{cm}}^{\text{in}}) \times \left[\frac{1}{F(1, v)} \cdot \frac{\text{PS}(n\pi^+ \rightarrow p\pi^0)}{\text{PS}(p\pi^- \rightarrow n\pi^0)} \right]_{k_{\text{cm}}^{\text{in}}}. \end{aligned} \quad (\text{C15})$$

The $1/F(1, v)$ factor accounts for the absence of the Coulomb enhancement in the $n\pi^+$ initial state. On the other hand, the phase space correction is approximately 1.

The cross section of fitted $p\pi^- \rightarrow n\pi^0$ scattering in the center-of-mass frame is shown in Fig. 21, left, and the thermally averaged cross section is on the right. The inferred cross sections for $n\pi^+ \rightarrow p\pi^0$ are presented in Fig. 22.

$p\pi^- \rightarrow n\gamma$ and $n\pi^+ \rightarrow p\gamma$ (Figs. 23 and 24): There is no direct measurement of these processes, so we use the reverse process, whose fitting functions are given in the PWA [128–130].

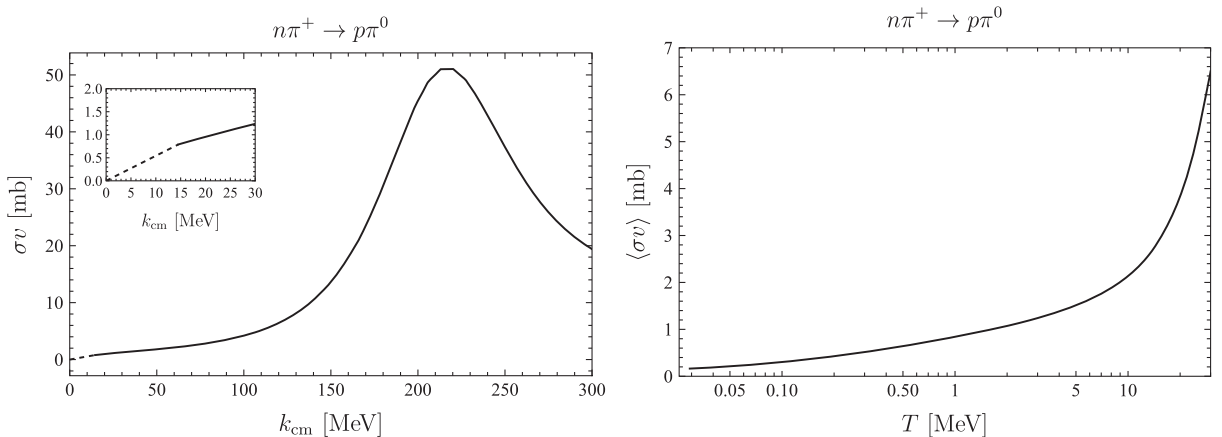


FIG. 22. Similar plots as in Fig. 21, but for $n\pi^+ \rightarrow p\pi^0$. In the left panel, the dashed line is a linear interpolation from SAID program data to the threshold value 0.

Since the threshold behavior is crucial in applying time reversal, we need to ensure the threshold energy encoded in the PWA cross sections is consistent with our input parameters, such as the nucleon and pion masses. We fit the PWA cross sections by $\sigma \propto v_{\pi,f} - v_{n(p),f}$ to find the threshold energy (see the small panels in the upper plots of Figs. 23 and 24). These thresholds slightly mismatch those derived from the up-to-date nucleon and pion masses. Thus, we slightly adjust our mass parameters to be consistent with PWA cross sections (only in these channels).

Then, we infer the cross section by the time reversal transformation as

$$\begin{aligned} \sigma(p\pi^- \rightarrow n\gamma)(E_{\text{cm}}) \\ = \sigma(n\gamma \rightarrow p\pi^-)(E_{\text{cm}}) \times \left[2 \cdot \frac{\text{PS}(p\pi^- \rightarrow n\gamma)}{\text{PS}(n\gamma \rightarrow p\pi^-)} \right]_{E_{\text{cm}}}, \quad (\text{C16}) \end{aligned}$$

$$\begin{aligned} \sigma(n\pi^+ \rightarrow p\gamma)(E_{\text{cm}}) \\ = \sigma(p\gamma \rightarrow n\pi^+)(E_{\text{cm}}) \times \left[2 \cdot \frac{\text{PS}(n\pi^+ \rightarrow p\gamma)}{\text{PS}(p\gamma \rightarrow n\pi^+)} \right]_{E_{\text{cm}}}, \quad (\text{C17}) \end{aligned}$$

for a given center-of-mass energy E_{cm} . The factor of 2 comes from the photon degree of freedom. The cross

sections as well as thermally averaged cross sections are plotted in Figs. 23 and 24.

Our result agrees well with the threshold cross section inferred from the Panofsky ratio [132,133] (see Ref. [134] for a review),

$$P = \frac{\sigma(p\pi^- \rightarrow n\pi^0)_{\text{th}}}{\sigma(p\pi^- \rightarrow n\gamma)_{\text{th}}} = 1.546, \quad (\text{C18})$$

from which one can obtain $\sigma v(p\pi^- \rightarrow n\gamma)_{\text{Panofsky}} = 0.57$ mb.

3. Injected kaons

The $2 \rightarrow 2$ scattering processes are considered in our analysis, but their cross sections are not fully known for the momentum from the threshold to $\mathcal{O}(1)$ GeV because relevant p -wave contributions were not consistently investigated and the measurements involving n or K_L are limited. Here, we utilize the known datasets with $k_{\text{lab}} < 2$ GeV cut, relate the amplitude by isospin, and perform a simultaneous fit of the scattering lengths up to p -wave for the first time.

Our method involves 14 parameters. Four of them are adopted from Ref. [118], and the remaining ten are fitted. We consider scattering amplitudes initiated by $\bar{K}N$ and KN

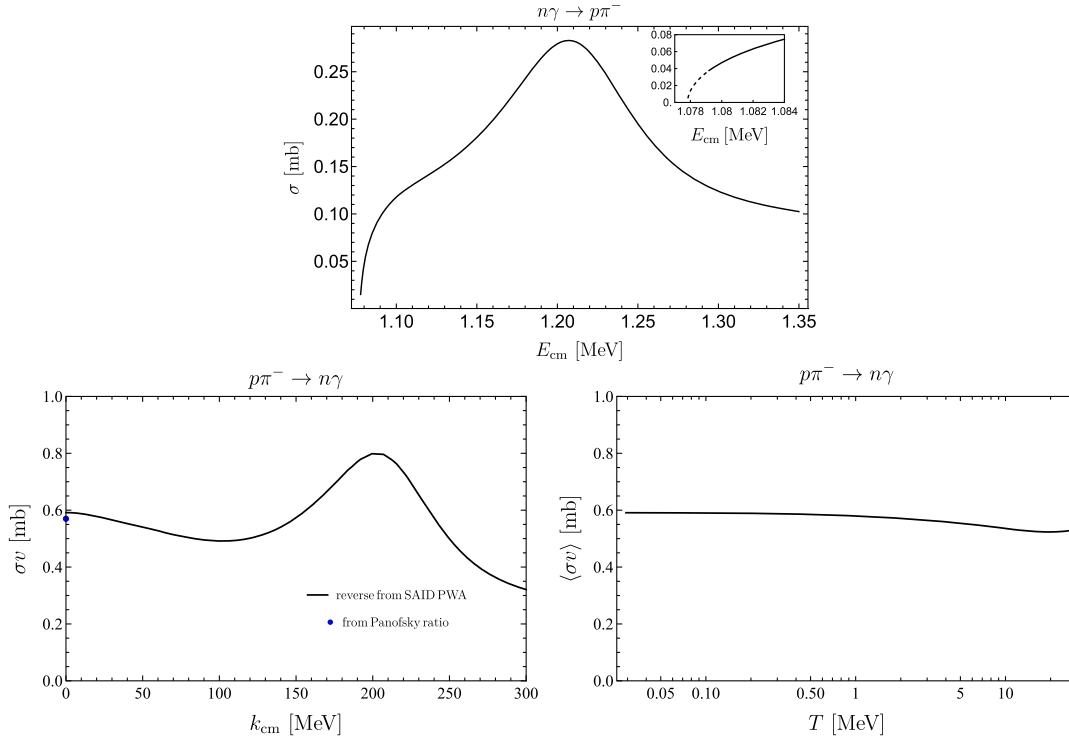


FIG. 23. Cross sections for $n\gamma \rightarrow p\pi^-$ (top) and $p\pi^- \rightarrow n\gamma$ (bottom). In the top panel, the data are from the SAID program. The inset in the top right shows a magnified view near the threshold. The dashed line (in the top small panel) is the interpolation fitted by $\sigma \propto v_{\pi,f} - v_{p,f}$. The bottom left plot is for the inverse process, which is what we obtained from the time reversal transformation. The blue point is the threshold value obtained from the Panofsky ratio, a method people usually use, with which our fitting agrees. The bottom right plot is a thermally averaged cross section in the relevant temperature range.

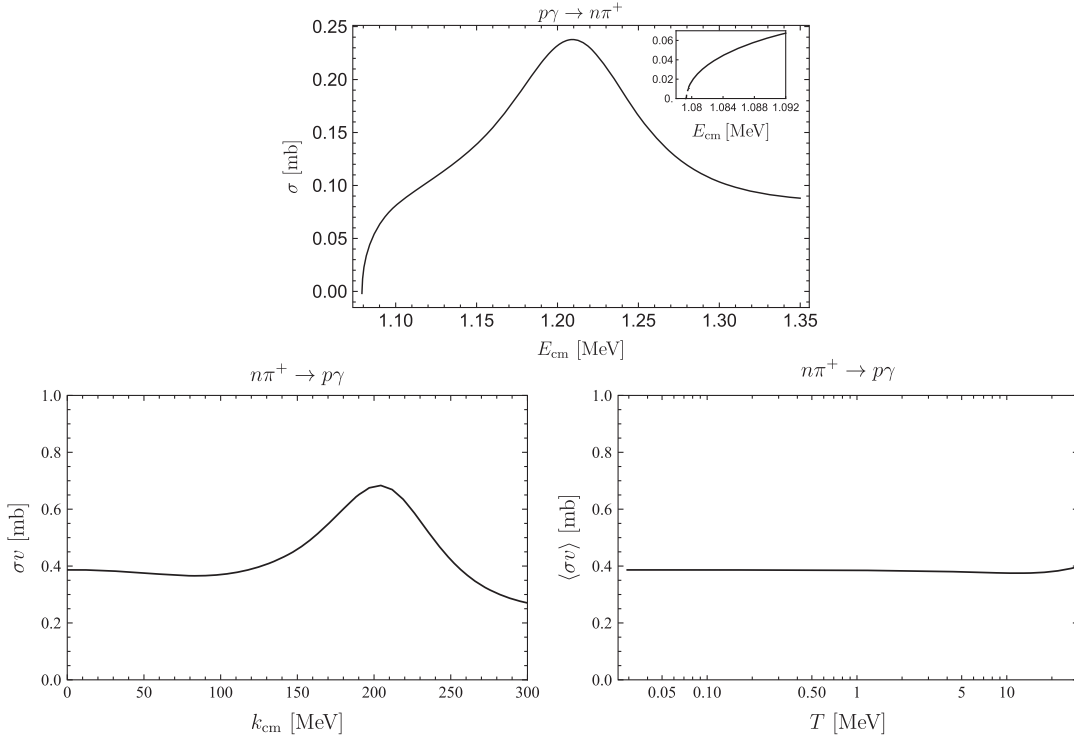


FIG. 24. Similar plots as in Fig. 23, but for $p\gamma \rightarrow n\pi^+$ (top) and $n\pi^+ \rightarrow p\gamma$ (bottom). In the top panel, the data are from the SAID program. The inset in the top right shows a zoomed view near the threshold. The dashed line in the top panel is the interpolation fitted by $\sigma \propto v_{\pi,f} - v_{n,f}$.

with the isospin channels $I = 0$ and 1, including s - and p -waves. In the $\bar{K}N$ -initiated processes, the scattering lengths are complex, resulting in eight parameters, and the s -wave parameters are given in Ref. [118], leaving four p -wave parameters. Similarly, in the KN -initiated processes, we have four scattering lengths (s - or p -waves and $I = 0$ or 1), but they are taken to be real, as hyperons are not produced. To improve the overall fit, we introduce a linear momentum dependence to the s -wave scattering length, such as $a_s = a_s^{(0)}(1 + k_{\text{cm}}^{\text{in}} a_s^{(1)})$. Thus, the KN processes are fixed by six fitting parameters.

The adopted s -wave scattering lengths of $\bar{K}N$ amplitudes from Ref. [118] are

$$A_{0,s}^{\bar{K}N} = a_{0,s}^{\bar{K}N} + ib_{0,s}^{\bar{K}N} \simeq (-1.74 + 0.70i) \text{ fm}, \quad (\text{C19})$$

$$A_{1,s}^{\bar{K}N} = a_{1,s}^{\bar{K}N} + ib_{1,s}^{\bar{K}N} \simeq (-0.05 + 0.63i) \text{ fm}, \quad (\text{C20})$$

where various hyperon productions and also $pK^- \rightarrow n\bar{K}^0$ with $k_{\text{lab}} < 0.3 \text{ GeV}$ were fitted.

In the following, we present the formulae we use for the scattering cross sections, datasets, and fitting scheme.

$\bar{K}N$ cross sections: For the s -wave, we rely on the K -matrix analysis with parameters obtained in Ref. [118], which includes the effects coming from charge and mass differences between n and p as well as K^- and $\bar{K}^0$ [135].

We add the p -wave separately, and we check that this contribution is very small in the datasets used in [118].

Let us first consider $I_3 = 0$ of hyperon production channels, where $(\bar{K}N)_0 = pK^-$ and $n\bar{K}^0$,

$$pK^- \quad \text{or} \quad n\bar{K}^0 \rightarrow \begin{cases} \pi^\pm \Sigma^\mp \\ \pi^0 \Sigma^0 \\ \pi^0 \Lambda. \end{cases} \quad (\text{C21})$$

When we replace $\bar{K}^0 \rightarrow K_L$, we evaluate the cross sections with an additional factor of 1/2. From the representation in the isospin space with proper Clebsch-Gordan coefficients, we can parametrize cross sections as

$$\begin{aligned} \sigma((\bar{K}N)_{0,s} \rightarrow \pi^\pm \Sigma^\mp) &= \frac{1}{6} \sigma_{0,s}^{(\bar{K}N)_0} + \frac{1}{4} (1 - \epsilon) \sigma_{1,s}^{(\bar{K}N)_0} \\ &\pm \left[\frac{1}{6} (1 - \epsilon) \sigma_{0,s}^{(\bar{K}N)_0} \sigma_{1,s}^{(\bar{K}N)_0} \right]^{1/2} \cos \phi \\ &+ \frac{1}{2} \sigma_{0,p}^{(\bar{K}N)_0}, \end{aligned} \quad (\text{C22})$$

$$\sigma((\bar{K}N)_0 \rightarrow \pi^0 \Sigma^0) = \frac{1}{6} \sigma_{0,s}^{(\bar{K}N)_0} + \frac{1}{2} \sigma_{0,p}^{(\bar{K}N)_0}, \quad (\text{C23})$$

$$\sigma((\bar{K}N)_0 \rightarrow \pi^0 \Lambda) = \frac{1}{2} \epsilon \sigma_{1,s}^{(\bar{K}N)_0} + \frac{3}{2} \sigma_{1,p}^{(\bar{K}N)_0}, \quad (\text{C24})$$

where $\epsilon \simeq 0.34$ denotes the ratio of $\pi\Lambda$ production over the total hyperon production ($\pi\Lambda$ and $\pi\Sigma^0$) within $I = 1$ s -wave cross sections. The similar quantity with $I = 1$ and p -wave is assumed to be 1 since it is dominated by the $\pi\Lambda$ channel [118]. For the s -wave results from [118], $\sigma_{0(1),s}^{(\bar{K}N)_0}$ are given by

$$\sigma_{0(1),s}^{(\bar{K}N)_0} = \begin{cases} \frac{4\pi b_{0(1),s}^{\bar{K}N}}{k} F(1, v) \left| \frac{1 - ik_0 A_{1(0),s}^{\bar{K}N}}{D} \right|^2 & \text{for } pK^-, \\ \frac{4\pi b_{0(1),s}^{\bar{K}N}}{k_0} \left| \frac{1}{1 - ik_0 A_{0(1)}^{\bar{K}N}} \right|^2 & \text{for } n\bar{K}^0, \end{cases} \quad (C25)$$

with k and k_0 being the pK^- and $n\bar{K}^0$ momenta in the center-of-mass frame for a given center-of-mass energy (k_0 is taken as $i|k_0|$ below the $n\bar{K}^0$ threshold) [135]. $b_{0,1}^{\bar{K}N}$ is the imaginary part of scattering length $A_{0,1}^{\bar{K}N}$. Here, D is given by

$$D = 1 - \frac{i}{2} (A_0^{\bar{K}N} + A_1^{\bar{K}N}) [k_0 + kF(1, v)(1 - i\lambda)] - k_0 kF(1, v)(1 - i\lambda) A_0^{\bar{K}N} A_1^{\bar{K}N}, \quad (C26)$$

$$\lambda = -\frac{2}{kBF(1, v)} \left[\log(2kR) + \text{Re} \left[\frac{\Gamma'(i/kB)}{\Gamma(i/kB)} \right] + 2\gamma_E \right], \quad (C27)$$

where γ_E is the Euler constant, $B = (\alpha_{\text{EM}}\mu)^{-1}$ is the Bohr radius of the pK^- system, $R \simeq 0.4$ fm is the interaction radius, and

$$\phi = \phi_{\text{th}} + \text{Arg} \left[\frac{1 - ik_0 A_1^{\bar{K}N}}{1 - ik_0 A_0^{\bar{K}N}} \right]. \quad (C28)$$

From the fitting to experimental data of pK^- cross sections in Ref. [118], $\phi_{\text{th}} \simeq -52.9^\circ$.

The p -wave cross sections, $\sigma_{0(1),p}^{(\bar{K}N)_0}$, are given by

$$\sigma_{0(1),p}^{(\bar{K}N)_0} = \begin{cases} 4\pi F(1, v)(1 + \eta^2) \frac{k \text{Im}[(A_{0(1),p}^{\bar{K}N})^3]}{|1 - ik^3 (A_{0(1),p}^{\bar{K}N})^3|^2} & \text{for } pK^-, \\ 4\pi \frac{k_0 \text{Im}[(A_{0(1),p}^{\bar{K}N})^3]}{|1 - ik_0^3 (A_{0(1),p}^{\bar{K}N})^3|^2} & \text{for } n\bar{K}^0. \end{cases} \quad (C29)$$

We verify that this p -wave is subleading for $k_{\text{lab}} < 280$ MeV, which agrees with the statement in [118]. The factor of $(1 + \eta^2)$ for pK^- corresponds to the correction to the Coulomb factor for p -wave contributions [136].

The cross sections of $pK^- \rightarrow \pi Y$ and $nK_L \rightarrow \pi Y$ scattering in the center-of-mass frame as well as the thermally averaged cross sections are shown in Figs. 25 and 26.

Hyperons can also be produced by the $I_3 = \pm 1$ processes where $(\bar{N}K)_+ = p\bar{K}^0$ and $(\bar{K}N)_- = nK^-$,

$$\begin{aligned} \sigma(p\bar{K}^0 \rightarrow \pi^0 \Sigma^+) &= \sigma(p\bar{K}^0 \rightarrow \pi^+ \Sigma^0) \\ &= (1 - \epsilon) \frac{2\pi}{k_{\text{cm}}^{\text{in}} |1 - ik_{\text{cm}}^{\text{in}} A_{1,s}^{\bar{K}N}|^2} b_{1,s}^{\bar{K}N}, \end{aligned} \quad (C30)$$

$$\begin{aligned} \sigma(p\bar{K}^0 \rightarrow \pi^+ \Lambda) &= \epsilon \frac{4\pi}{k_{\text{cm}}^{\text{in}} |1 - ik_{\text{cm}}^{\text{in}} A_{1,s}^{\bar{K}N}|^2} b_{1,s}^{\bar{K}N} \\ &+ 12\pi \frac{k_{\text{cm}}^{\text{in}} \text{Im}[(A_{1,p}^{\bar{K}N})^3]}{|1 - i(k_{\text{cm}}^{\text{in}})^3 (A_{1,p}^{\bar{K}N})^3|^2}, \end{aligned} \quad (C31)$$

$$\begin{aligned} \sigma(nK^- \rightarrow \pi^0 \Sigma^-) &= \sigma(nK^- \rightarrow \pi^- \Sigma^0) \\ &= (1 - \epsilon) \frac{2\pi}{k_{\text{cm}}^{\text{in}} |1 - ik_{\text{cm}}^{\text{in}} A_{1,s}^{\bar{K}N}|^2} b_{1,s}^{\bar{K}N}, \end{aligned} \quad (C32)$$

$$\begin{aligned} \sigma(nK^- \rightarrow \pi^- \Lambda) &= \epsilon \frac{4\pi}{k_{\text{cm}}^{\text{in}} |1 - ik_{\text{cm}}^{\text{in}} A_{1,s}^{\bar{K}N}|^2} b_{1,s}^{\bar{K}N} \\ &+ 12\pi \frac{k_{\text{cm}}^{\text{in}} \text{Im}[(A_{1,p}^{\bar{K}N})^3]}{|1 - i(k_{\text{cm}}^{\text{in}})^3 (A_{1,p}^{\bar{K}N})^3|^2}. \end{aligned} \quad (C33)$$

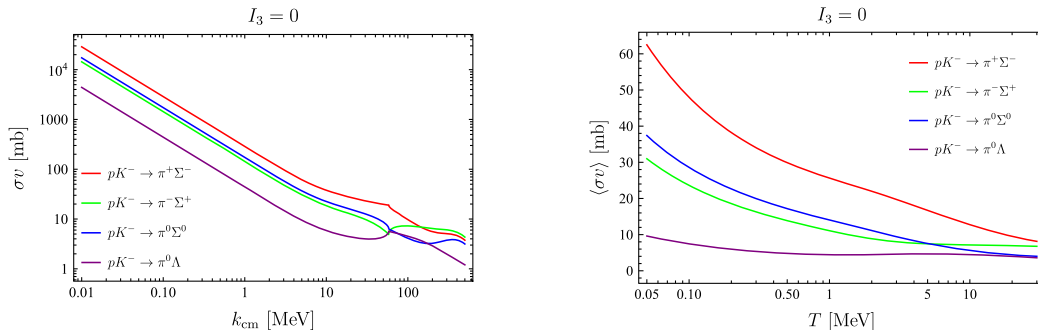


FIG. 25. Cross sections for $pK^- \rightarrow \pi Y$. The kink at $k_{\text{cm}} = 58.23$ MeV is the division of k_0 being real or imaginary.

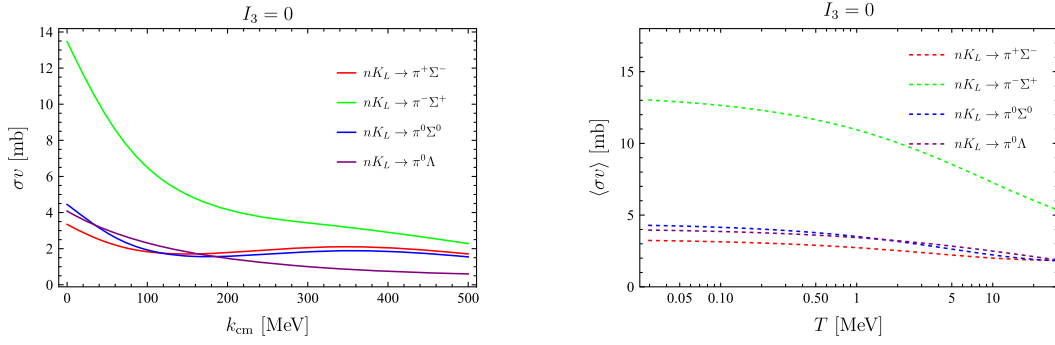


FIG. 26. Cross sections for $nK_L \rightarrow \pi Y$. Note that the right panel assumes the Maxwell-Boltzmann distribution, therefore, it is not used in our calculation because the K_L 's are not thermalized. One should use the momentum spectrum determined from the axion decay. The cross sections we use are shown in Fig. 33.

The K_L cross section is just given by $\sigma(NK_L \rightarrow \pi Y) = \frac{1}{2}\sigma(N\bar{K}^0 \rightarrow \pi Y)$.

The cross sections of $p\bar{K}^0 \rightarrow \pi Y$ and $nK^- \rightarrow \pi Y$ scattering in the center-of-mass frame as well as the thermally averaged cross sections are shown in Figs. 27 and 28.

Now we consider $N\bar{K} \rightarrow N'\bar{K}'$ processes. To proceed, it is useful to define scattering amplitude expressions as

$$T_s(A_s) = \frac{A_s}{1 - ik_{\text{cm}}^{\text{in}} A_s}, \quad (\text{C34})$$

$$T_p(A_p) = T_s((k_{\text{cm}}^{\text{in}})^2 A_p^3). \quad (\text{C35})$$

Note that, in our convention, A_p still remains in the dimension of length, while A_p^3 has the dimension of volume.

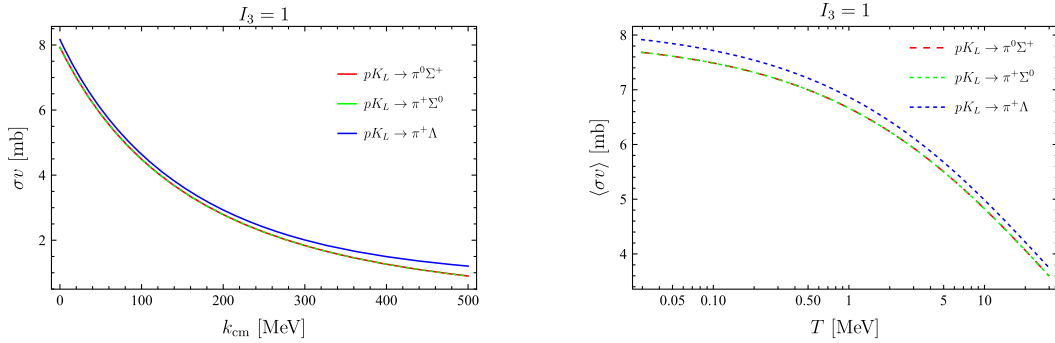


FIG. 27. Cross sections for $pK_L \rightarrow \pi Y$. The cross sections for $pK_L \rightarrow \pi^+\Sigma^0$ and $pK_L \rightarrow \pi^0\Sigma^+$ are the same up to the mass difference. The right panel is not used, as explained in the caption of Fig. 26. The cross sections we use are shown in Fig. 34.

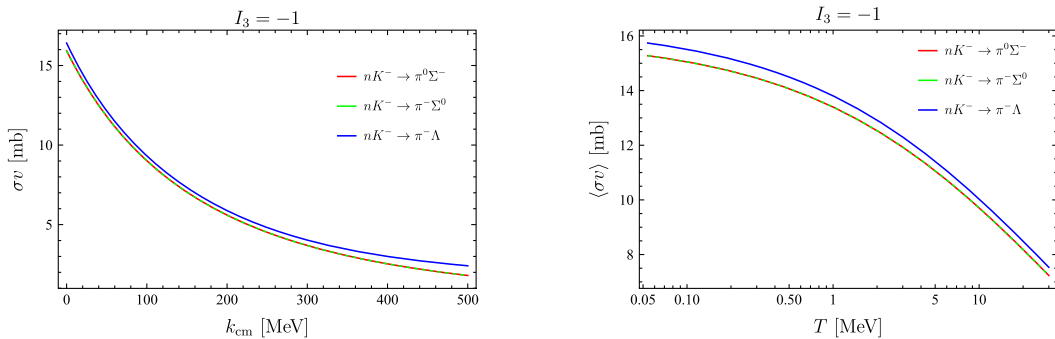


FIG. 28. Cross sections for $nK^- \rightarrow \pi Y$. The cross sections for $nK^- \rightarrow \pi^-\Sigma^0$ and $nK^- \rightarrow \pi^0\Sigma^-$ are the same up to the small mass difference.

The charge exchange processes, $pK^- \leftrightarrow n\bar{K}^0$, are given by

$$\sigma(pK^- \rightarrow n\bar{K}^0) = \frac{\pi k_0 F(1, v)}{k} \left(\left| \frac{A_{1,s}^{\bar{K}N} - A_{0,s}^{\bar{K}N}}{D} \right|^2 + 3(1 + \eta^2) |T_p(A_{1,p}^{\bar{K}N}) - T_p(A_{0,p}^{\bar{K}N})|^2 \right), \quad (\text{C36})$$

$$\sigma(n\bar{K}^0 \rightarrow pK^-) = \frac{\pi k F(1, v)}{k_0} \left(\left| \frac{A_{1,s}^{\bar{K}N} - A_{0,s}^{\bar{K}N}}{D} \right|^2 + 3(1 + \eta^2) |T_p(A_{1,p}^{\bar{K}N}) - T_p(A_{0,p}^{\bar{K}N})|^2 \right), \quad (\text{C37})$$

with v being the relative velocity in the pK^- system. $\bar{K}^0$ in the final state gives both K_L and K_S as $\sigma(pK^- \rightarrow n\bar{K}_{L(S)}) = \frac{1}{2}\sigma(pK^- \rightarrow n\bar{K}^0)$. The cross sections of $pK^- \leftrightarrow n\bar{K}^0$ scattering in the center-of-mass frame as well as the thermally averaged cross sections are shown in Fig. 29.

The elastic scattering of pK^- is also induced by the $\bar{K}N$ amplitude,

$$\sigma(pK^- \rightarrow pK^-) = \pi F(1, v)^2 \{ |T_s(A_{1,s}^{\bar{K}N}) + T_s(A_{0,s}^{\bar{K}N})|^2 + 3(1 + \eta^2) |T_p(A_{1,p}^{\bar{K}N}) + T_p(A_{0,p}^{\bar{K}N})|^2 \}. \quad (\text{C38})$$

The cross section is shown in Fig. 32 (second row, left).

KN cross sections: Unlike the $N\bar{K}$ system, hyperon production processes are forbidden and, therefore, only four states are possible: pK^+ , pK^0 , nK^+ , and nK^0 (K^0 is written with respect to K_L and K_S). Due to the absence of

the hyperon channels, we consider both s - and p -waves with real scattering lengths,

$$\sigma(pK^+ \rightarrow pK^+) = 4\pi F(-1, v)^2 \{ |T_s(a_{1,s}^{KN})|^2 + 3(1 + \eta^2) |T_p(a_{1,p}^{KN})|^2 \}, \quad (\text{C39})$$

$$\sigma(pK^0 \rightarrow nK^+) = \pi \frac{k_{\text{cm}}^{\text{out}}}{k_{\text{cm}}^{\text{in}}} \{ |T_s(a_{1,s}^{KN}) - T_s(a_{0,s}^{KN})|^2 + 3 |T_p(a_{1,p}^{KN}) - T_p(a_{0,p}^{KN})|^2 \}, \quad (\text{C40})$$

$$\sigma(nK^+ \rightarrow pK^0) = \pi \frac{k_{\text{cm}}^{\text{out}}}{k_{\text{cm}}^{\text{in}}} \{ |T_s(a_{1,s}^{KN}) - T_s(a_{0,s}^{KN})|^2 + 3 |T_p(a_{1,p}^{KN}) - T_p(a_{0,p}^{KN})|^2 \}, \quad (\text{C41})$$

where $k_{\text{cm}}^{\text{out}}$ is the outgoing momentum in the center-of-mass frame. As before, when $\bar{K}^0$ is replaced with $K_{L,S}$, the cross section is multiplied by 1/2. For the s -wave, we introduce k dependence as $a_s = a_s^{(0)}(1 + k_{\text{cm}}^{\text{in}} a_s^{(1)})$ to describe the pK^- elastic scattering better.

The cross sections of $pK^0 \leftrightarrow n\bar{K}^+$ scattering in the center-of-mass frame as well as the thermally averaged cross sections are shown in Fig. 30. The cross section of pK^+ elastic scattering is shown in Fig. 32 (second row, right).

$\bar{K}N$ and KN processes: Both $\bar{K}N$ and KN amplitudes are necessary for the K_L elastic scattering processes and the regeneration processes,

$$\begin{aligned} \sigma(pK_L \rightarrow pK_S) &= \pi \left| \frac{1}{2} (T_s(a_{1,s}^{KN}) + T_s(a_{0,s}^{KN})) - T_s(A_{1,s}^{\bar{K}N}) \right|^2 \\ &\quad + 3\pi \left| \frac{1}{2} (T_p(a_{1,p}^{KN}) + T_p(a_{0,p}^{KN})) - T_p(A_{1,p}^{\bar{K}N}) \right|^2, \end{aligned} \quad (\text{C42})$$

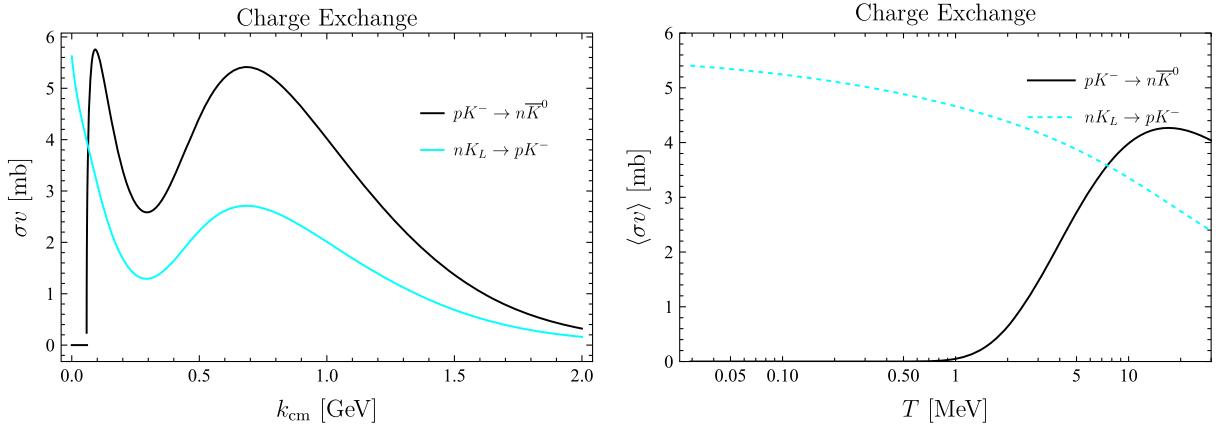


FIG. 29. Cross sections for $pK^- \leftrightarrow n\bar{K}_L$ as a function of k_{cm} (left) and the thermally averaged cross sections (right). Thermally averaged cross section for $nK_L \rightarrow pK^-$ is not used, as explained in the caption of Fig. 26. The correct cross sections are shown in Fig. 33.

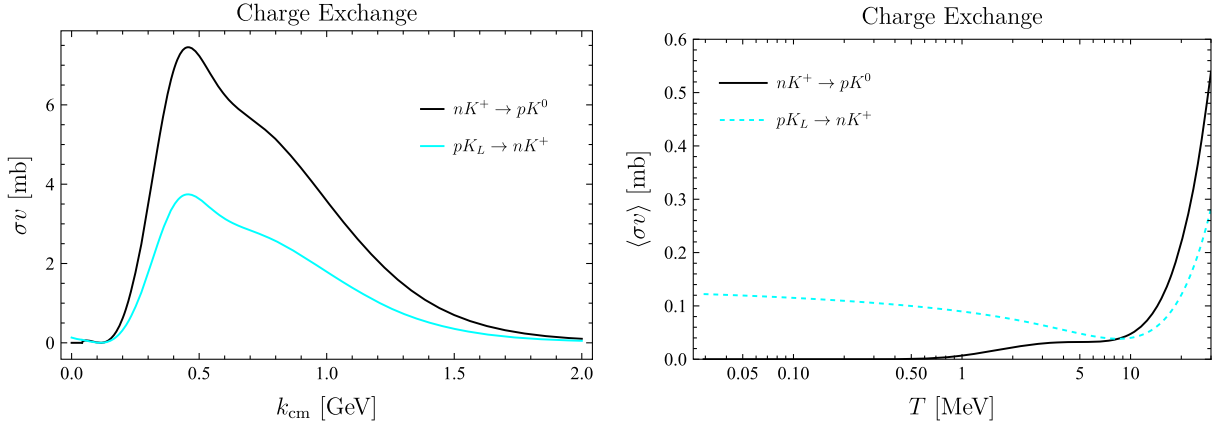


FIG. 30. Cross sections for $pK_L \leftrightarrow nK^+$ as a function of k_{cm} (left) and the thermally averaged cross sections (right). A thermally averaged cross section for $pK_L \rightarrow pK^+$ is not used, as explained in the caption of Fig. 26. The correct cross sections are shown in Fig. 34.

$$\begin{aligned} \sigma(pK_L \rightarrow pK_L) &= \pi \left| \frac{1}{2} (T_s(a_{1,s}^{KN}) + T_s(a_{0,s}^{KN})) + T_s(A_{1,s}^{\bar{K}N}) \right|^2 \\ &\quad + 3\pi \left| \frac{1}{2} (T_p(a_{1,p}^{KN}) + T_p(a_{0,p}^{KN})) + T_p(A_{1,p}^{\bar{K}N}) \right|^2, \end{aligned} \quad (\text{C43})$$

$$\begin{aligned} \sigma(nK_L \rightarrow nK_S) &= \pi \left| \frac{1}{2} (T_s(A_{0,s}^{\bar{K}N}) + T_s(A_{1,s}^{\bar{K}N})) - T_s(a_{1,s}^{KN}) \right|^2 \\ &\quad + 3\pi \left| \frac{1}{2} (T_p(A_{0,p}^{\bar{K}N}) + T_p(A_{1,p}^{\bar{K}N}) - T_p(a_{1,p}^{KN})) \right|^2, \end{aligned} \quad (\text{C44})$$

$$\begin{aligned} \sigma(nK_L \rightarrow nK_L) &= \pi \left| \frac{1}{2} (T_s(A_{0,s}^{\bar{K}N}) + T_s(A_{1,s}^{\bar{K}N})) + T_s(a_{1,s}^{KN}) \right|^2 \\ &\quad + 3\pi \left| \frac{1}{2} (T_p(A_{0,p}^{\bar{K}N}) + T_p(A_{1,p}^{\bar{K}N}) + T_p(a_{1,p}^{KN})) \right|^2. \end{aligned} \quad (\text{C45})$$

Including the p -wave contribution is necessary to capture the features of charge exchange and regeneration processes. The cross sections of those processes are given in Fig. 31.

Fitting method and parameters: In order to determine ten parameters, we utilize seven datasets and evaluate the combined χ^2 .

For the pure $\bar{K}N$ reactions, the measured processes are

$$pK^- \rightarrow pK^-, \quad pK^- \rightarrow n\bar{K}^0, \quad p\bar{K}^0 \rightarrow \pi^+\Lambda, \quad (\text{C46})$$

and the corresponding datasets are from the 2022 edition of the PDG [137], Ref. [121], and Fig. 30 of Ref. [122] (originally Ref. [138]), respectively.

For the pure KN reactions, we can use

$$pK^+ \rightarrow pK^+, \quad pK^0 \rightarrow nK^+. \quad (\text{C47})$$

The corresponding datasets are from the 2022 edition of the PDG [137] and Fig. 35 of Ref. [122] (Ref. [139]), respectively.

Two pK_L scattering processes are measured,

$$\begin{aligned} pK_L &\rightarrow pK_S, \\ \text{Inclusive } pK_L (k_{\text{lab}} < 0.3 \text{ GeV}). \end{aligned} \quad (\text{C48})$$

The corresponding datasets are from Figs. 27 of Ref. [122] (Ref. [123]) and 4 of Ref. [119] (which includes data from Ref. [120]), respectively.

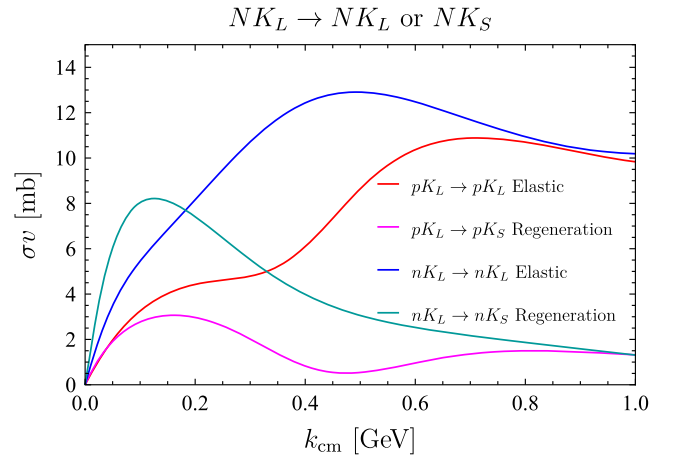


FIG. 31. Cross sections for elastic processes $NK_L \rightarrow NK_L$ and regeneration processes $NK_L \rightarrow NK_S$ as a function of k_{cm} . N can be a proton or a neutron.

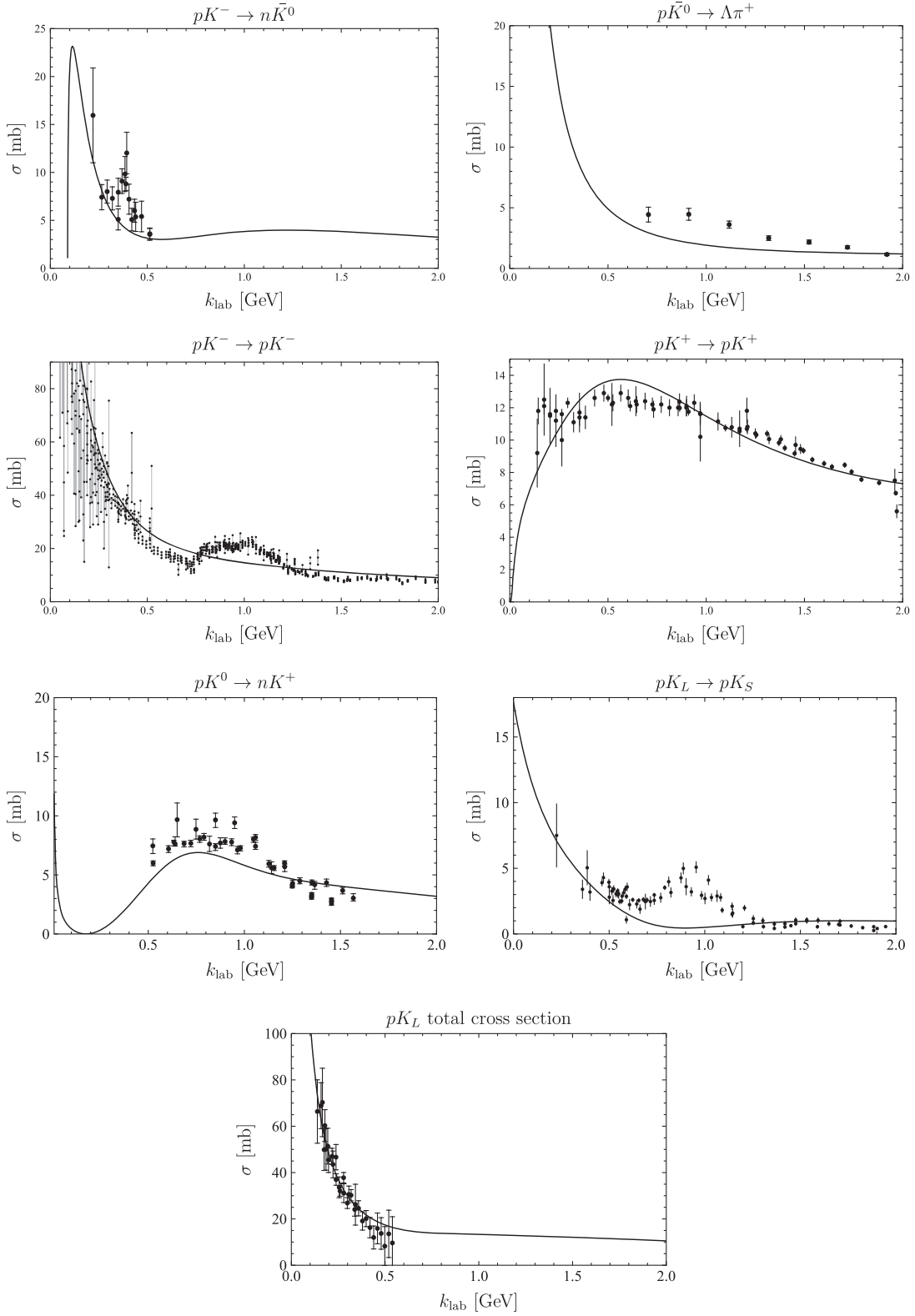


FIG. 32. Ten parameters of our scattering lengths are determined by fitting seven different datasets, as indicated in each panel. The solid line shows our fit, and the data points with error bars represent the experimental measurements. See the main text for further details. Since we focus on $2 \rightarrow 2$ scattering processes of NK or $N\bar{K}$, the momentum range is restricted to $k_{\text{lab}} < 2$ GeV.

To perform the fit, we consider both statistical and systematic (if reported) uncertainties, and construct χ^2 summing over the data points with $k_{\text{lab}} < 2$ GeV.

Generating 200,000 initial seed points, we look for the local minimum of χ^2 , and pick the parameter set that gives the smallest χ^2 , that is,

$$\begin{aligned} a_{0,s}^{(0),KN} &= 0.48 \text{ fm}, & a_{1,s}^{(0),KN} &= 0.23 \text{ fm}, \\ a_{0,s}^{(1),KN} &= -0.69 \text{ fm}, & a_{1,s}^{(1),KN} &= 0.49 \text{ fm}, \\ a_{0,p}^{KN} &= 0.28 \text{ fm}, & a_{1,p}^{KN} &= 0.15 \text{ fm}, \\ (A_{0,p}^{\bar{K}N})^3 &= (-0.012 + 0.053i) \text{ fm}^3, \\ (A_{1,p}^{\bar{K}N})^3 &= (0.0017 + 0.00050i) \text{ fm}^3. \end{aligned} \quad (\text{C49})$$

The fitted curves and the data with error bars are shown in Fig. 32. Recall that four parameters are from Eqs. (C19) and (C20).

These data are fitted fairly well, as the data points with errors are shown if a measurement exists. We do not evaluate the uncertainties because this is beyond the scope of this paper.

4. Injected baryons

Since the axion has neither a baryon number nor an electric charge, the net proton number and neutron number from one axion decay must be zero. This implies that X_n is not modified if an injected antiproton (or antineutron) annihilates with a proton (or neutron), while X_n receives an effective modification only when an antiproton (or antineutron) annihilates with a neutron (or proton). So, we need to know four possible annihilation channels: $p\bar{p}$, $n\bar{p}$, $p\bar{n}$, and $n\bar{n}$ into $\pi\pi$ (annihilation into $\gamma\gamma$ is quantum electrodynamics-suppressed).

For this, we follow the analysis presented in Ref. [124],

$$\sigma_{\text{ann}} = \frac{\pi}{k^2} \sum_{L=0}^{L_{\text{max}}} (2L+1) T_L(k) G_L(k), \quad (\text{C50})$$

where

$$T_L = \frac{4s_L K R}{\Delta_L^2 + (s_L + K R)^2} \quad (\text{C51})$$

for $K = \sqrt{k^2 + 2\mu V_0}$ and

$$s_L = R \left(\frac{g_L(df_L/dr) - f_L(dg_L/dr)}{g_L^2 + f_L^2} \right)_{r=R} \quad (\text{C52})$$

$$\Delta_L = R \left(\frac{g_L(df_L/dr) + f_L(dg_L/dr)}{g_L^2 + f_L^2} \right)_{r=R}, \quad (\text{C53})$$

$$f_L(r) = \left(\frac{\pi k r}{2} \right)^{1/2} J_{L+1/2}(kr), \quad (\text{C54})$$

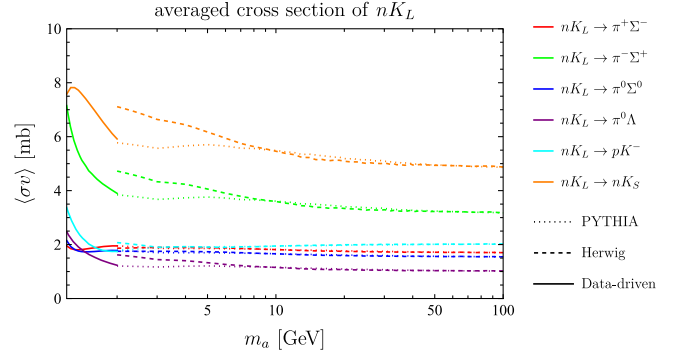


FIG. 33. nK_L cross sections averaged by using the kinematics from the axion decay. For $m_a < 2$ GeV, the data-driven method is used. For $m_a > 2$ GeV, two generators, PYTHIA and Herwig, are used. PYTHIA's result agrees better with the data-driven method at 2 GeV.

$$g_L(r) = -\left(\frac{\pi k r}{2} \right)^{1/2} Y_{L+1/2}(kr). \quad (\text{C55})$$

Here, $J_n(kr)$ and $Y_n(kr)$ are, respectively, Bessel and Neumann functions. The Coulomb correction is included in G_L ($G_L = 1$ for $n\bar{n}$, $p\bar{n}$, and $n\bar{p}$),

$$G_L(k) = \frac{(L^2 + \xi^2)((L-1)^2 + \xi^2) \cdots (1 + \xi^2)}{(L!)^2} F(Z, v). \quad (\text{C56})$$

R and V_0 are the parameters of a square well potential, $V(r) = -V_0\Theta(R-r)$, whose physical meanings can be interpreted as the nuclear contact radius and the strong interaction potential depth. As analyzed in Ref. [124], these parametrizations agree with experimental data well using $R = 0.97$ fm and $V_0 = 85$ MeV.

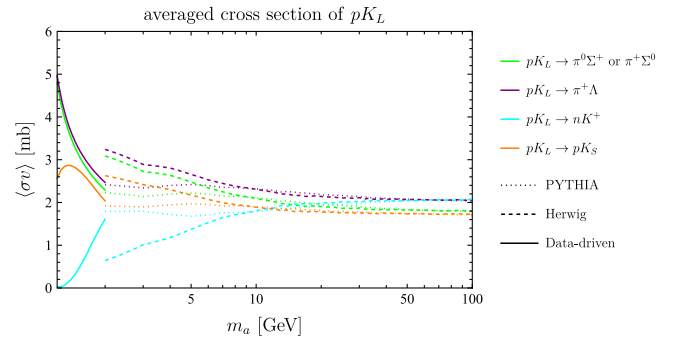


FIG. 34. pK_L cross sections averaged by using the kinematics from the axion decay. For $m_a < 2$ GeV, the data-driven method is used. For $m_a > 2$ GeV, two generators, PYTHIA and Herwig, are used. PYTHIA's result agrees better with the data-driven method at 2 GeV.

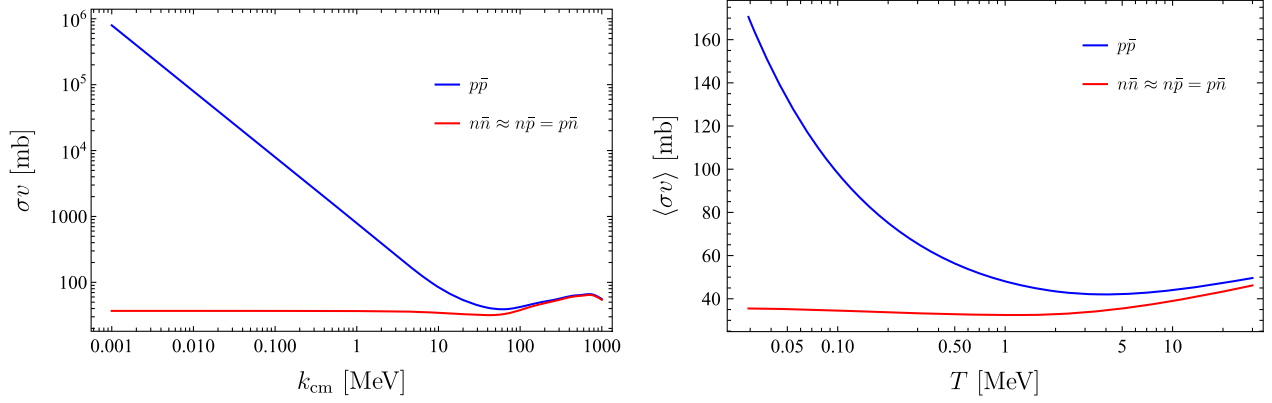


FIG. 35. Cross sections for $N\bar{N}'$ annihilation. We make an estimation that $n\bar{n}$, $n\bar{p}$, and $p\bar{n}$ have the same cross sections.

APPENDIX D: RESHAPING K_L DISTRIBUTION

Based on our assessment of various $K_L N$ scattering cross sections, we find that elastic scattering is significant in the high-momentum region, $1 \text{ GeV} \lesssim E_{K_L} \lesssim 2 \text{ GeV}$. Although elastic scattering does not contribute to the Boltzmann equations as a number-changing process, it does modify the K_L energy spectrum relative to the distribution originating from axion decay. In this appendix, we present prescriptions for incorporating elastic scattering effects and obtaining the reshaped K_L spectrum, which is relevant for subsequent scattering processes in which K_L is depleted.

Firstly, we ignore both the Hubble expansion and K_L decay, which are good approximations since the timescale of K_L scattering with a nucleon is quite short. Then, a K_L with energy E has two branches: a fraction $r(E)$ redistributes to different energies due to elastic scattering, and the remaining fraction, $1 - r(E)$, disappears due to number-changing scattering processes. The redistributed spectrum undergoes another iteration, and eventually, the remaining component becomes negligible after enough iterations.

A specific algorithm is as follows. We bin the K_L energy spectrum from threshold to $E_{K_L} = 2.023 \text{ GeV}$ (equivalent to $k_{\text{lab}} = 2 \text{ GeV}$), with a bin size of 50 MeV (25 MeV for $m_a \leq 2 \text{ GeV}$). Let n_i denote the number of K_L particles in the i th bin, such that the total $\sum_i n_i$ is $N_{a \rightarrow K_L}$. Then, n_i either migrates to other bins $n_{j \leq i}$ through elastic scattering, or is added to n_i^{dis} , which is initially zero,

$$\text{Elastic scattering: } n_j = \sum_{i=1}^{N_{\text{bins}}} n_i r(E_i) f(E_i \rightarrow E_j), \quad (\text{D1})$$

$$\text{Disappearance: } n_i^{\text{dis}} + = n_i (1 - r(E_i)). \quad (\text{D2})$$

Here, $f(E_i \rightarrow E_j)$ is the probability distribution for elastic scattering (assuming isotropic in the center-of-mass frame), where the K_L energy changes from E_i to E_j . The function $r(E_i)$ depends on the cross section ratios and also on X_n ,

$$r(E) = \frac{X_n \sigma_{nK_L}^{\text{elastic}} + (1 - X_n) \sigma_{pK_L}^{\text{elastic}}}{X_n \sigma_{nK_L}^{\text{tot}} + (1 - X_n) \sigma_{pK_L}^{\text{tot}}}. \quad (\text{D3})$$

Fortunately, $r(E)$ is less sensitive to X_n because the cross section ratios are accidentally similar, $\sigma_{nK_L}^{\text{elastic}} / \sigma_{nK_L}^{\text{tot}} \simeq \sigma_{pK_L}^{\text{elastic}} / \sigma_{pK_L}^{\text{tot}}$. We choose $X_n = 0.5$ as a representative value. The elastically scattered component, n_j , is recycled in the next migration step, and n_i^{dis} is incrementally built up in each iteration.

After about six iterations, the K_L disappears, and the reshaping procedure is complete. The modified spectrum used for computing averaged cross sections is stored in n_i^{dis} , where $\sum_i n_i^{\text{dis}} \approx N_{a \rightarrow K_L}$. Note that the updated K_L spectrum is still very different from the thermal distribution.

- [1] M. H. Reno and D. Seckel, Primordial nucleosynthesis: The effects of injecting hadrons, *Phys. Rev. D* **37**, 3441 (1988).
- [2] K. Kohri, Primordial nucleosynthesis and hadronic decay of a massive particle with a relatively short lifetime, *Phys. Rev. D* **64**, 043515 (2001).

- [3] K. Kohri and J. Yokoyama, Primordial black holes and primordial nucleosynthesis. 1. Effects of hadron injection from low mass holes, *Phys. Rev. D* **61**, 023501 (2000).
- [4] M. Kawasaki, K. Kohri, and N. Sugiyama, MeV scale reheating temperature and thermalization of neutrino background, *Phys. Rev. D* **62**, 023506 (2000).

- [5] K. Jedamzik, Did something decay, evaporate, or annihilate during big bang nucleosynthesis?, *Phys. Rev. D* **70**, 063524 (2004).
- [6] M. Kawasaki, K. Kohri, and T. Moroi, Hadronic decay of late—decaying particles and big-bang nucleosynthesis, *Phys. Lett. B* **625**, 7 (2005).
- [7] M. Kawasaki, K. Kohri, and T. Moroi, Big-Bang nucleosynthesis and hadronic decay of long-lived massive particles, *Phys. Rev. D* **71**, 083502 (2005).
- [8] K. Kohri, T. Moroi, and A. Yotsuyanagi, Big-bang nucleosynthesis with unstable gravitino and upper bound on the reheating temperature, *Phys. Rev. D* **73**, 123511 (2006).
- [9] K. Jedamzik, Big bang nucleosynthesis constraints on hadronically and electromagnetically decaying relic neutral particles, *Phys. Rev. D* **74**, 103509 (2006).
- [10] M. Kawasaki, K. Kohri, T. Moroi, and A. Yotsuyanagi, Big-Bang nucleosynthesis and gravitino, *Phys. Rev. D* **78**, 065011 (2008).
- [11] R. H. Cyburt, J. Ellis, B. D. Fields, F. Luo, K. A. Olive, and V. C. Spanos, Nucleosynthesis constraints on a massive gravitino in neutralino dark matter scenarios, *J. Cosmol. Astropart. Phys.* **10** (2009) 021.
- [12] R. H. Cyburt, J. Ellis, B. D. Fields, F. Luo, K. A. Olive, and V. C. Spanos, Nuclear reaction uncertainties, massive gravitino decays and the cosmological lithium problem, *J. Cosmol. Astropart. Phys.* **10** (2010) 032.
- [13] R. H. Cyburt, J. Ellis, B. D. Fields, F. Luo, K. A. Olive, and V. C. Spanos, Gravitino decays and the cosmological lithium problem in light of the LHC Higgs and supersymmetry searches, *J. Cosmol. Astropart. Phys.* **05** (2013) 014.
- [14] M. Kawasaki, K. Kohri, T. Moroi, and Y. Takaesu, Revisiting big-bang nucleosynthesis constraints on long-lived decaying particles, *Phys. Rev. D* **97**, 023502 (2018).
- [15] T. Hasegawa, N. Hiroshima, K. Kohri, R. S. L. Hansen, T. Tram, and S. Hannestad, MeV-scale reheating temperature and thermalization of oscillating neutrinos by radiative and hadronic decays of massive particles, *J. Cosmol. Astropart. Phys.* **12** (2019) 012.
- [16] L. Angel, G. Arcadi, M. M. A. Paixão, and F. S. Queiroz, Updated BBN bounds on hadronic injection in the early universe: The gravitino problem, [arXiv:2501.09120](https://arxiv.org/abs/2501.09120).
- [17] A. Fradette and M. Pospelov, BBN for the LHC: Constraints on lifetimes of the Higgs portal scalars, *Phys. Rev. D* **96**, 075033 (2017).
- [18] A. Fradette, M. Pospelov, J. Pradler, and A. Ritz, Cosmological constraints on very dark photons, *Phys. Rev. D* **90**, 035022 (2014).
- [19] J. Berger, K. Jedamzik, and D. G. E. Walker, Cosmological constraints on decoupled dark photons and dark Higgs, *J. Cosmol. Astropart. Phys.* **11** (2016) 032.
- [20] A. Boyarsky, M. Ovchinnikov, O. Ruchayskiy, and V. Syvolap, Improved big bang nucleosynthesis constraints on heavy neutral leptons, *Phys. Rev. D* **104**, 023517 (2021).
- [21] Y.-M. Chen and Y. Zhang, BBN constraint on heavy neutrino production and decay, *Phys. Rev. D* **111**, 123024 (2025).
- [22] K. Jedamzik and M. Pospelov, Big bang nucleosynthesis and particle dark matter, *New J. Phys.* **11**, 105028 (2009).
- [23] B. Henning and H. Murayama, Constraints on light dark matter from big bang nucleosynthesis, [arXiv:1205.6479](https://arxiv.org/abs/1205.6479).
- [24] A. Omar and A. Ritz, BBN constraints on the hadronic annihilation of sub-GeV dark matter, [arXiv:2510.11791](https://arxiv.org/abs/2510.11791).
- [25] M. Pospelov and J. Pradler, Metastable GeV-scale particles as a solution to the cosmological lithium problem, *Phys. Rev. D* **82**, 103514 (2010).
- [26] R. D. Peccei and H. R. Quinn, *CP* conservation in the presence of instantons, *Phys. Rev. Lett.* **38**, 1440 (1977).
- [27] R. D. Peccei and H. R. Quinn, Constraints imposed by *CP* conservation in the presence of instantons, *Phys. Rev. D* **16**, 1791 (1977).
- [28] S. Weinberg, A new light boson?, *Phys. Rev. Lett.* **40**, 223 (1978).
- [29] F. Wilczek, Problem of strong *P* and *T* invariance in the presence of instantons, *Phys. Rev. Lett.* **40**, 279 (1978).
- [30] J. E. Kim, Weak interaction singlet and strong *CP* invariance, *Phys. Rev. Lett.* **43**, 103 (1979).
- [31] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Can confinement ensure natural *CP* invariance of strong interactions?, *Nucl. Phys.* **B166**, 493 (1980).
- [32] M. Dine, W. Fischler, and M. Srednicki, A simple solution to the strong *CP* problem with a harmless axion, *Phys. Lett.* **104B**, 199 (1981).
- [33] A. R. Zhitnitsky, On possible suppression of the axion hadron interactions. (In Russian), *Sov. J. Nucl. Phys.* **31**, 260 (1980).
- [34] E. Goudzovski *et al.*, New physics searches at kaon and hyperon factories, *Rep. Prog. Phys.* **86**, 016201 (2023).
- [35] S. Dimopoulos, A solution of the strong *CP* problem in models with scalars, *Phys. Lett.* **84B**, 435 (1979).
- [36] S. H. H. Tye, A superstrong force with a heavy axion, *Phys. Rev. Lett.* **47**, 1035 (1981).
- [37] P. Agrawal and K. Howe, Factoring the strong *CP* problem, *J. High Energy Phys.* **12** (2018) 029.
- [38] V. A. Rubakov, Grand unification and heavy axion, *JETP Lett.* **65**, 621 (1997).
- [39] A. Valenti, L. Vecchi, and L.-X. Xu, Grand color axion, *J. High Energy Phys.* **10** (2022) 025.
- [40] B. Holdom and M. E. Peskin, Raising the axion mass, *Nucl. Phys.* **B208**, 397 (1982).
- [41] B. Holdom, Strong QCD at high-energies and a heavy axion, *Phys. Lett.* **154B**, 316 (1985); **156B**, 452(E) (1985).
- [42] M. Dine and N. Seiberg, String theory and the strong *CP* problem, *Nucl. Phys.* **B273**, 109 (1986).
- [43] J. M. Flynn and L. Randall, A computation of the small instanton contribution to the axion potential, *Nucl. Phys.* **B293**, 731 (1987).
- [44] K. Choi, C. W. Kim, and W. K. Sze, Mass renormalization by instantons and the strong *CP* problem, *Phys. Rev. Lett.* **61**, 794 (1988).
- [45] K. Choi and H. D. Kim, Small instanton contribution to the axion potential in supersymmetric models, *Phys. Rev. D* **59**, 072001 (1999).
- [46] T. Gherghetta, V. V. Khoze, A. Pomarol, and Y. Shirman, The axion mass from 5D small instantons, *J. High Energy Phys.* **03** (2020) 063.
- [47] Z. Berezhiani, L. Gianfagna, and M. Giannotti, Strong *CP* problem and mirror world: The Weinberg-Wilczek axion revisited, *Phys. Lett. B* **500**, 286 (2001).

- [48] H. Fukuda, K. Harigaya, M. Ibe, and T. T. Yanagida, Model of visible QCD axion, *Phys. Rev. D* **92**, 015021 (2015).
- [49] A. Hook, S. Kumar, Z. Liu, and R. Sundrum, High quality QCD axion and the LHC, *Phys. Rev. Lett.* **124**, 221801 (2020).
- [50] K. J. Kelly, S. Kumar, and Z. Liu, Heavy axion opportunities at the DUNE near detector, *Phys. Rev. D* **103**, 095002 (2021).
- [51] 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, *J. High Energy Phys.* **09** (2018) 051.
- [52] F. Ertas and F. Kahlhoefer, On the interplay between astrophysical and laboratory probes of MeV-scale axion-like particles, *J. High Energy Phys.* **07** (2020) 050.
- [53] R. Aaij *et al.* (LHCb Collaboration), Search for hidden-sector bosons in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays, *Phys. Rev. Lett.* **115**, 161802 (2015).
- [54] R. Aaij *et al.* (LHCb Collaboration), Search for long-lived scalar particles in $B^+ \rightarrow K^+ \chi(\mu^+ \mu^-)$ decays, *Phys. Rev. D* **95**, 071101 (2017).
- [55] M. Bauer, M. Neubert, S. Renner, M. Schnubel, and A. Thamm, Flavor probes of axion-like particles, *J. High Energy Phys.* **09** (2022) 056.
- [56] G. Aad *et al.* (ATLAS Collaboration), Search for boosted diphoton resonances in the 10 to 70 GeV mass range using 138 fb⁻¹ of 13 TeV pp collisions with the ATLAS detector, *J. High Energy Phys.* **07** (2023) 155.
- [57] A. M. Sirunyan *et al.* (CMS Collaboration), Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV, *J. High Energy Phys.* **01** (2018) 097.
- [58] X. Cid Vidal, A. Mariotti, D. Redigolo, F. Sala, and K. Tobioka, New axion searches at flavor factories, *J. High Energy Phys.* **01** (2019) 113.
- [59] R. Aaij *et al.* (LHCb Collaboration), Search for resonances decaying to photon pairs with masses between 4.9 and 19.4 GeV, *arXiv:2507.14390*.
- [60] D. Aloni, Y. Soreq, and M. Williams, Coupling QCD-scale axionlike particles to gluons, *Phys. Rev. Lett.* **123**, 031803 (2019).
- [61] S. Chakraborty, M. Kraus, V. Loladze, T. Okui, and K. Tobioka, Heavy QCD axion in $b \rightarrow s$ transition: Enhanced limits and projections, *Phys. Rev. D* **104**, 055036 (2021).
- [62] E. Bertholet, S. Chakraborty, V. Loladze, T. Okui, A. Soffer, and K. Tobioka, Heavy QCD axion at Belle II: Displaced and prompt signals, *Phys. Rev. D* **105**, L071701 (2022).
- [63] J. P. Lees *et al.* (BABAR Collaboration), Search for an axionlike particle in B meson decays, *Phys. Rev. Lett.* **128**, 131802 (2022).
- [64] F. Bergsma *et al.* (CHARM Collaboration), Search for axion like particle production in 400-GeV proton—copper interactions, *Phys. Lett.* **157B**, 458 (1985).
- [65] J. Blumlein *et al.*, Limits on neutral light scalar and pseudoscalar particles in a proton beam dump experiment, *Z. Phys. C* **51**, 341 (1991).
- [66] R. Acciarri *et al.* (ArgoNeuT Collaboration), First constraints on heavy QCD axions with a liquid argon time projection chamber using the ArgoNeuT experiment, *Phys. Rev. Lett.* **130**, 221802 (2023).
- [67] Y. Afik, B. Döbrich, J. Jerhot, Y. Soreq, and K. Tobioka, Probing long-lived axions at the KOTO experiment, *Phys. Rev. D* **108**, 055007 (2023).
- [68] E. Cortina Gil *et al.* (NA62 Collaboration), Search for hadronic decays of feebly-interacting particles at NA62, *Eur. Phys. J. C* **85**, 571 (2025).
- [69] D. I. Dunskey, L. J. Hall, and K. Harigaya, Dark radiation constraints on heavy QCD axions, *J. High Energy Phys.* **04** (2024) 130.
- [70] P. F. Depta, M. Hufnagel, and K. Schmidt-Hoberg, Robust cosmological constraints on axion-like particles, *J. Cosmol. Astropart. Phys.* **05** (2020) 009.
- [71] C. Balázs *et al.*, Cosmological constraints on decaying axion-like particles: A global analysis, *J. Cosmol. Astropart. Phys.* **12** (2022) 027.
- [72] D. Cadamuro and J. Redondo, Cosmological bounds on pseudo Nambu-Goldstone bosons, *J. Cosmol. Astropart. Phys.* **02** (2012) 032.
- [73] M. Millea, L. Knox, and B. Fields, New bounds for axions and axion-like particles with keV-GeV masses, *Phys. Rev. D* **92**, 023010 (2015).
- [74] H.-C. Cheng, L. Li, and E. Salvioni, A theory of dark pions, *J. High Energy Phys.* **01** (2022) 122.
- [75] D. Bisht, S. Chakraborty, and A. Samanta, A comprehensive study of ALPs from B -decays, *J. High Energy Phys.* **07** (2025) 092.
- [76] Y. Bai, T.-K. Chen, J. Liu, and X. Ma, Wess-Zumino-Witten interactions of axions: Three-flavor, *arXiv:2505.24822*.
- [77] R. Balkin, T. Coren, Y. Soreq, and M. Williams, A covariant description of the interactions of axion-like particles and hadrons, *arXiv:2506.15637*.
- [78] K. G. Chetyrkin, B. A. Kniehl, M. Steinhauser, and W. A. Bardeen, Effective QCD interactions of CP odd Higgs bosons at three loops, *Nucl. Phys.* **B535**, 3 (1998).
- [79] C. Bierlich *et al.*, A comprehensive guide to the physics and usage of PYTHIA 8.3, *SciPost Phys. Codebases* **2022**, 8 (2022).
- [80] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber, Herwig 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes), *J. High Energy Phys.* **01** (2001) 010.
- [81] M. Bahr *et al.*, Herwig++ physics and manual, *Eur. Phys. J. C* **58**, 639 (2008).
- [82] J. Bellm *et al.*, Herwig 7.0/herwig++ 3.0 release note, *Eur. Phys. J. C* **76**, 196 (2016).
- [83] C. Pitrou, A. Coc, J.-P. Uzan, and E. Vangioni, Precision big bang nucleosynthesis with improved Helium-4 predictions, *Phys. Rep.* **754**, 1 (2018).
- [84] S. Navas *et al.* (Particle Data Group Collaboration), Review of particle physics, *Phys. Rev. D* **110**, 030001 (2024).
- [85] R. Srianand, N. Gupta, P. Petitjean, P. Noterdaeme, and C. Ledoux, Detection of 21-cm, H2 and deuterium absorption at $z > 3$ along the line-of-sight to J1337 + 3152, *Mon. Not. R. Astron. Soc.* **405**, 1888 (2010).
- [86] M. Valerdi, A. Peimbert, M. Peimbert, and A. Sixtos, Determination of the primordial helium abundance based

- on NGC 346, an H II region of the small magellanic cloud, *Astrophys. J.* **876**, 98 (2019).
- [87] V. Fernández, E. Terlevich, A. I. Díaz, and R. Terlevich, A bayesian direct method implementation to fit emission line spectra: Application to the primordial he abundance determination, *Mon. Not. R. Astron. Soc.* **487**, 3221 (2019).
- [88] O. A. Kurichin, P. A. Kislitsyn, V. V. Klimenko, S. A. Balashev, and A. V. Ivanchik, A new determination of the primordial helium abundance using the analyses of H II region spectra from SDSS, *Mon. Not. R. Astron. Soc.* **502**, 3045 (2021).
- [89] T. Hsyu, R. J. Cooke, J. X. Prochaska, and M. Bolte, The PHLEK survey: A new determination of the primordial helium abundance, *Astrophys. J.* **896**, 77 (2020).
- [90] M. Valerdi, A. Peimbert, and M. Peimbert, Chemical abundances in seven metal-poor HII regions and a determination of the primordial helium abundance, *Mon. Not. R. Astron. Soc.* **505**, 3624 (2021).
- [91] E. Aver, D. A. Berg, A. S. Hirschauer, K. A. Olive, R. W. Pogge, N. S. J. Rogers, J. J. Salzer, and E. D. Skillman, A comprehensive chemical abundance analysis of the extremely metal poor Leoncino Dwarf galaxy (AGC 198691), *Mon. Not. R. Astron. Soc.* **510**, 373 (2021).
- [92] A. Matsumoto *et al.*, EMPRESS. VIII. A new determination of primordial he abundance with extremely metal-poor galaxies: A suggestion of the lepton asymmetry and implications for the Hubble tension, *Astrophys. J.* **941**, 167 (2022).
- [93] N. Aghanim *et al.* (Planck Collaboration), Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, A6 (2020).
- [94] A. Salvio, A. Strumia, and W. Xue, Thermal axion production, *J. Cosmol. Astropart. Phys.* **01** (2014) 011.
- [95] F. D'Eramo, F. Hajkarim, and S. Yun, Thermal axion production at low temperatures: A smooth treatment of the QCD phase transition, *Phys. Rev. Lett.* **128**, 152001 (2022).
- [96] F. D'Eramo, F. Hajkarim, and S. Yun, Thermal QCD axions across thresholds, *J. High Energy Phys.* **10** (2021) 224.
- [97] Z. G. Berezhiani, A. S. Sakharov, and M. Y. Khlopov, Primordial background of cosmological axions, *Sov. J. Nucl. Phys.* **55**, 1063 (1992).
- [98] S. Chang and K. Choi, Hadronic axion window and the big bang nucleosynthesis, *Phys. Lett. B* **316**, 51 (1993).
- [99] S. Hannestad, A. Mirizzi, and G. Raffelt, New cosmological mass limit on thermal relic axions, *J. Cosmol. Astropart. Phys.* **07** (2005) 002.
- [100] F. D'Eramo, L. J. Hall, and D. Pappadopulo, Multiverse dark matter: SUSY or axions, *J. High Energy Phys.* **11** (2014) 108.
- [101] M. Kawasaki, M. Yamada, and T. T. Yanagida, Observable dark radiation from a cosmologically safe QCD axion, *Phys. Rev. D* **91**, 125018 (2015).
- [102] R. Z. Ferreira, A. Notari, and F. Rompineve, Dine-Fischler-Srednicki-Zhitnitsky axion in the CMB, *Phys. Rev. D* **103**, 063524 (2021).
- [103] 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).
- [104] E. Masso, F. Rota, and G. Zsembinski, On axion thermalization in the early universe, *Phys. Rev. D* **66**, 023004 (2002).
- [105] P. Graf and F. D. Steffen, Thermal axion production in the primordial quark-gluon plasma, *Phys. Rev. D* **83**, 075011 (2011).
- [106] R. Z. Ferreira and A. Notari, Observable windows for the QCD axion through the number of relativistic species, *Phys. Rev. Lett.* **120**, 191301 (2018).
- [107] F. Arias-Aragón, F. D'eraimo, R. Z. Ferreira, L. Merlo, and A. Notari, Cosmic imprints of XENON1T axions, *J. Cosmol. Astropart. Phys.* **11** (2020) 025.
- [108] W. Giarè, E. Di Valentino, A. Melchiorri, and O. Mena, New cosmological bounds on hot relics: Axions and neutrinos, *Mon. Not. R. Astron. Soc.* **505**, 2703 (2021).
- [109] A. Notari, F. Rompineve, and G. Villadoro, Improved hot dark matter bound on the QCD axion, *Phys. Rev. Lett.* **131**, 011004 (2023).
- [110] F. Bianchini, G. G. di Cortona, and M. Valli, QCD axion: Some like it hot, *Phys. Rev. D* **110**, 123527 (2024).
- [111] https://github.com/Kohsaku-Tobioka/axion_BBN
- [112] O. Pisanti, A. Cirillo, S. Esposito, F. Iocco, G. Mangano, G. Miele, and P. D. Serpico, PArthENoPE: Public algorithm evaluating the nucleosynthesis of primordial elements, *Comput. Phys. Commun.* **178**, 956 (2008).
- [113] R. Consiglio, P. F. de Salas, G. Mangano, G. Miele, S. Pastor, and O. Pisanti, PArthENoPE reloaded, *Comput. Phys. Commun.* **233**, 237 (2018).
- [114] S. Gariazzo, P. F. de Salas, O. Pisanti, and R. Consiglio, PArthENoPE revolutions, *Comput. Phys. Commun.* **271**, 108205 (2022).
- [115] K. Saikawa and S. Shirai, Primordial gravitational waves, precisely: The role of thermodynamics in the standard model, *J. Cosmol. Astropart. Phys.* **05** (2018) 035.
- [116] S. Ganguly, T. H. Jung, and S. Yun, Consistent N_{eff} fitting in big bang nucleosynthesis analysis, [arXiv:2507.23354](https://arxiv.org/abs/2507.23354).
- [117] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (John Wiley and Sons, New York, 1972).
- [118] A. D. Martin and G. G. Ross, K matrix analysis of the low-energy data for K^-p and K_2^0p reactions, *Nucl. Phys.* **B16**, 479 (1970).
- [119] W. E. Cleland, B. Goz, D. Freytag, T. J. Devlin, R. J. Esterling, and K. G. Vosburgh, Measurement of the $K_L^0 - p$ and $K_L^0 - d$ total cross-sections, *Phys. Rev. D* **12**, 1247 (1975).
- [120] G. A. Sayer, E. F. Beall, T. J. Devlin, P. Shepard, and J. Solomon, Measurements of total cross sections for K_2^0 mesons on protons and selected nuclei from 168 to 343 MeV/c and measurement of the K_2^0 mean life, *Phys. Rev.* **169**, 1045 (1968).
- [121] M. Ferro-Luzzi, R. D. Tripp, and M. B. Watson, Excited hyperon of mass 1520 MeV, *Phys. Rev. Lett.* **8**, 28 (1962).
- [122] S. Adhikari *et al.* (GlueX Collaboration), Strange hadron spectroscopy with a secondary KL beam at GlueX, [arXiv:1707.05284](https://arxiv.org/abs/1707.05284).
- [123] P. Capiluppi, G. Giacomelli, G. Mandrioli, A. M. Rossi, P. Serra-Lugaresi, and L. Zitelli, A compilation of $K_0(L)$ p cross-sections.

- [124] T.-G. Lee and C.-Y. Wong, Nuclear annihilation by antinucleons, *Phys. Rev. C* **93**, 014616 (2016); **95**, 029901(E) (2017).
- [125] K. Abazajian *et al.* (CMB-S4 Collaboration), Snowmass 2021 CMB-S4 white paper, [arXiv:2203.08024](https://arxiv.org/abs/2203.08024).
- [126] P. Gondolo and G. Gelmini, Cosmic abundances of stable particles: Improved analysis, *Nucl. Phys.* **B360**, 145 (1991).
- [127] E. Fermi, An attempt of a theory of beta radiation. 1., *Z. Phys.* **88**, 161 (1934).
- [128] R. L. Workman, R. A. Arndt, W. J. Briscoe, M. W. Paris, and I. I. Strakovsky, Parameterization dependence of T matrix poles and eigenphases from a fit to πN elastic scattering data, *Phys. Rev. C* **86**, 035202 (2012).
- [129] R. A. Arndt, I. I. Strakovsky, and R. L. Workman, The SAID PWA program, *Int. J. Mod. Phys. A* **18**, 449 (2003).
- [130] The George Washington University, The SAID PWA program, <https://gwdac.phys.gwu.edu/> (accessed in 2025).
- [131] J. Gasser, V. E. Lyubovitskij, and A. Rusetsky, Hadronic atoms in QCD + QED, *Phys. Rep.* **456**, 167 (2008).
- [132] W. K. H. Panofsky, R. L. Aamodt, and J. Hadley, The gamma-ray spectrum resulting from capture of negative π -mesons in hydrogen and deuterium, *Phys. Rev.* **81**, 565 (1951).
- [133] J. Spuller, D. Berghofer, M. D. Hasinoff, R. Macdonald, D. F. Measday, M. Salomon, T. Suzuki, J. M. Poutissou, R. Poutissou, and J. K. P. Lee, A remeasurement of the Panofsky ratio, *Phys. Lett. B* **67**, 479 (1977).
- [134] T. Flugel, The pion beta decay experiment and a remeasurement of the Panofsky ratio, thesis, Federal Institute of Technology Zurich (ETHZ), 3, 1999.
- [135] R. H. Dalitz and S. F. Tuan, The phenomenological description of $-K$ -nucleon reaction processes, *Ann. Phys. (N.Y.)* **10**, 307 (1960).
- [136] S. Cassel, Sommerfeld factor for arbitrary partial wave processes, *J. Phys. G* **37**, 105009 (2010).
- [137] R. L. Workman *et al.* (Particle Data Group Collaboration), Review of particle physics, *Prog. Theor. Exp. Phys.* **2022**, 083C01 (2022).
- [138] R. Yamartino, G. W. Brandenburg, W. B. Johnson, D. W. G. S. Leith, J. S. Loos, G. Luste, J. A. J. Matthews, K. Moriyasu, W. M. Smart, and F. C. Winkelmann, A study of the reactions $\bar{K}_p^0 \rightarrow \Lambda \pi^+$ and $\bar{K}_p^0 \rightarrow \Sigma^0 \pi^+$ from 1-GeV/c to 12-GeV/c, *Phys. Rev. D* **10**, 9 (1974).
- [139] J. C. M. Armitage *et al.*, A study of $K^0 p$ charge exchange scattering from 0.6-GeV/c to 1.5-GeV/c, *Nucl. Phys.* **B123**, 11 (1977).