Suppressing the QCD axion abundance by hidden monopoles

Masahiro Kawasaki, Fuminobu Takahashi, Masaki Yamada

Abstract

We study the Witten effect of hidden monopoles on the QCD axion dynamics, and show that its abundance as well as isocurvature perturbations can be significantly suppressed if there is a sufficient amount of hidden monopoles. When the hidden monopoles make up a significant fraction of dark matter, the Witten effect suppresses the abundance of axion with the decay constant smaller than $10^{12}$ GeV. The cosmological domain wall problem of the QCD axion can also be avoided, relaxing the upper bound on the decay constant when the Peccei–Quinn symmetry is spontaneously broken after inflation.

1. Introduction

The smallness of the strong CP phase is an outstanding mystery in particle physics, and in particular, it lacks any obvious anthropic explanation. The most natural solution is the Peccei–Quinn (PQ) mechanism, where an anomalous global symmetry is assumed to be broken spontaneously [1]. The associated pseudo-Nambu-Goldstone boson, called the axion, obtains an effective mass via the QCD instanton effect [2,3] and the CP phase is dynamically canceled at the resulting potential minimum. Accordingly, the axion coherent oscillation is necessarily produced by the misalignment mechanism and it contributes to dark matter (DM) [4] (see Refs. [5–7] for recent reviews). However, there are some problems in axion cosmology.

When the PQ symmetry is spontaneously broken before inflation, the abundance of the QCD axion–dark matter depends on the axion decay constant $f_a$ and the initial misalignment angle $\theta_i$ [8]:

$$\Omega_a h^2 \simeq 0.2 \theta_i^2 \left( \frac{f_a}{10^{12} \text{ GeV}} \right)^{1.19},$$

where $h$ is the present-day Hubble parameter in units of $100 \text{ km s}^{-1} \text{Mpc}^{-1}$, and anharmonic effects are neglected. One can see that the observed DM abundance can be naturally explained by the coherent oscillations of the axion with $\theta_i = O(1)$ and $f_a \sim 10^{12}$ GeV. On the other hand, for $f_a \gg 10^{12}$ GeV as suggested by the string theory [9,10], the initial misalignment angle must be finely tuned as $\theta_i \ll 1$ to avoid the overclosure of the Universe. In addition, there is another problem related to the energy scale of inflation. Since the axion is massless during inflation, it acquires quantum fluctuations, $\delta a \simeq H_{\text{inf}}/2\pi$, where $H_{\text{inf}}$ is the Hubble parameter during inflation [11]. As a result, the axion DM has isocurvature fluctuation, which is constrained by observations, setting a tight upper bound on the inflation scale [12]. In particular, there is a strong tension between the axion DM and high-scale inflation. There have been proposed various ways to avoid or ameliorate the isocurvature limit on the axion DM; restoration of the PQ symmetry [13,14], entropy dilution [15–17], time-dependent axion decay constant [13,18–21], stronger QCD in the early Universe [22,23] (see also Ref. [24]), explicit breaking of the PQ symmetry [25–28], and a non-minimal kinetic term of the PQ scalar [29].

When the PQ symmetry is spontaneously broken after inflation, an axionic string and wall system appears at the QCD phase transition [30]. When the domain wall number is unity, the string–wall network collapses due to the tension of domain walls. The axions are produced from those topological defects as well as the misalignment mechanism and the total abundance is given by [31]

$$\Omega_a h^2 \simeq 4 \left( \frac{f_a}{10^{12} \text{ GeV}} \right)^{1.19}.$$

On the other hand, when the domain wall number is greater than unity, the cosmic string and domain wall system is stable and soon

$\Omega_a h^2 \simeq 10^{11}$.
dominates the Universe [32,33]. The resulting Universe would be highly inhomogeneous, so that such scenario is excluded. This is known as the axion domain wall problem.

In this Letter we provide a novel solution to the above cosmological problems based on an axion coupling to monopoles of a hidden U(1)EM gauge symmetry. In the presence of a CP violating θ-term, monopoles acquire a non-zero electric charge and become dyons due to the Witten effect [34]. A non-zero θ costs more energy as the mass of dyon is heavier than the monopole mass. As a result, if θ is replaced with a dynamical axion field, the axion acquires an extra potential. The Witten effect on the QCD axion was studied in Ref. [35], where monopoles were assumed to have an ordinary electromagnetic U(1)EM charge. The effect, however, turned out to be extremely small because of the tight observational constraints on the abundance of the monopoles with U(1)EM charge. Here we focus on hidden magnetic monopoles that are much less constrained, and study their Witten effect on the QCD axion dynamics, also taking account of the adiabatic suppression mechanism as well as various theoretical/cosmological possibilities.

We find that the axion abundance (and therefore the isocurvature perturbation) is suppressed efficiently and the domain wall problem is avoided. Instead of the axion, the hidden monopoles (as well as vector fields) may make up the significant fraction of DM. Interestingly, they have non-negligible self-interactions, which could ameliorate small-scale tensions of the ΛCDM [36].

### 2. The Witten effect on the QCD axion

First, we summarize properties of the axion for later use. The axion obtains an effective mass from QCD instanton effects [3]. At a temperature higher than ΛQCD, it is given by

\[ m_{a,\text{QCD}}(T) \simeq c_T \frac{\Lambda_{\text{QCD}}^4}{T} \left( \frac{T}{\Lambda_{\text{QCD}}} \right)^{-n}, \]

where \( c_T \simeq 1.68 \times 10^{-7} \), \( n = 6.68 \), and \( \Lambda_{\text{QCD}} = 400 \text{ MeV} \) [5,37]. The axion starts to oscillate around the CP conserving minimum at the temperature of

\[ T_{\text{osc},0} \simeq \Lambda_{\text{QCD}} \left( \frac{90c_TM^2_{\Pi}}{\pi^2 g_*(T_{\text{osc},0})T^2} \right)^{1/(4+n)}, \]

\[ \simeq 2.8 \text{ GeV} \left( \frac{f_a}{10^{10} \text{ GeV}} \right)^{-0.187}, \]

where \( M_{\Pi} \simeq 2.4 \times 10^{18} \text{ GeV} \) is the reduced Planck mass. The parameter \( g_*(T) \) is the effective number of relativistic particles in the plasma and we use \( g_*(T_{\text{osc},0}) \approx 85 \). For \( T \lesssim 100 \text{ MeV} \), the axion mass is approximately given by

\[ m_a|_{T=0} \simeq \frac{z}{(1+z)^2} m_\pi f_\pi f_a, \]

where \( z \simeq 0.56 \) is the ratio of u- and d-quark masses, and \( m_\pi \simeq 140 \text{ MeV} \) is the pion mass. The axion coherent oscillation is necessarily induced during the QCD phase transition, and its abundance is given by Eq. (1).

Now let us explain the Witten effect in a hidden Abelian gauge theory [34]. Supposing that there are no charged particles, the Lagrangian is given by

\[ \mathcal{L} = -\frac{i}{4} F_{\mu\nu} F^{\mu\nu} - \frac{e^2\theta}{64\pi^2} \epsilon_{\mu\nu\sigma\rho} F^{\mu\nu} F^{\sigma\rho}, \]

where \( e \) is the gauge coupling constant of the hidden gauge theory. One of the Maxwell’s equation is given by

\[ \partial_\mu \left( F_{\mu\nu} + \frac{e^2\theta}{16\pi^2} \epsilon_{\mu\nu\sigma\rho} F^{\sigma\rho} \right) = 0, \]

so that the Gauss’s law for the electric field is now modified as

\[ \nabla \cdot E + \frac{e^2}{8\pi^2} \nabla \cdot (\theta B) = 0, \]

where \( E_i \equiv F_{0i} \) and \( B_i \equiv -1/2 \epsilon_{ijk} F^{jk} \). Although \( E \) and \( B \) are of the hidden electric and magnetic fields, we use them for notational simplicity. When we introduce a magnetic monopole with a magnetic charge \( g \), the Gauss’s law for the magnetic field is given by

\[ \nabla \cdot B = g (m_{M} - n_{M}), \]

where \( n_{M} = (4\pi/\alpha) \) is the number density of (anti-)monopoles. Then, Eq. (9) implies that the monopole carries also an electric charge \( q \), which is proportional to \( \theta \). In fact, the usual charge quantization condition, \( q/e = n \), is extended to

\[ q + \frac{eg}{8\pi^2} \theta = n, \]

where \( n \) is an integer. The periodicity of \( \theta \rightarrow \theta + 2\pi \) can be seen if one substitutes a magnetic charge of the monopole, \( g = 4\pi\theta/e \).

Thus, the monopole becomes a dyon due to the Witten effect. The mass of dyon mass is heavier than the monopole mass, and so, a non-zero \( \theta \) costs more energy.

The Witten effect on the axion dynamics was studied by Fischer and Preskill [35], where, instead of the \( \theta \) parameter, the axion is coupled to the gauge field as

\[ \mathcal{L}_\theta = -\frac{e^2}{64\pi^2} \alpha \frac{e}{f_a} \epsilon_{\mu\nu\sigma\rho} F^{\mu\nu} F^{\sigma\rho}. \]

Here we have assumed that the domain wall number is unity for the above coupling. The potential energy of a single monopole was estimated to be

\[ V_M \approx \beta f_a \frac{a^2}{J_5}, \]

\[ \beta = \frac{\alpha}{32\pi^2} \frac{1}{r_c f_a}, \]

where \( \alpha \approx e^2/4\pi \), \( a \) denotes the asymptotic field value of the axion, and \( r_c \) is the radius of the monopole core. The origin of \( a \) is chosen so that it coincides with \( \theta = 0 \). When we consider a ‘t Hooft–Polyakov monopole, \( r_c \) is the inverse of the mass of heavy gauge fields, \( m_W \). As a result, the energy density of the axion ground state in a plasma with monopoles and anti-monopoles is given by

\[ U = n_M V_0, \]

where \( n_M = n_{M^+} + n_{M^-} \). Thus, the axion effectively obtains a mass of

\[ m_{2,M}(T) = 2\beta \frac{n_M(T)}{f_a}. \]

The monopole number density \( n_M \) will be evaluated in Sec. 4.

### 3. Axion dynamics

Now let us consider the dynamics of axion in a plasma with monopoles. Once monopoles are produced in thermal plasma, its number density decreases as \( R^{-3} \) where \( R \) is the scale factor. This means that the ratio \( m_{2,M}/R^2 \) increases with time during the radiation dominated era. The Witten effect becomes relevant for the axion dynamics when \( m_{2,M}(T_{\text{osc},1}) \gtrsim R^2(T_{\text{osc},1}) \) is satisfied. Here, the temperature \( T_{\text{osc},1} \) is written as

\[ 1 \text{ in our convention, half-integer electric charges are allowed.} \]

\[ 2 \text{ We define } n_M \text{ as the sum of the number densities of monopoles and anti-monopoles, so that the axion mass squared is different from the one in Ref. [35] by a factor two.} \]
Hereafter, we focus on the case that the axion mass squared Eq. (14) is much larger than the Hubble parameter squared at the time of $T = T_{\text{osc},0}$, i.e., we consider the case of $T_{\text{osc},0} \ll T_{\text{osc},1}$.

Next we consider the cosmological history of axion. If the PQ symmetry is broken before inflation, the axion stays at a certain initial phase until the temperature decreases to $T_{\text{osc},1}$ when the axion starts to oscillate. The resulting axion-number-to-entropy ratio is given by

$$n_a \sim \frac{H_{\text{osc},1} \theta_{ini}^2 f_a^2/2}{s(T_{\text{osc},1})} \approx \frac{45}{32 \pi^2 g_* T_{\text{osc},1} M_{\text{Pl}}} \theta_{ini}^2 f_a^2,$$  

(17)

where $\theta_{ini}$ is an initial phase of axion. On the other hand, if the PQ symmetry is spontaneously broken after inflation, cosmic strings form after the phase transition. When the temperature decreases to $T_{\text{osc},1}$, each cosmic string becomes attached by a domain wall due to the axion mass coming from the Witten effect. Then the cosmic strings and domain walls disappear soon due to the tension of the domain wall.\(^3\) As a result, the axion is produced from the decay of those topological defects and its abundance is expected to be approximately given by Eq. (18) with the replacement of $\theta_{ini} \to 20$, based on the axion abundances (1) and (2). The subsequent evolution does not depend on whether the PQ symmetry is broken before or after inflation.

Around the time of QCD phase transition, the axion mass increases as Eq. (3) due to QCD instanton effects. When $m_{a,QCD}^2(T_{\text{osc},2}) > m_{a,QCD}^2(T_{\text{osc},1})$, the potential minimum of axion changes adiabatically to the vacuum at which the strong CP phase is canceled if $m_{a,QCD}^2 / H_2^2 \gtrsim 10^4$, i.e.,

$$f_a \lesssim 10^{12} \text{GeV} \left( \frac{\Omega_a h^2}{0.12} \right)^{0.55},$$

(19)

Since the potential minimum changes adiabatically, the axion number density (in the comoving volume) is approximately conserved during this epoch and Eq. (18) remains valid even after the QCD phase transition [38,39]. Let us emphasize that we do not have to assume any fine-tuning between the axion VEVs at the potential minima due to the Witten effect and QCD instanton effect. We generically expect $\mathcal{O}(1)$ difference between the two minima, but no extra axion oscillations are induced around the QCD phase transition due to the adiabatic suppression mechanism. The axion abundance is thus given by

$$\Omega_a h^2 \approx 3 \times 10^{-14} \theta_{ini}^2 f_a / T_{\text{osc},1} \approx 3 \times 10^{-4} \theta_{ini}^2 \left( \frac{0.12}{\Omega_a h^2} \right) \left( \frac{f_a}{10^{12} \text{GeV}} \right)^3,$$

(20)

where we use $g_* = 106.75$. Interestingly, the axion abundance is inversely proportional to that of the hidden monopoles. As long as $\mathcal{U}(1)_H$ is unbroken, the monopoles are stable and contribute to DM.

\(^3\) This is not the case if the domain wall number in the interaction (11) is greater than unity. Note that the domain wall number associated with the QCD instanton effects can be greater than unity as the axion potential from the QCD instanton is still negligible at $T = T_{\text{osc}}$.

---

Fig. 1. Relation between the axion and monopole abundances. We take $\theta_{ini} f_a^2 / \alpha^2 = (10^{12} \text{GeV})^3$ (green curve), $2 \times 10^{12} \text{GeV}^3$ (blue curve), and $4 \times 10^{12} \text{GeV}^3$ (red curve). The diagonal (magenta) dashed line represents the observed DM abundance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We plot Eq. (21) in Fig. 1, where we take $\theta_{ini} f_a^2 / \alpha^2 = 10^{12} \text{GeV}^3$ (green curve), $2 \times 10^{12} \text{GeV}^3$ (blue curve), and $4 \times 10^{12} \text{GeV}^3$ (red curve). The observed DM abundance can be explained at the intersection points of each solid curve and the diagonal (magenta) dashed line representing $\Omega_{\text{DM}} h^2 = 0.12$. In order for the abundance of axions and monopoles not to exceed the total DM abundance, the axion decay constant is bounded from above:

$$f_a \lesssim 4 \times 10^{12} \text{GeV} \left( \frac{\alpha}{\theta_{ini}} \right)^{2/3}.$$  

(22)

Since the axion abundance is related to the initial axion angle $\theta_{ini}$, the quantum fluctuation of axion during inflation induces isocurvature modes to the CMB temperature fluctuations. The isocurvature constraint is written as [12]

$$f_a \lesssim 3.4 \times 10^{12} \text{GeV} \left( \frac{\Omega_a}{\Omega_{\text{DM}}} \right)^{1/3} \theta_{ini}.$$  

(23)

Thus, the isocurvature constraint is weakened if we can make the axion abundance smaller than the observed DM abundance. The Witten effect indeed suppresses the axion abundance, thereby relaxing the isocurvature constraint on the decay constant and the inflation scale.

4. Monopole abundance

As an example, let us consider ‘t Hooft–Polyakov monopoles associated with the spontaneous breaking of a hidden SU(2)$_H$ gauge symmetry down to U(1)$_H$ by the vacuum expectation value of an adjoint Higgs ($\phi_H = (0, 0, v)$, where $v$ is the symmetry breaking scale. In this case, there are charged massive gauge field $W^i$ of mass $m_W = ev$ and a massless hidden photon. The mass of magnetic monopole is given by $\sim 4\pi v / \alpha$. Assuming the second order phase transition, the monopole abundance is determined by the Kibble–Zurek mechanism:

$$Y_M \approx 10^{-2} \left( \frac{30 \pi v}{\sqrt{8 \pi M_{\text{Pl}}}} \right)^{3/2 (1 + v)}.$$  

(24)

where $v$ is a critical exponents [40]. At the tree level, $v = 1/2$, and it increases to $v \sim 0.7$ including quantum corrections.

In the minimal set-up without any other light charged particles, the massive gauge bosons $W^\pm$ are stable because they are...
the lightest particles charged under the unbroken $U(1)_{A}$. The abundances of the $W^{\pm}$ and monopoles were evaluated in Refs. [36,41]. In this case, the annihilation of monopoles sets an upper bound on its number density as [36]

$$y_{M}^{\text{max}} = \frac{2\pi B}{g^{2}x_{B}} \sqrt{\frac{45}{4\pi^{3}g_{*}^{3}}} \frac{m_{M}}{\sqrt{8\pi M_{Pl}}}$$  \hspace{1cm} (25)$$

$$B = \frac{6\zeta(3)}{\pi^{2}}$$  \hspace{1cm} (26)$$

where $x_{B} \equiv m_{M}/T$ and $x_{M} = \text{Min}(x_{B}, x_{M})$. The parameter $x_{M}$ is determined by the temperature at which the free streaming length of monopoles exceeds a capture radius of a monopole–anti-monopole bound state, while $x_{B}$ is determined by the temperature at which the massive gauge bosons become non-relativistic.\(^{4}\)

The total abundance of the $W^{\pm}$ and monopoles is consistent with observations for e.g. $\nu \sim 10^{5}$ GeV for $\alpha = O(1)$, where the fraction of the monopole DM is $O(10)\%$. Interestingly, both monopoles and $W$ have non-negligible self-interactions, and the small-scale tensions such as the ‘core-vs.-cusp’ and ‘too-big-to-fail’ problems may be ameliorated [36,41].

Finally, we comment on a kinetic mixing between the hidden and $U(1)_{V}$ gauge bosons. Since we have introduced the adjoint Higgs field $\phi_{h}$ in the hidden sector, we can write the following operator [42]:

$$\phi_{h}^{*} C_{\mu}^{\nu} F_{\mu\nu} \sim \frac{\nu_{h}}{M} C_{\mu}^{\nu} F_{\mu\nu},$$  \hspace{1cm} (27)$$

where $C_{\mu}^{\nu}$ and $F_{\mu\nu}$ are $SU(2)_{h}$ and $U(1)_{V}$ gauge fields, respectively. The parameter $M$ is a cutoff scale. This operator leads to the kinetic mixing of $\nu_{h}/M$, which is of order $10^{-13}$ for $\nu \sim 10^{5}$ GeV and $M = M_{Pl}$. Here we should note that the monopole in the hidden sector has an $O(1)$ electric charge of hidden sector due to the Witten effect. This is because the axion stays at the minimum determined by the QCD instanton effect, which is generically deviated from the one determined by the Witten effect. Through the above kinetic mixing effect, the monopole as well as $W^{\pm}$ acquire a fractional SM electric charge of order $\nu/M$. While the current bound on the mini-charged DM is satisfied for $\nu = 10^{5}$ GeV and $M = M_{Pl}$, the kinetic mixing may provide an interesting probe of such DM candidates.

5. Discussion and conclusions

In this Letter, we have proposed a novel mechanism to suppress the axion abundance based on the Witten effect. If the QCD axion couples to a hidden Abelian gauge field, the axion obtains a large effective mass in a plasma with monopoles. In particular we have focused on the case in which the Witten effect becomes important before the QCD phase transition.

If the PQ symmetry is broken before inflation, the axion starts to oscillate earlier than usual, and the axion abundance is suppressed. If the PQ symmetry is broken after inflation, there also appear axionic domain walls due to the Witten effect. The domain walls are bounded by the cosmic strings associated with the spontaneous breaking of the PQ symmetry. As a result, the string–wall network soon collapses due to the tension of the domain walls.

In either case, the axion abundance is determined when the effective mass becomes comparable to the Hubble parameter before the QCD phase transition. Although the axion potential minimum is shifted to the strong CP conserving one during the QCD phase transition, no extra coherent oscillations are induced as long as the effective mass is much larger than the Hubble parameter. This is because the axion follows its time-dependent minimum adiabatically so that its number density in the comoving volume is conserved. Thus, the final axion abundance can be significantly suppressed compared to the standard scenario.

We have found that the final axion abundance is inversely proportional to the monopole abundance (cf. Eq. (21)). The monopoles are stable and its abundance is bounded above by the observed DM density. The suppression mechanism works when the axion decay constant is smaller than order $10^{12}$ GeV, if the DM is mainly composed of hidden monopoles. The monopoles (as well as the massive vector bosons in the ’t Hooft–Polyakov monopole) have self-interactions, which may relax small-scale problems of $\Lambda$CDM.

We have discussed the ’t Hooft–Polyakov monopole as an explicit example. The observed DM abundance can be explained by both the monopoles and massive gauge bosons for e.g. $\nu \sim 10^{5}$ GeV and $\alpha \sim O(0.1)$, for which the monopoles occupy $O(10)\%$ of the total DM density [36,41]. It is also conceivable that, if the UV (electric) theory becomes strongly coupled, light monopoles appear in the low-energy magnetic dual theory, and they may constitute even larger fraction of the DM.

Finally we mention a possibility that the monopoles decay and disappear after the QCD phase transition. Suppose that the $U(1)_{A}$ gauge symmetry is spontaneously broken by another Higgs field $\phi_{A}$ at an energy scale of $\nu'$ (1 MeV $\lesssim \nu' \lesssim 100$ MeV). (For instance one may consider a doublet under $SU(2)_{h}$.) In this case, cosmic strings form at the phase transition and each monopole and anti-monopole pair is connected by a single cosmic string. This is known as the monopole confinement by the electron condensation in the Abelian gauge theory [43]. This implies that monopoles and anti-monopoles annihilate each other and disappear soon after the phase transition. The produced Higgs field may decay into lighter SM particles through the portal coupling with the SM Higgs field, and such Higgs portal coupling may induce the invisible decay of the SM Higgs. Thus we can avoid the upper bound on the monopole density. Furthermore the massive gauge bosons also decay into the Higgs field $\phi_{A}$. Therefore, there is no remnant in the hidden sector, which implies that the suppression of the axion abundance may be possible for $f_{s} \gg 10^{12}$ GeV. However, the monopole abundance (before the spontaneous break down of $U(1)_{h}$) may be modified because the energy dissipation rate of monopoles depend on the existence of light charged particles. We leave further detailed analysis of such scenario for future work.

Acknowledgements

M.Y. thanks W. Buchmüller for kind hospitality at DESY. This work is supported by MEXT KAKENHI Grant Numbers 15H05889 (M.K. and F.T.) and 23104008 (F.T.), JSPS KAKENHI Grant Numbers 25400248 (M.K.), 24740135 (F.T.), 26247042 (F.T.), and 26287039 (F.T.), JSPS Research Fellowships for Young Scientists (No. 25.8715 (M.Y.), World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan, and the Program for the Leading Graduate Schools, MEXT, Japan (M.Y.).

References

[1] R.D. Peccei, H.R. Quinn, Phys. Rev. Lett. 38 (1977) 1440; R.D. Peccei, H.R. Quinn, Phys. Rev. D 16 (1977) 1791.
[2] S. Weinberg, Phys. Rev. Lett. 40 (1978) 223; F. Wilczek, Phys. Rev. Lett. 40 (1978) 279.
[3] G. ’t Hooft, Phys. Rev. Lett. 37 (1976) 8; G. ’t Hooft, Phys. Rev. D 14 (1976) 3432; G. ’t Hooft, Phys. Rev. D 18 (1978) 2199 (Erratum).
[4] J. Preskill, M.B. Wise, F. Wilczek, Phys. Lett. B 120 (1983) 127;
