# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>Overview of Cosmology</td>
<td>9</td>
</tr>
<tr>
<td>1.1</td>
<td>The expanding Universe</td>
<td>9</td>
</tr>
<tr>
<td>1.1.1</td>
<td>General relativity and Friedmann’s equation</td>
<td>9</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Redshift, distances and horizon</td>
<td>12</td>
</tr>
<tr>
<td>1.2</td>
<td>Big Bang Nucleosynthesis</td>
<td>15</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Thermalization and decoupling</td>
<td>16</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Cosmological constituents</td>
<td>18</td>
</tr>
<tr>
<td>1.2.3</td>
<td>The primordial abundance of the light elements</td>
<td>19</td>
</tr>
<tr>
<td>1.3</td>
<td>Cosmic Microwave Background Radiation</td>
<td>22</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Recombination</td>
<td>22</td>
</tr>
<tr>
<td>1.3.2</td>
<td>Photon decoupling</td>
<td>25</td>
</tr>
<tr>
<td>1.4</td>
<td>CMBR Temperature Anisotropies</td>
<td>27</td>
</tr>
<tr>
<td>1.4.1</td>
<td>Anisotropy observables</td>
<td>27</td>
</tr>
<tr>
<td>1.4.2</td>
<td>Boltzmann and Einstein equations in the perturbed Universe</td>
<td>28</td>
</tr>
<tr>
<td>1.4.3</td>
<td>CMBR anisotropies reveal CDM</td>
<td>35</td>
</tr>
<tr>
<td>1.5</td>
<td>Evidence for CDM from galactic rotation curves</td>
<td>37</td>
</tr>
<tr>
<td>2</td>
<td>Introduction to axion physics</td>
<td>39</td>
</tr>
<tr>
<td>2.1</td>
<td>The strong CP problem</td>
<td>39</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Lagrangian of the standard model</td>
<td>39</td>
</tr>
<tr>
<td>2.1.2</td>
<td>The $U(1)_A$ problem</td>
<td>41</td>
</tr>
<tr>
<td>2.1.3</td>
<td>The $\theta$ Vacuum and Instantons</td>
<td>42</td>
</tr>
<tr>
<td>2.1.4</td>
<td>Solutions to the strong CP problem</td>
<td>45</td>
</tr>
<tr>
<td>2.1.5</td>
<td>Axion models</td>
<td>46</td>
</tr>
<tr>
<td>2.2</td>
<td>Axion properties</td>
<td>49</td>
</tr>
<tr>
<td>2.2.1</td>
<td>The axion mass</td>
<td>49</td>
</tr>
<tr>
<td>2.2.2</td>
<td>The axion couplings</td>
<td>50</td>
</tr>
<tr>
<td>2.3</td>
<td>Axion astrophysics and cosmology</td>
<td>50</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Constraint due to cosmology</td>
<td>50</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Constraint due to astrophysics</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>Axion searches</td>
<td>55</td>
</tr>
<tr>
<td>3.1</td>
<td>Axion dark matter search</td>
<td>55</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Theory and experiment</td>
<td>55</td>
</tr>
<tr>
<td>3.2</td>
<td>Solar axion searches</td>
<td>57</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Solar axion production</td>
<td>58</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Detector using Bragg scattering</td>
<td>60</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Axion helioscope</td>
<td>60</td>
</tr>
<tr>
<td>3.3</td>
<td>Laser experiments</td>
<td>62</td>
</tr>
</tbody>
</table>
3.3.1 Photon regeneration ........................................... 62
3.3.2 Polarization experiments ..................................... 62

4 Axion Bose-Einstein condensation ................................. 64
  4.1 Review of cold axion properties ............................... 64
  4.2 Axion thermalization in the particle kinetic and condensed regimes ........................................... 65
    4.2.1 Evolution equations for nonrelativistic axions ........................................... 65
    4.2.2 The particle kinetic regime .................................... 69
    4.2.3 The condensed regime ........................................ 71
    4.2.4 Cold axions form a BEC ..................................... 74
  4.3 Thermal contact with other species ................................ 74
    4.3.1 Evolution equations for other species .......................... 74
    4.3.1.1 Baryons .................................................. 76
    4.3.1.2 Hot axions ............................................... 77
    4.3.1.3 Photons .................................................. 78
    4.3.2 Possible outcomes ........................................... 78

5 Implications for observation ......................................... 81
  5.1 A non-rethermalizing axion BEC behaves as ordinary CDM ........................................... 81
  5.2 Tidal torquing, inner caustics and axion BEC .................... 84
  5.3 Axion BEC and cosmological parameters ......................... 85
    5.3.1 Possibility of photon cooling .................................. 87
    5.3.2 Effect on the other light element primordial abundances .................. 89
    5.3.3 Effective number of neutrino species ............................ 90

6 Conclusions .......................................................... 92

APPENDIX

A Detection of axion-like particles by interferometry .................. 93
  A.1 Introduction ..................................................... 93
  A.2 Design of experiment ............................................ 96
  A.3 Discussion and conclusions ..................................... 101

B Cosmic ray protons illuminate dark matter axions .................... 104
  B.1 Introduction ..................................................... 104
  B.2 Theoretical analysis of photon emission by protons in a pseudoscalar field .................. 105
    B.2.1 Matrix elements ............................................. 106
    B.2.2 Differential cross section .................................. 108
    B.2.3 Emission rate of the photons ................................ 108
    B.2.4 Energy spectrum of photons ................................ 110
  B.3 Observational Consequences ..................................... 112
    B.3.1 Insufficient sensitivity to detect the QCD axion .................. 112
    B.3.2 Constraining the parameter space of ALPs .......................... 112
I started work on the field of dark matter and cosmology with Dr. Sikivie three years ago with a goal to distinguish observationally axions or axion-like particles (ALPs) from other dark matter candidates such as weakly interacting massive particles (WIMPs) and sterile neutrinos. The subject is exciting because if one can determine the identity of the dark matter, it will be a mile-stone of physics beyond the standard model.

On the high energy frontier, the standard model with three generation fermions is firmly established. However, it is not complete because the theory does not contain a plausible dark matter candidate, with properties required from observation, and the theory has fine-tuning problems such as the strong CP problem.

On the cosmology and astrophysics frontiers, new observations of the dynamics of galaxy clusters, the rotation curves of galaxies, the abundances of light elements, gravitational lensing, and the anisotropies of the CMBR reach unprecedented accuracy. They imply cold dark matter (CDM) is 23% of the total energy density of the universe.

Although many "beyond the standard model" theories may provide proper candidates to serve as CDM particles, the axion is especially compelling because it not only serves as the CDM particle, but also solves the strong CP problem. The axion was initially motivated by the strong CP problem, namely the puzzle why there is no CP violation in the strong interactions. Peccei and Quinn solved the problem by introducing a new $U_{PC}$ symmetry, and later Weinberg and Wilczek pointed out that
the spontaneous breaking of $U_{PC}(1)$ symmetry leads to a new pseudoscalar particle, the axion[1][2][3]. Axion models were proposed in which the symmetry breaking scale may be much larger than the electroweak scale, in which case the axion is very light and couples extremely weakly to ordinary matter. Furthermore, it was realized [4] that the cold axions, produced by the misalignment mechanism during the QCD phase transition, have the right properties to be cold dark matter.

It was also realized that the existence of axions or axion like particles (ALPs) can be probed experimentally by exploiting their coupling to photons[30]. The ADMX experiment is a realization of the concept of the axion haloscope, in which halo axions in an electromagnetic cavity permeated by magnetic field are induced to convert to microwave photons, which may then be picked up by an antenna. The CERN Axion Solar Telescope (CAST) and the Tokyo Helioscope are axion helioscopes which convert axions from the Sun into X-rays in a magnetic field. A third type of experiment is called photon regeneration[99]. In these experiments photons in a laser beam are converted to axions in a magnetic field. The axions travel unimpeded through a wall, behind which is an identical setup of magnets, where some axions are converted back to photons which can be detected.

The three major candidates for CDM, axions/ALPs, WIMPs, and sterile neutrinos, were thought until recently to be indistinguishable by purely astronomical and cosmological observations. However, axions/ALPs are very different from the other two in terms of statistical mechanics properties. Axions/ALPs are spin 0 particles and form a highly degenerate Bose fluid while the typical WIMPs such as neutralinos and sterile neutrinos are fermions and are not degenerate. It was recently found [5] that cold axions are not only a highly degenerate bose fluid, but also form a Bose-Einstein condensate (BEC). Therefore, if the CDM particles are indeed axions/ALPs, there is an opportunity to distinguish them from the other candidates on observational grounds.
In this thesis, chapter 1 is an introduction to cosmology and the evidence for CDM. Chapter 2 introduces axion/ALPs physics. Chapter 3 provides an introduction to the axion/ALPs detection experiments and the constraints on axion/ALP parameter space. Chapter 4 shows quantitatively that cold axions do form a BEC. Chapter 5 discusses the observational consequences of axion BEC. Finally, chapter 6 draws a brief conclusion of the implications of axion BEC. In addition, appendix A investigates a new method to detect the axion like particles and appendix B studies the signals produced by cosmic ray propagating in an axion background.

Qiaoli Yang
Gainesville, Florida
2011
CHAPTER 1
OVERVIEW OF COSMOLOGY

1.1 The expanding Universe

With the discovery of the theory of General Relativity and Hubble’s law, modern cosmology was born. People realized that the universe is expanding and was once much hotter and denser. This picture of an expanding Universe is now called the Big Bang theory and is widely accepted as the standard model of cosmology thanks to abundant observational support.

1.1.1 General relativity and Friedmann’s equation

General Relativity is a geometric theory of gravity. It has two pillars: 1) the metric of space-time determined by the stress-energy tensor of everything present; 2) geodesic motion of all particles moving in that space-time. The Einstein equation is the first pillar:

\[ G^\alpha_\beta = 8\pi G T^\alpha_\beta \]  

(1–1)

where

\[ G^\alpha_\beta = R^\alpha_\beta - \frac{1}{2} \delta^\alpha_\beta R - \Lambda \delta^\alpha_\beta. \]  

(1–2)

\[ R^\alpha_\beta \] is the Ricci tensor, \( R \) is the scalar curvature, and \( \Lambda \) is the cosmological constant.

For a homogeneous isotropic universe, one can show that the metric tensor (FLRW metric) has the general form:

\[ ds^2 = -dt^2 + a^2(t) \left( \frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right), \]  

(1–3)

where \( a(t) \) is a function of time, and \( k \) is a real number. From Eq.(1-3), we can see that the spatial distance between two points is proportional to \( a(t) \), so \( a(t) \) is called the scale factor. (The scale factor describes the evolution of universe in absence of density perturbation.) It can be shown that \( k \) is the spatial curvature. For example, when \( k = 0 \), space is flat.
Provided the content of the universe can be described as a perfect fluid, the stress-energy tensor at RHS of Eq.(1-1) can be written as:

$$T^\alpha_\beta = (\rho + p)\delta^\alpha_\beta + \rho\eta^\alpha_\beta , \quad (1-4)$$

where \(\rho\) and \(p\) are the energy density and pressure. Combining Eq.(1-1), Eq.(1-3), and Eq.(1-4) one gets two independent equations:

$$\frac{3}{a^2}(a^2 + k) = 8\pi G\rho + \Lambda \quad (1-5)$$

and

$$\frac{\dot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \quad (1-6)$$

Eq.(1-5) is called the Friedmann's equation and takes the standard form:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8}{3}\pi G\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \quad (1-7)$$

Eq.(1-5), Eq.(1-6) can be combined to get:

$$\frac{d}{dt}(a^3 \rho) + \rho \frac{da^3}{dt} = 0 \quad (1-8)$$

By combining Eq.(1-8) with equations relating the pressure to the energy density, one gets the dependence of the density upon the scale factor. For dust like matter \(p = 0\) and therefore

$$\rho_m(t) = \rho_{m,0}/a^3 \quad (1-9)$$

For radiation \(p = \rho/3\) and hence

$$\rho_r(t) = \rho_{r,0}/a^4 \quad (1-10)$$

Therefore one may rewrite Eq.(1-6) as:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8}{3}\pi G\left(\frac{\rho_r}{a^4} + \frac{\rho_{m,0}}{a^3}\right) - \frac{k}{a^2} + \frac{\Lambda}{3} \quad (1-11)$$
We see that, as \( a(t) \) increases first the radiation term dominates, then the matter term dominates, then the space-curvature term \( k \), and finally the cosmological constant dominates. The Hubble parameter is defined as: \( H = \dot{a}(t)/a(t) \). Rewrite the Friedmann’s equation as:

\[
  k = \left( \frac{8\pi G}{3} \rho + \frac{\Lambda}{3} - H^2 \right) a^2 = \frac{8\pi G a^2}{3} \left( \rho + \frac{\Lambda}{8\pi G} - \rho_c \right) \tag{1-12}
\]

where \( \rho_c = 3H^2/8\pi G \) is the critical density. Let us define the total energy density:

\[
  \rho_t = \rho + \frac{\Lambda}{8\pi G} \quad \text{and} \quad \Omega = \rho_t/\rho_c.
\]

One can see that:

\[
  \text{if} \quad \Omega > 1, \quad k > 0, \quad \Omega = 1, \quad k = 0, \quad \Omega < 1, \quad k < 0.
\]

Thus the value of the cosmological parameter \( \Omega \) determines the geometry of universe. If \( \Omega > 1 \), the universe is closed; \( \Omega = 1 \) corresponds to a flat universe; and for \( \Omega < 1 \), the universe is open and has hyperbolic geometry. From Eq.(1-12), we can identify three types of energy density which contribute to \( \Omega \): radiation \( \rho_r(t) \); matter \( \rho_m(t) \); and dark energy \( \rho_d = \frac{\Lambda}{8\pi G} \propto \text{constant} \). From observations of CMBR, it is concluded that \( \Omega_t = 1.0023^{+0.0056}_{-0.0054} \). Therefore, we can take \( k = 0 \). Thus:

\[
  \left( \frac{\dot{a}}{a} \right)^2 = \frac{8}{3} \pi G \left( \frac{\rho_r,0}{a^4} \right) \tag{1-13}
\]

for the radiation dominated era in which case the scale factor \( a(t) \propto t^{1/2} \); and

\[
  \left( \frac{\dot{a}}{a} \right)^2 = \frac{8}{3} \pi G \left( \frac{\rho_m,0}{a^3} \right) \tag{1-14}
\]

for the matter dominated era in which case the scale factor \( a(t) \propto t^{2/3} \); and

\[
  \left( \frac{\dot{a}}{a} \right)^2 = \frac{8}{3} \pi G (\rho_\Lambda) \tag{1-15}
\]

for the dark energy dominated era in which case the scale factor \( a(t) \propto e^t \).
1.1.2 Redshift, distances and horizon

Photon propagation satisfies the on-shell condition

\[
0 = -dt^2 + a^2(t) \left( \frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right). \tag{1-16}
\]

Therefore we have:

\[
\frac{dt}{a(t)} = -\frac{dr}{\sqrt{1 - kr^2}}. \tag{1-17}
\]

Consider a comoving source emitting photons of frequency \( \nu = 1/\text{dt} \) from time \( t \) to \( t + \text{dt} \) and the photons reach a comoving observer with frequency \( \nu_0 = 1/\text{dt}_0 \) between time \( t_0 \) and \( t_0 + \text{dt}_0 \). Then it follows:

\[
\int_t^{t_0} \frac{dw}{a(w)} = \int_{t + \text{dt}}^{t_0 + \text{dt}_0} \frac{dw}{a(w)}. \tag{1-18}
\]

Since \( dt \) and \( dt_0 \) are small numbers, we have:

\[
\frac{dt}{dt_0} = \frac{a(t)}{a(t_0)}, \tag{1-19}
\]

which leads to

\[
\frac{\nu_0}{\nu} = \frac{a(t)}{a(t_0)}. \tag{1-20}
\]

In cosmology literature, people use \( z = \frac{\nu_0}{\nu} - 1 \) to denote the redshift. So we find:

\[
1 + z = \frac{a(t_0)}{a(t)}, \tag{1-21}
\]

which shows that the photons are redshifted while they propagate from the source to observer due to the expansion of universe.

On the other hand, one can define the luminosity distance between the source and observer:

\[
d_L = \sqrt{\frac{L}{4\pi B}}, \tag{1-22}
\]

where \( L \) is the emitting power of source, and \( B \) is the observed apparent luminosity.
Consider a source which emits $N$ photons with frequency $\nu$ during time period $dt$,

$$L = \frac{N h \nu}{dt}. \quad (1-23)$$

The observed luminosity at time $t_0$ is:

$$B = \frac{N h \nu}{4 \pi a(t_0)^2 r^2 dt_0}. \quad (1-24)$$

Combining Eq.(1-23) and Eq.(1-24) we have:

$$d_L = a(t_0) r (1 + z), \quad (1-25)$$

which relates $d_L$ and $z$.

Since $dt = da/\dot{a}$, $a(t_0) = 1$ and $a(t_1) = (1 + z)^{-1}$, the relation between $r$ and $z$ is

$$- \int_0^r \frac{dr}{\sqrt{1 - kr^2}} = \int_0^{t_0} \frac{dt}{a(t)} = \frac{1}{a(t_0)} \int_{(1+z)^{-1}}^1 \frac{da}{\dot{a}}. \quad (1-26)$$

Combing the Friedmann equation, Eq.(1-25), Eq.(1-26) and the fact that $\Omega_{r0} \ll 1$ we have:

$$H_0 d_L = \frac{1 + z}{|\Omega_k|^{1/2}} \sin n \left\{ |\Omega_k| \int_0^z \frac{dx}{(1 + x^2)(1 + \Omega_m x) - x(2 + x)\Omega_\Lambda} \right\} \quad (1-27)$$

where $H_0 = \dot{a}(t_0)/a(t_0)$ is the Hubble constant, $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$, and

$$\sin n = \sin, \quad \text{if } k > 0$$

$$= 1, \quad \text{if } k = 0$$

$$= \sinh, \quad \text{if } k < 0. \quad (1-28)$$

If $z$ is small, one may expand the RHS of Eq.(1-27):

$$H_0 d_L = z + \frac{1}{2} (1 + \Omega_\Lambda - \Omega_m/2) z^3 + \ldots \quad (1-29)$$

where the leading term is Hubble's law. When $z$ is not small, $d_L$ and $z$ will no longer be a linear function of $z$. Therefore by fitting the luminosity distance-redshift relation
function with observations, $\Omega_\Lambda$ and $\Omega_m$ can be determined. In the late 90’s Perlmutter, Schmidt, and Riess et al. measured the luminosity distance-redshift relation of type Ia supernovae and got very accurate results. From CMBR, we know that the universe is flat, so $\Omega_m + \Omega_\Lambda = 1$ and $k = 0$. When the distance-redshift curve is fitted, the result shows a dark energy dominated universe, with $\Omega_\Lambda \sim 74\%$.

![Hubble diagram from distant Type Ia supernovae. (From Hicken et al. [6])](image)

Figure 1-1. Hubble diagram from distant Type Ia supernovae. (From Hicken et al. [6])

Another important concept is the event horizon. The event horizon is the distance light traveled, so for events beyond that distance, causal connection is impossible. The size of the event horizon is very important for the evolution of inhomogeneities and structure formation as we will see. The relation between horizon coordinate $r_h$ and photons propagation time $t_0$ is:

$$\int_0^{t_0} \frac{dt}{a(t)} = \int_0^{r_h} \frac{dr}{\sqrt{1 - kr^2}} .$$  \hspace{1cm} (1–30)
So the size of event horizon is \( a(t_0)d_h \):

\[
d_h = a(t_0) \int_0^{t_0} \frac{dr}{\sqrt{1 - kr^2}} = a(t_0) \int_0^{t_0} \frac{dt}{a(t)}.
\]

(1–31)

For example, during matter dominated era, we have \( a(t) \propto t^{2/3} \); so the event horizon is of order \( 3t \).

### 1.2 Big Bang Nucleosynthesis

From the previous section, we see that the dark energy \( \Omega_\Lambda = 74\% \) and the matter \( \Omega_{m0} = 26\% \). It is a natural question to ask what are the constituents of matter \( \Omega_{m0} \) in today’s universe. Part of matter is baryonic. However the energy density of baryonic
matter is order of $4\% \rho_c$ as inferred from big bang nucleosynthesis (BBN). The part of matter $\Omega_m$ that is not baryonic is called "dark matter". Part of dark matter can be massive neutrinos however they are not the major component because they are too "hot" for structure formation as we will see below. So a new kind of matter is required that does not exist in the standard model.

Before going to the details of the cosmic constituents, let us give a brief view of the history of the early universe according to today's understanding:

1. $t = 10^{-43} \text{s} (T = 10^{19} \text{Gev})$ Planck epoch during which the theory of quantum gravity is necessary to understand physics.

2. $t = 10^{-43} - 10^{-7} \text{s} (T = 10^{10} \text{GeV-1GeV})$ General Relativity is valid. However physics beyond the standard model is required to describe phenomena such as the baryon asymmetry and inflation.

3. $t = 10^{-7} \text{s} (T = 1 \text{GeV})$ Quarks and gluons become confined, cold axions are produced.

4. $t = 0.2 \text{s} (T = 2 - 1 \text{MeV})$ Neutrinos decouple and the ratio of neutrons to protons "freezes out".

5. $t = 1 \text{s} (T = 0.5 \text{MeV})$ Electron-positron pairs annihilate, increasing the photon temperature compared to the temperature of neutrinos.

6. $t = 200 - 300 \text{s} (T = 0.05 \text{MeV})$ Nuclear reactions produce light elements.

7. $t = 10^{11} \text{s} (T = 1 \text{eV})$ Matter-radiation equality.

8. $t = 10^{12} - 10^{13} \text{s}$ The universe becomes transparent due to recombination.

9. $t = 10^{16} - 10^{17} \text{s}$ Structure formation, dark energy domination.

### 1.2.1 Thermalization and decoupling

To understand the thermal evolution of universe, one needs to know when particle species become decoupled from the other constituents. The condition of particles
remaining in thermal equilibrium is:

\[ \Gamma >> \frac{1}{\Delta t} \sim H \]  \hspace{1cm} (1–32)

Therefore, thermalization means the particle collision rate is much bigger than the expansion rate of universe. In the particle kinetic regime, the collision rate of particles is:

\[ \Gamma = n < v \sigma(v) > \]  \hspace{1cm} (1–33)

where \( n \) is the particle number density, \( v \) is particle's relative velocity and \( \sigma(v) \) is the cross-section. \(<\cdot\cdot\cdot\>>\) means average over velocities for given temperature. Therefore the collision rate is a function of temperature. For example, the neutrinos are coupled to leptons and baryons via:

\[
\begin{align*}
n + \nu & \leftrightarrow \bar{\nu} + l \\
\bar{\nu} + \bar{l} & \leftrightarrow \nu + l \\
p + \bar{\nu} & \leftrightarrow n + \bar{l} \\
n + \nu & \leftrightarrow p + l .
\end{align*}
\]  \hspace{1cm} (1–34)

The cross-section is \( \sigma \sim G_F^2 T^2 \), where \( G_F \) is the Fermi coupling. The velocity of neutrinos is \( v \sim 1 \) and their particle density is \( n \sim T^3 \). Therefore the collision rate is:

\[ \Gamma \sim G_F^2 T^5 \]. The Hubble rate is \( H = \dot{a}/a \sim 1/t \sim G^{1/2} T^2 \). So we have:

\[ \frac{\Gamma}{H} \sim G_F^2 G^{-1/2} T^3 . \]  \hspace{1cm} (1–35)

From Eq.(1-35) we can see that the ratio of the collision rate to the Hubble rate is decreasing as the temperature drops and we may estimate the decoupling temperature of neutrinos as \( T \sim G_F^{-2/3} G^{1/6} \sim 1 \text{MeV} \).
### 1.2.2 Cosmological constituents

For massless particles, the energy density for given temperature and zero chemical potential is:

\[
\rho = g \int \frac{d^3p}{(2\pi)^3} \frac{p}{e^{p/T} + 1}
\]

where \( g \) is the number of spin degrees of freedom, \( T \) is the temperature, \( - \) for bosons and \(+\) for fermions. (The chemical potential is zero or approximately zero when the particle number is not conserved in the early universe. Furthermore for photons, the CMBR spectrum gives a limit \( \mu / T < 9 \times 10^{-5} \).) After integration one gets:

\[
\rho_b = \frac{\pi^2}{30} g_b T^4 \quad \text{for bosons}\]

\[
\rho_f = \frac{\pi^2}{30} \frac{7}{8} g_f T^4 \quad \text{for fermions}.
\]

Since we already showed that for relativistic particles \( \rho \propto 1/a^4 \), their temperature \( T \propto 1/a \).

The photons are hotter than the neutrinos because annihilation of electrons and positrons inject energy into the photons when the neutrinos have already decoupled. One can calculate the temperature ratio between photons and neutrinos by means of entropy conservation:

\[
\frac{T_\nu}{T_\gamma} = (\frac{g_1}{g_0})^{1/3} = (\frac{4}{11})^{1/3}
\]

where \( g_0 \) and \( g_1 \) are the effective number of degrees of freedom before and after electron-positron annihilation. There are three generations of neutrinos. Therefore the total radiation energy density is:

\[
\rho_r = \frac{\pi^2}{15} \cdot \left( 1 + 3.04 \times 7 \left( \frac{4}{11} \right)^{4/3} \right) T^4.
\]

The photon temperature today is \( T = 2.73 \text{K} \). So we can easily calculate the energy density of radiation today \( \rho_r \sim 4.7 \times 10^{-34} \text{g/cm}^3 \). Since the critical density today is
$\rho_c = \frac{3\hbar^2}{8\pi G} \sim 2 \times 10^{-29}\text{g/cm}^3$, the ratio of radiation energy density to critical density is of order $10^{-5}$.

1.2.3 The primordial abundance of the light elements

The most abundant baryonic matter in the universe is Hydrogen followed by Helium-4 and other light elements such as Deuterium, Helium-3 and Lithium. After the temperature drops to $0.05\text{MeV}$, the primordial nucleosynthesis of light elements begins. Typical reactions are:

$$n + \nu \leftrightarrow p + e, \quad (1-41)$$
$$n + e \leftrightarrow p + \bar{\nu}, \quad (1-42)$$

for production of neutrons, and

$$p + n \leftrightarrow D + \gamma,$$
$$D + D \rightarrow ^{3}\text{He} + n,$$
$$^{3}\text{He} + n \rightarrow ^{3}\text{T} + p,$$
$$^{3}\text{He} + ^{2}D \rightarrow ^{4}\text{He} + p,$$
$$^{4}\text{He} + ^{3}\text{T} \rightarrow ^{7}\text{Li} + \gamma,$$

$$\cdots \quad (1-43)$$

Let us first estimate the ratio of neutrons to protons at the beginning of nucleosynthesis. The neutrons are unstable with a mean life time $\tau = 887\text{s}$, so $n_n / n_p = n_n(t_d) / n_p(t_d)e^{-\Delta t / \tau}$, where $t_d$ is the freeze-out time of neutrons, and $\Delta t \sim 100\text{s}$ is duration between freeze-out time and the beginning of nucleosynthesis.

When the temperature is order of MeV, both protons and neutrons are nonrelativistic particles. The number density of such particles is:

$$n \simeq g \left(\frac{Tm}{2\pi}\right)^{3/2}\exp\left(-\frac{m-\mu}{T}\right)(1 + \frac{15T}{8m}) \quad (1-44)$$
where \( g = 2 \) for protons and neutrons. When the number density of neutrons freezes out, neutrons and protons have the same temperature and chemical potential. Therefore we have:

\[
\frac{n_n(t_d)}{n_p(t_d)} \simeq \left( \frac{m_n}{m_p} \right)^{3/2} \exp\left( -\frac{m_n - m_p}{T_f} \right) \simeq \exp\left( -\frac{m_n - m_p}{T_f} \right)
\]  

(1–45)

where \( T_f \sim 0.8\,\text{MeV} \) is the temperature of neutrinos at their freeze out, and \( m_n - m_p = 1.29\,\text{MeV} \) is the mass difference between neutrons and protons. We conclude that the neutrons to protons ratio is order of \( e^{-1.29/0.8} = 0.199 \) when neutrons freeze-out, and \( X_n \equiv n_n / (n_n + n_p) = 0.166 \). A more detailed calculation shows \( X_n \simeq 0.158 \) [11].

The baryon excess remains after baryon anti-baryon annihilations in the early universe. As the temperature drops below the electron mass, photon collisions no longer produce electron positron pairs. Let us define the ratio of baryon number to the number of photons as \( \eta = n_B / n_\gamma \). It is convenient to introduce \( r_{h0} = 10^{10} \eta \) since \( \eta \) is a very small number.

The collision between neutrons and protons can form deuterons but the deuterons also dissociate due to surrounding high energy photons. After the temperature has dropped sufficiently below the binding energy of the deuteron, the newly formed deuterons are no longer broken by the photons and net deuterons are produced. Once deuterons are produced, \(^3\text{He}, \(^4\text{He}, \) can also be quickly produced by processes described in Eq.(1-43) and finally, stable \(^7\text{Li} \) is produced. There is no primordial production of stable elements heavier than \(^7\text{Li} \).

In summary, primordial nucleosynthesis produces stable elements such as \( D, \(^3\text{He}, \(^4\text{He}, \) and \(^7\text{Li} \). There are in total eight particles involving in the processes: \( n, p, \(^2\text{D}, \(^3\text{T}, \(^3\text{He}, \(^4\text{He}, \(^7\text{Li}, \) and \(^7\text{Be} \). The evolution equations of their number densities are:

\[
dn_i / dt = -3Hn_i + \sum_{a,j} n_a n_j < v \sigma_{aj \rightarrow i} >,
\]  

(1–46)
Figure 1-3. The predicted abundances as functions of $\eta$. From K. A. Olive, G. Steigman, T. P. Walker [8]

where the $n_i$ are the number densities of respective particles. The equations are valid only after the deuterons are no longer dissociated by photons, which time depends on the parameter $\eta_{10}$.

The equations can only be solved numerically. $Y_p \equiv n^{4He}/n_H$, $y_i \equiv n_i/n_H$ are the relative abundances of elements to Hydrogen. From WMAP, Spergel. et al. derive $\eta_{10} =$
6.3 ± 0.3. With this input, one finds that $Y_P = 0.2485 \pm 0.0008$, $y_{Li}/10^{-10} = 4.67 \pm 0.64$, $y_{D}/10^{-5} = 2.45 \pm 0.20$, $y_{H}/10^{-5} = 1.03 \pm 0.04$. The predicted abundances are roughly consistent with observations, but do not fit exactly. We will see in later chapters that the existence of cold axions may provide the key to solve the conflict between observations and predictions.

1.3 Cosmic Microwave Background Radiation

The discovery of cosmic microwave background radiation (CMBR) marked the birth of modern cosmology. The CMBR is radiation left from the hot plasma of the early universe. It provides evidence that the universe was once much hotter and denser.

1.3.1 Recombination

Charged electrons and nucleons become bound between $10^5$ and $3 \times 10^5$ years after the big bang. Radiation decouples from matter then. The main recombination process is:

$$e^- + p \leftrightarrow H + \gamma.$$  \hspace{1cm} (1-47)

The process of (1-47) is reversible which means that for given temperature, both neutral and ionized atoms exist, and the ratio of the two depends on the photon temperature. Let us define the ionization ratio:

$$X_p = \frac{n_p}{n_p + n_H},$$  \hspace{1cm} (1-48)

where for simplicity we neglect other light elements such as He. $1 - X_p$ is the neutral hydrogen percentage. When $X_p$ is small enough, one can regard the recombination process complete. Notice that the baryon to photon ratio $\eta = \frac{n_B}{n_\gamma}$ is a constant since $n_N$ and $n_\gamma$ both decrease as $a^{-3}(t)$, so $\eta \sim 10^{-10}$ today as well as in the early universe. Now, let us find the relationship between $X_p$ and $T$.  

22
Figure 1-4. The conflict between observations and predictions. Solid curves show predictions of the standard big bang nucleosynthesis, solid and dashed vertical lines indicate the value of $\eta$ and its 1 $\sigma$ deviations from WMAP. Boxes show the observed light element abundances and their 1 $\sigma$ deviations. From Y. I. Izotov, T. X. Thuan and G. Stansinska[9]
In the Saha approximation, protons, free electrons and neutral hydrogen atoms satisfy the Boltzmann distribution:

\[
\begin{align*}
    n_p &= 2 \left( \frac{m_p T}{2\pi} \right)^{3/2} e^{\frac{\mu_p - m_p}{T}} \quad (1-49) \\
    n_e &= 2 \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{\frac{\mu_e - m_e}{T}} \quad (1-50) \\
    n_H &= 4 \left( \frac{m_H T}{2\pi} \right)^{3/2} e^{\frac{\mu_H - m_H}{T}}. \quad (1-51)
\end{align*}
\]

where \( \mu_i \) are their chemical potentials, \( m_i \) are the particle masses; \( \mu_p + \mu_e = \mu_H \), \( n_p = n_e \). These equations assume thermal equilibrium, which at later time of recombination is no longer accurate, One needs a more delicate method such as numerical simulation to completely describe the recombination process. Using the definition of \( \eta \) and photon number density \( n_\gamma = 2.4 \frac{T^3}{\pi^2} \), we have:

\[
    n_p + n_H = \eta n_\gamma = \eta \left( \frac{2.4}{\pi^2} T^3 \right). \quad (1-52)
\]

Combining (1–49), (1–51) and \( \mu_p + \mu_e = \mu_H \) we have:

\[
    \frac{n_p^2}{n_H} = \frac{n_p n_e}{n_H} = \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{\frac{m_p + m_e - m_H}{T}}. \quad (1-53)
\]

From (1–64), (1–66) one gets:

\[
    \frac{1 - X_p}{X_p^2} = 1.1 \times 10^{-8} e^{13.6 eV / T} \eta (T/eV)^{3/2}. \quad (1-54)
\]

(1–54) tells us explicitly the relation between photon temperature and ionization fraction of matter. Using the WMAP value of \( \eta \) we find \( X_p \) is 0.1 when \( T \sim 0.3 eV \), therefore one can regard the recombination complete at \( T_{rec} = 0.3 eV \). Since:

\[
    1 + z = a(t_0)/a(t) = T / T_0, \quad (1-55)
\]

and \( T_0 = 2.73 K = 2.35 \times 10^{-4} eV \), we have \( z_{rec} = 1251. \)
Figure 1-5. Free electron fraction $X_e$ as a function of redshift. The solid line is the exact solution and the dashed line is the Saha approximation. $X_e \approx X_p$. From S. Dodelson[10]

1.3.2 Photon decoupling

In the last section, we discussed the recombination of charged particles. The photons can propagate freely once recombination is complete. There are two major processes involved:

$$\gamma + e^- \rightarrow e^- + \gamma$$
$$\gamma + p^+ \rightarrow p^+ + \gamma.$$  \hspace{1cm} (1–56)

Since the Thomson scattering cross-section $\sigma$ is inversely proportional to the mass squared, only photon-electron scattering matters here. The collision rate of photons and
electrons is:

\[ \Gamma = cn_e \sigma, \tag{1–57} \]

where \( c = 1 \) is the speed of light, and \( \sigma = 1.7 \times 10^3 \text{GeV}^{-2} \). During recombination, the number density of free electrons drops down, so the collision rate decreases as well. When the collision rate is smaller than the hubble rate:

\[ \Gamma / H < 1, \tag{1–58} \]

the photons decouple from matter.

Since the recombination happens during the matter dominated era, we can write the Friedmann equation as:

\[ H^2 = \frac{8\pi G}{3} \rho_m = H_0^2 \Omega_{m0} \frac{\rho_m}{\rho_{m0}}, \tag{1–59} \]

where \( \rho_{m0} \) is the matter density today. Using \( \rho_m/\rho_{m0} = T^3/T_0^3 \), we can rewrite (1–59) as:

\[ H^2 = \Omega_{m0} H_0^2 (T/T_0)^3. \tag{1–60} \]

Putting numbers to (1–60) one gets:

\[ H = 8.1 \times 10^{-43}(\frac{T}{T_0})^{3/2} \text{GeV}. \tag{1–61} \]

(1–57) can also be written in terms of \( T, T_0 \):

\[ \Gamma = X_p \eta n_e \sigma = 5.4 \times 10^{-36} X_p \eta (\frac{T}{T_0})^3 \text{GeV}. \tag{1–62} \]

Therefore, by using (1–61), (1–62) and \( \Gamma / H < 1 \) we find that photons decouple when \( X_p \sim 4 \times 10^{-3} \) which is equivalent to \( T_{\text{dec}} = 0.25 \text{eV} \) or \( z_{\text{dec}} = 1061 \). After that photons propagate freely. The photon frequency spectrum is a Planck distribution because they are in thermal equilibrium with the matter plasma at the last scattering surface. Therefore when \( z = z_{\text{dec}} \), one expects the spectrum of photons to be: \( n(\nu) = 2\nu^3(e^{\nu/T_{\text{dec}}} - 1)^{-1} \). The frequency of photons redshifts according to \( \nu = (1 + z_{\text{dec}}) \nu_{\text{dec}} \).
The CMB spectrum observed by COBE shows a perfect Planck distribution. Uncertainties are a small fraction of the line thickness. From D. J. Fixsen et al.[13]

Therefore the effective temperature of photons today (see Fig. 1-6) is:

\[ T_0 = \frac{T_{\text{dec}}}{1 + z_{\text{dec}}}. \]  (1–63)

Penzias and Wilson first observed this relic radiation of the early universe. They got the 1978 Nobel Prize in physics.

1.4 CMBR Temperature Anisotropies

1.4.1 Anisotropy observables

The observed CMBR effective temperature \( T \) can be written as a function of direction \( T[\hat{n}(\theta, \phi)] \). The fluctuation of the temperature is: \( T = T_0 + \Delta T(\hat{n}) \), where
Let us define:
\[ \Theta(\hat{n}) \equiv \frac{\Delta T(\hat{n})}{T_0}. \] (1–64)

\( \Theta(\hat{n}) \) can be expanded into spherical harmonics:
\[ a_{lm} = \int d\Omega Y_{lm}^*(\hat{n}) \Theta(\hat{n}). \] (1–65)
The \( a_{lm} \) have the property:
\[ \langle a_{lm} a_{l'm'}^* \rangle = \delta_{ll'} \delta_{mm'} C_i, \] (1–66)
where we assume that the fluctuations have statistical isotropy, so the \( C_i \) are independent of \( m \). From (1–64) - (1–66) we have:
\[ \langle \Delta T(\hat{n}_1) \Delta T(\hat{n}_2) \rangle = T_0^2 \sum_l\frac{2l + 1}{4\pi} C_l P_l(\cos \theta), \] (1–67)
where \( \theta \) is the angular separation between \( \hat{n}_1 \) and \( \hat{n}_2 \); the \( P_l(x) \) are the Legendre polynomials.

### 1.4.2 Boltzmann and Einstein equations in the perturbed Universe

In chapters 1.1-1.3 we discussed the evolution of the homogeneous universe. To discuss anisotropies of the CMBR, one needs to add perturbations to the energy-stress tensor and to the metric. The decomposition theorem states that the perturbations can be divided into scalar, vector and tensor, and that each type of perturbation evolves independently. The vector and tensor perturbations play a subdominant role in structure formation. Therefore we only consider scalar perturbations here. It can be shown that for scalar perturbations the metric can always be written in the form:
\[ ds^2 = -\ddot{a}^2(\eta)(1 + 2\psi)d\eta^2 + (1 + 2\phi)d\vec{x}^2, \] (1–68)
where \( \eta \) is conformal time which is defined by: \( \eta = \int dt/a(t) \), \( a \) is the scale factor, \( \phi \) is the Newtonian potential and \( \psi \) is the perturbation to the spatial curvature. This choice of metric is called "conformal Newton gauge".
Let us consider the Boltzmann equation for the photons first. The Boltzmann equation is

\[ \frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} dx^i + \frac{\partial f}{\partial p} dp + \frac{\partial f}{\partial \hat{p}^i} d\hat{p}^i, \]

(1–69)

where \( \hat{p} \) is unit vector in the direction of momentum \( \vec{p} \). Since both \( \frac{\partial f}{\partial \hat{p}^i} \) and \( \frac{d\hat{p}^i}{dt} \) are first order, we can neglect the last term of 1–69, as it is of second order.

For photons \( p^2 = 0 \) and hence \( dx^i / dt = p^i / p^0 = \hat{p}^i \sqrt{1 + \psi / a \gamma} \) \( \approx \hat{p}^i (1 + \psi - \phi) / a \). The geodesic equation leads to \( dp / p dt = -H + \partial \phi / \partial t - \hat{p}^i \partial \psi / a \partial x^i \). So we have:

\[ \frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} \frac{\hat{p}^i}{a} - p \frac{\partial f}{\partial p} \left( H + \frac{-\partial \phi}{\partial t} + \frac{\hat{p}^i \partial \psi}{a \partial x^i} \right), \]

(1–70)

where we have neglected \( \partial f / \partial x^i \hat{p}^i / a \) since it is a higher order term. Let us assume the perturbed photon phase space distribution function \( f \) may be written:

\[ f(\vec{x}, \vec{p}, \eta) = \left[ \exp \frac{\vec{p}}{T(\eta)} \left[ 1 + \Theta(\vec{x}, \vec{p}, \eta) \right] - 1 \right]^{-1}, \]

(1–71)

where the perturbation is characterized by \( \Theta \), and \( \Theta \) is function of \( \vec{x}, \eta, \vec{p} \).

By expanding the distribution function near \( \Theta = 0 \) we have:

\[ f = f^0 - \Theta(\vec{x}, \vec{p}, \eta) p \frac{\partial f^0}{\partial p}, \]

(1–72)

where \( f^0 = \left( \exp[\vec{p} / T] - 1 \right)^{-1} \). Putting 1–72 to 1–70, and using \( T \partial f^0 / \partial t = -p \partial f^0 / \partial p \), \( T \propto 1 / a \), one finds:

\[ \frac{df}{dt} = -p \frac{\partial f^0}{\partial p} \left[ \frac{\partial \Theta}{\partial t} + \frac{\hat{p}^i \partial \Theta}{a \partial x^i} + \frac{\partial \phi}{\partial t} + \frac{\hat{p}^i \partial \psi}{a \partial x^i} \right], \]

(1–73)

up to first order.
For the \( C[f] \) terms, we have\[17 \] [18]:

\[
C[f] = \sum_{\vec{m}\vec{q}} |\text{Amplitude}|^2 \{ f(\vec{q}) f(\vec{p}) (1 - f_e(\vec{q}))(1 + f(\vec{m})) - f(\vec{q}) f(\vec{m})(1 - f_e(\vec{q}))(1 + f(\vec{p}))\}
\]

\[
= \frac{1}{2\rho} \int \frac{d^3q d^3q_1 d^3p_1}{(2\pi)^6 E(q) E(q_1) E(p_1)} |M|^2 (2\pi)^4 \delta^4(\rho + q - p_1 - q_1) F, \tag{1-74}
\]

where \( |M|^2 \) is the matrix element of Thompson scattering and

\[
F = (1 - f_e(\vec{q}))(1 + f(\vec{p})) f_e(\vec{q}_1) f(\vec{p}_1) - (1 - f_e(\vec{q}_1))(1 + f(\vec{p}_1)) f_e(\vec{q}) f(\vec{p})
\approx [f_e(\vec{q}_1) f(\vec{p}_1) - f_e(\vec{q}) f(\vec{p})]. \tag{1-75}
\]

since \( f_e \) (electron distribution function) is much less than 1. \( f_e(\vec{q}_1) \approx f_e(\vec{q}) \) for the epoch of interest since photon momentum is much smaller than that of electrons. The \( \delta^3 \) function can be easily integrated since \( E_e \approx m_e \) and \( |M|^2 \) is a constant:

\[
C = \frac{\pi}{8m_e^2 \rho} \int D\vec{q} D\vec{p}_1 \frac{1}{\rho_1} \delta[p + \frac{q^2}{2m_e} - p_1 - \frac{(\vec{q} + \vec{p} - \vec{p}_1)^2}{2m_e}] |M|^2 (f_e(\vec{q} + \vec{p} - \vec{p}_1) f(\vec{p}_1) - f_e(\vec{q}) f(\vec{p})). \tag{1-76}
\]

where \( D\vec{q} = d^3q/(2\pi)^3 \). Next since \( f_e(\vec{q} + \vec{p} - \vec{p}_1) \approx f_e(\vec{q}) \) we have:

\[
C = \frac{\pi}{8m_e^2 \rho} \int D\vec{q} D\vec{p}_1 \frac{1}{\rho_1} f_e(q) |M|^2 [\delta(p - p_1) + \frac{(\vec{p} - \vec{p}_1) \cdot \vec{q} d\delta(p - p_1)}{m_e} f(\vec{p}_1) - f(\vec{p})], \tag{1-77}
\]

where we expand the \( \delta \) function around \( p - p_1 \) and \( q^2 - (\vec{q} + \vec{p} - \vec{p}_1)^2 \approx 2(\vec{p}_1 - \vec{p}) \cdot \vec{q} \) for the scattering. Putting \( |M|^2 = 8\pi \sigma_T m_e^2 \) and noticing that \( f_e(\vec{q}) \approx n_e \delta(\vec{q} - m_e \vec{v}_b) \), (\( \vec{v}_b \) is the bulk velocity of the electrons) we conclude:

\[
C = \frac{\pi^2 n_e \sigma_T}{p} \int D\vec{p}_1 \frac{1}{\rho_1} [\delta(p - p_1) + \frac{\vec{p} \cdot \vec{v}_b d\delta(p - p_1)}{dp_1} ] (f(\vec{p}_1) - f(\vec{p})), \tag{1-78}
\]

where \( n_e \) is electron number density. Let us define \( \Theta_0 = 1/(4\pi) \int d\Omega \Theta(\hat{p}, \vec{x}, t) \).

Combining (1–78), (1–72), we have:

\[
C = \frac{n_e \sigma_T}{p} \int \rho_1 d\rho_1 [\delta(p - p_1)(-\rho_1 \frac{\partial f^0_0}{\partial \rho_1} \Theta_0 + \rho \frac{\partial f^0_0}{\partial p} \Theta(\hat{p})) + \vec{p} \cdot \vec{v}_b \frac{d\delta(p - p_1)}{dp_1} (f^0_0(p_1) - f^0_0(p))]. \tag{1-79}
\]
Then integrating the second term by parts we finally obtain:

$$C = -\rho \frac{\partial f^0}{\partial \rho} n_e \sigma T \left[ -\Theta(\hat{p}, \vec{x}, \eta) + \Theta_0 + \hat{p} \cdot \vec{\nu}_b \right].$$  \hfill (1–80)

Therefore we can write the first order Boltzmann equation for photons:

$$\frac{\partial (\Theta + \phi)}{\partial \eta} + \vec{p} \cdot \vec{\nabla} (\Theta + \psi) = n_e \sigma T a [\Theta_0 - \Theta + \hat{p} \cdot \vec{\nu}_b],$$  \hfill (1–81)

where we use conformal time $\eta$ for convenience. It is easier to solve this linear partial differential equation in its Fourier modes, since different modes are decoupled. Let us define $\Theta(x, \hat{p}, \eta) = \int d^3k/(2\pi)^3 e^{i\vec{k} \cdot \vec{x}} \Theta(\vec{k}, \hat{p}, \eta)$, and assume $\Theta(\vec{k}, \hat{p}, \eta) = \Theta(\vec{k}, \mu, \eta)$ where $\mu = \vec{k} \cdot \hat{p}$. In first order for scalar perturbations, $\vec{\nu}_b(\vec{k}) = \vec{k} \nu_b(\vec{k})$. Then we have:

$$d\Theta(\vec{k}, \mu, \eta)/d\eta + ik\mu \Theta + d\phi(\vec{k}, \eta)/d\eta + ik\mu \psi(\vec{k}, \eta) = n_e \sigma T a [\Theta_0 - \Theta + \mu \nu_b(\vec{k})].$$  \hfill (1–82)

The Boltzmann equation for neutrinos can be obtained by following similar steps as for the photons but with $C = 0$ and $f = [\exp(\rho/\Theta) + 1]^{-1}$, noticing that for the epoch of interest the neutrinos are relativistic. Therefore the equation of neutrinos is:

$$d\nu/d\eta + ik\mu \nu + d\phi(\vec{k})/d\eta + ik\mu \psi(\vec{k}) = 0,$$  \hfill (1–83)

where $\nu$ is defined similarly as $\Theta$ for photons.

Cold dark matter and baryons are non-relativistic during the era we are interested in. For non-relativistic particles, we do not need to assume a particular form of thermal distribution function because the thermal motion of the particles, which is order of $T \sim m^2 \ll 1 * m$, can be neglected. We use only $n(\vec{x}, t)$ and $\vec{v}(\vec{x}, t)$ and the corresponding equations of motion to describe the evolution of such particle systems.

The dark matter particles are massive. So we have: $g_{\mu\nu} p^\mu p^\nu = -m^2$, which gives a constraint on the four momenta. Let us choose $E$ and $\hat{p}$ as independent variables so:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} \frac{dx^i}{dt} + \frac{\partial f}{\partial E} \frac{dE}{dt} + \frac{\partial f}{\partial \hat{p}^i} \frac{d\hat{p}^i}{dt},$$  \hfill (1–84)
where the last term can be neglected since it is of second order. By using the on-shell condition and geodesic equation for massive particles, we get:

\[
\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{x}} \frac{\mathbf{p}}{aE} - \frac{\partial f}{\partial E} \left( \frac{\mathbf{p}^2}{E} + \frac{\mathbf{p} \partial \phi}{E \partial t} + \frac{p \partial \psi}{a \partial x^i} \right). \tag{1–85}
\]

We can multiply 1–85 by the phase space volume and integrate:

\[
\int \frac{d^3 p}{(2\pi)^3} \left[ \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{x}^i} \frac{\mathbf{p}^i}{aE} - \frac{\partial f}{\partial E} \left( \frac{\mathbf{p}^2}{E} + \frac{\mathbf{p} \partial \phi}{E \partial t} + \frac{p \partial \psi}{a \partial x^i} \right) \right] = 0. \tag{1–86}
\]

For the first term in 1–86 we use \( n = \int [d^3 p/(2\pi)^3] f(\mathbf{p}/E) \) for the second term. Since \( dE/dp = p/E \), the third term can be integrated by part:

\[
\int [d^3 p/(2\pi)^3] |p^2/E| \delta f/\delta E = \int [d^3 p/(2\pi)^3] |p^2/E| \partial f/\partial p = -3n. \tag{1–87}
\]

Neglecting all terms of order \( \mathbf{x}^i \) or higher, we have the first moment of the Boltzmann equation:

\[
\frac{\partial n}{\partial t} + \frac{1}{a} \frac{\partial (nv^i)}{\partial x^i} + 3[H + \frac{\partial \phi}{\partial t}]n = 0. \tag{1–88}
\]

Letting \( n_{dm} = n_{dm}^0[1 + \delta(\mathbf{x}, t)] \), one gets the zero order and first order equations for cold dark matter:

\[
\frac{\partial n_{dm}^0}{\partial t} + 3Hn_{dm}^0 = 0 \tag{1–89}
\]

\[
\frac{\partial \delta}{\partial t} + \frac{1}{a} \frac{\partial \nu^i}{\partial x^i} + 3 \frac{\partial \phi}{\partial t} = 0. \tag{1–90}
\]

Assuming \( \nu(\mathbf{k}) = \hat{\kappa} \nu(\mathbf{k}) \), we have the first order equation in \( k \) space

\[
\frac{d \delta}{d \eta} + ik \nu + 3 \frac{d \phi}{d \eta} = 0. \tag{1–91}
\]

In similar fashion, we can multiply 1–85 by \( d^3 \rho(\mathbf{p}/E) \hat{\mathbf{p}} / (2\pi)^3 \) and integrate. Neglecting all terms of order \( (\mathbf{p}/E)^2 \) or higher we have the first moment of the Boltzmann equation:

\[
\frac{\partial (nv^i)}{\partial t} + 4Hnv^i + \frac{n}{a} \frac{\partial \psi}{\partial x^i} = 0. \tag{1–92}
\]
Combined with the zero order equation, 1–91 leads to:

$$\frac{\partial \nu^i}{\partial t} + H \nu^i + \frac{1}{a} \frac{\partial \psi}{\partial x^i} = 0,$$  \hspace{1cm} (1–92)

or in Fourier space

$$\frac{dv}{d\eta} + \frac{d\alpha}{ad\eta} v + i k \psi = 0.$$  \hspace{1cm} (1–93)

This equation together with 1–90 completes the Boltzmann equations for cold dark matter.

The electrons and protons are generally named "baryons" in cosmology literature although electrons are actually leptons. Let us define their perturbations:

$$\frac{\rho_e - \rho_e^0}{\rho_e^0} = \frac{\rho_p - \rho_p^0}{\rho_p^0} \equiv \delta_b.$$  \hspace{1cm} (1–94)

Here we used the fact that the Coulomb scattering rate is much larger than the expansion rate at all epochs of interest, so electrons and protons are tightly coupled.

The same reasoning also implies $\vec{\nu}_e = \vec{\nu}_p \equiv \vec{\nu}_b$.

The left hand sides of 1–90 and 1–93 are the same for "baryons" as for cold dark matter since both of them are non-relativistic. The right hand sides are the relevant collision terms. Let $C_{e\gamma}[f]$ denote electron-photon scattering. It is given by 1–74.

Proton-photon scattering is negligible because their Compton cross-section is much smaller. Let $C_{ep}[f]$ term denote electron-proton scattering. It can be written as 1–74 with Coulomb scattering amplitude $|M|_{ep}$. Because protons are much heavier than electrons, the $F$ term can be simplified to 1–75 by using a similar argument as for electron photon scattering. The Boltzmann equations for electrons and protons are respectively:

$$df_e/dt = C_{e\gamma}[f] + C_{ep}[f]$$  \hspace{1cm} (1–95)

$$df_p/dt = C_{pe}[f].$$  \hspace{1cm} (1–96)
Let $<>$ denote integration over the momentum appearing in the $C[f]$ term. Since $< C[f] >= 0$, the continuity equation for "baryons" is

$$\frac{d\delta_b}{d\eta} + i k \nu_b + 3 \frac{d\phi}{d\eta} = 0. \quad (1-97)$$

For the first moment of the Boltzmann equation, we first add 1–95 to 1–96 in order to annihilate the Coulomb scattering term after integration. The remaining $< C_{e, f}^i \rho^j >$ term can be calculated by using 1–80, which we rewrite here for convenience.

$$C = -p \frac{\partial f^0}{\partial p} n_0 \sigma_{\tau} \left[ -\Theta(\hat{p}, \vec{x}, \eta) + \Theta_0 + \hat{p} \cdot \vec{v}_b \right]. \quad (1-98)$$

In Fourier space it is:

$$C = -p \frac{\partial f^0}{\partial p} n_0 \sigma_{\tau} \left[ -\Theta(\hat{p}, \vec{k}, \eta) + \Theta_0 + \mu \nu_b(\vec{k}) \right]. \quad (1-99)$$

We have:

$$< C[f] \rho^j \cdot \hat{k} >= < C[f] \rho_{\mu} >= -n_0 \sigma_{\tau} \int \frac{dp}{2\pi^2} \rho^4 \frac{\partial f^0}{\partial p} \int_{-1}^{1} \frac{d\mu}{2} \mu \left[ \Theta_0 - \Theta(\mu) + \nu_b \mu \right]$$

$$= n_0 \sigma_{\tau} 4 \rho_\gamma \int_{-1}^{1} \frac{d\mu}{2} \mu \Theta(\mu) + \nu_b/3). \quad (1-100)$$

Define $\Theta_1 = i/2 \int d\mu \mu \Theta(\mu)$. Finally we have the first momentum of the Boltzmann equation for baryons:

$$\frac{d\nu_b}{d\eta} + \frac{da}{d\eta} \nu_b + i k \psi = -n_0 \sigma_{\tau} \frac{4 \rho_\gamma}{3 \rho_b} [3i \Theta_1 + \nu_b], \quad (1-101)$$

where we use conformal time for convenience.

The Einstein Equation is $G^\mu_{\nu} = 8\pi GT^\mu_{\nu}$. We can calculate the perturbed Einstein tensor in terms of $\psi \phi$. For the time-time component in $k$ space

$$\delta G^0_0 = -6H \partial \psi / \partial t + 6\psi H^2 - 2k^2 \phi / a^2. \quad (1-102)$$
The time-time component of the energy-momentum tensor is

\[ T^0_0 = - \sum_i g_i \int \frac{d^3 \rho}{(2\pi)^3} E_f(\bar{\rho}, \bar{x}, \eta) \]

\[ = - \rho_\gamma^0[1 + 4\Theta_0] - \rho_\nu^0[1 + 4\nu_0] - \rho_{dm}^0[1 + \delta] - \rho_b^0[1 + \delta_b]. \tag{1–103} \]

so we can write the perturbed time-time component of the Einstein Equation:

\[ k^2 \psi + 3\frac{\dot{a}}{a}(\dot{\phi} - \dot{\psi} \frac{\dot{a}}{a}) = 4\pi G a^2[4\rho_\gamma^0 \Theta_0 + 4\rho_\nu^0 \nu_0 + \rho_{dm}^0 \delta + \rho_b^0 \delta_b], \tag{1–104} \]

where \( \nu_0 \) is defined similarly to \( \Theta_0 \), \( \dot{a} = da/d\eta \).

We now consider the spatial part of the Einstein equations. The longitudinal traceless part of \( G^j_j \) in \( k \) space is:

\[ (\hat{k} \cdot \hat{k}' - \delta^j_j/3)G^j_j = \frac{2}{3a^2}k^2(\phi + \psi). \tag{1–105} \]

Then same part of \( T^j_j \) is:

\[ (\hat{k} \cdot \hat{k}' - (1/3)\delta^j_j)T^j_j = g_i \int \frac{d^3 \rho}{(2\pi)^3} \frac{\mu^2 \rho^2}{E(\rho)} (f_1(\bar{\rho}, \bar{x}, \eta) + f_\nu(\bar{\rho}, \bar{x}, \eta)). \tag{1–106} \]

We can write the perturbed spatial part of the Einstein Equation:

\[ k^2(\psi + \phi) = -32\pi G a^2[\rho_\gamma^0 \Theta_2 + \rho_\nu^0 \nu_2], \tag{1–107} \]

where \( \Theta_2 = -\int d\mu P_2(\mu) \Theta(k, \mu, \eta), \nu_2 = -\int d\mu P_2(\mu) \nu(k, \mu, \eta), P_2 = 3/2(\mu^2 - 1/3). \) We see that except for the multi-pole contributed by photons and neutrinos the two scalar perturbations are equal and opposite. So for the matter dominated era \( \psi = -\phi \).

### 1.4.3 CMBR anisotropies reveal CDM

To calculate the CMBR anisotropies, one chooses appropriate initial conditions and a set of cosmological parameters such as matter density \( \Omega_m \), baryon density \( \Omega_b \), cosmological constant \( \Omega_\Lambda \)... and solves numerically the dynamical equations. Different sets of cosmological parameters give different anisotropies of CMBR. By fitting these models with data obtained by WMAP, one obtains the abundance of baryons and of
Figure 1-7. CMB Temperature fluctuations [14]
non-relativistic matter in the Universe:

$$\Omega_b h^2 = 0.024 \pm 0.001 \quad \Omega_M h^2 = 0.14 \pm 0.02,$$  \hspace{1cm} (1–108)

which is clear evidence for the existence of non-baryonic cold dark matter.

### 1.5 Evidence for CDM from galactic rotation curves

The rotation curves of galaxies show evidence for dark matter. The observed rotation curves are approximately flat at large distances beyond the edge of the galactic disk. Using classical dynamics the rotation velocity is:

$$v(r) = \sqrt{\frac{GM(r)}{r}}$$  \hspace{1cm} (1–109)

for a spherically symmetric mass distribution, where $M(r)$ is the mass inside radius $r$. Therefore the velocities of stars should behave like $\sim 1/\sqrt{r}$ if the luminous matter were the only matter present. The observed $v(r)$ is approximately constant so it indicates an invisible halo with mass profile $M(r) \sim r$. 

---

Figure 1-8. The WMAP 7-year temperature power spectrum, along with the temperature power spectra from the ACBAR and QUaD experiments. The solid line shows the best-fitting 6-parameter flat CDM model to the WMAP data alone [15].
Figure 1-9. Rotation curve of NGC 6503. [16].
CHAPTER 2
INTRODUCTION TO AXION PHYSICS

2.1 The strong CP problem

2.1.1 Lagrangian of the standard model

The standard model of elementary particles is a very successful theory which
describes the fundamental interactions and the constituents of matter. It successfully
predicts or explains all phenomena observed in the laboratory up to the TeV energy
scale today. The standard model is based on quantized relativistic field theory. The
fundamental lagrangian is made of four parts: the Yang-Mills part \( \mathcal{L}_Y \), the Weyl-Dirac
part \( \mathcal{L}_W \), the Higgs part \( \mathcal{L}_H \), and the Yukawa coupling part \( \mathcal{L}_Y^\nu \). The Yang-Mills part is:

\[
\mathcal{L}_Y = -\frac{1}{4g_3^2} \sum G_{\mu\nu}^A G^{\mu\nu A} - \frac{1}{4g_2^2} \sum F_{\mu\nu}^a F^{\mu\nu a} - \frac{1}{4g_1^2} B_{\mu\nu} B^{\mu\nu},
\]

where \( g_i \) are dimensionless constants, \( A = 1\ldots8, a = 1\ldots3 \) and

\[
G_{\mu\nu}^A = \partial_\mu A_\nu^A - \partial_\nu A_\mu^A - f^{ABC} A_\mu^B A_\nu^C
\]

\[
F_{\mu\nu}^a = \partial_\mu F_\nu^a - \partial_\nu F_\mu^a - \epsilon^{abc} W_{\mu b}^b W_{\nu c}^c
\]

\[
B_{\mu\nu}^A = \partial_\mu B_\nu - \partial_\nu B_\mu
\]

where \( A_\mu^A, F_\mu^a \) and \( B_\mu \) are gauge fields for \( SU(3)_c, SU(2)_L \) and \( U(1) \). \( f^{ABC} \) and \( \epsilon^{abc} \) are
the structure functions of \( SU(3) \) and \( SU(2) \) respectively.

The Weyl-Dirac part describes the fermion fields and their coupling to gauge fields.
Let us first define:

\[
W_\mu = \frac{1}{2} W_\mu^a \tau^a
\]

\[
A_\mu = \frac{1}{2} A_\mu^A \lambda^A
\]

where \( \tau^a \) are the Pauli matrices and \( \lambda^A \) are the Gell-Mann matrices. The left-handed
fermions form \( SU(2) \) doublets, the right-handed fermions are \( SU(2) \) singlets, the quarks
are \( SU(3) \) triplets and the leptons are \( SU(3) \) singlets. One can write the covariant
where the Yukawa couplings \( \Gamma \). The Lagrangian is:

\[
\mathcal{L}_H = (D_\mu H) \dagger (D^\mu H) + \mu^2 H \dagger H - \lambda (H \dagger H)^2,
\]

where \( D_\mu H = (\partial_\mu + i \mathbf{W}_\mu + \frac{i}{2} B_\mu) H \). Finally the general form of the Yukawa part is:

\[
\mathcal{L}_{\text{Y}} = i l_i^T \sigma_2 \tilde{e}_i H^* Y^e_{\tilde{g}} + i Q_i^T \sigma_2 \tilde{d}_j H^* Y^d_{\tilde{g}} + i Q_i^T \sigma_2 \tilde{u}_j \tau_2 H Y^u_{\tilde{g}} + c.c.
\]

where the Yukawa couplings \( Y^e_{\tilde{g}} \) are \( 3 \times 3 \) matrices. Since any matrix can be written as \( M = U_1 D U_2 \), where \( U_{1,2} \) are unitary matrices and \( D \) is a real diagonal matrix, one can always make \( Y^e_{\tilde{g}} \) a real diagonal matrix by redefining the lepton fields. For the quark part there are two terms \( Y^d \ Y^u \) in 2–14, so one can only diagonalize one matrix. One can write:

\[
\mathcal{L}_{\text{Y}} = i Q_i^T \sigma_2 \tilde{d}_{ij} H^* + i Q_i^T \sigma_2 \tilde{u}_J \tau_2 H + c.c.
\]

where the \( y^d_{ij} \) are real and \( U_j \) is the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The CKM matrix contains three mixing angles and one phase. Therefore, the standard model
Lagrangian contains 3 gauge couplings $g_1, g_2, g_3$, four parameters for the CKM matrix, 9 masses and $\lambda$ and $\mu$ terms for the Higgs part. It is a total of 18 parameters, but we will see that there is one more term which is produced by quantum effects in the QCD vacuum. These 19 parameters form the complete parameter space of the standard model of particle physics.

2.1.2 The $U(1)_A$ problem

The masses of the two light quarks $u, d$ are order of MeV which is smaller than the QCD scale $\Lambda_{QCD} = 217\text{MeV}$. One may treat the two quarks as massless. This is called the chiral limit. The light quark QCD Lagrangian is then:

$$L_{QCD} \simeq -\frac{1}{4} G^A_{\mu\nu} G^{\mu\nu A} + \sum_{i=1}^{2} \bar{q}_i \gamma_\mu D^\mu q_i,$$  \hspace{1cm} (2–16)

where the $q_i$ are Dirac fermions. The Lagrangian has symmetry: $SU_L(2) \times SU_R(2) \times U(1) \times U_A(1)$, under which:

$$SU_{L,R}(3) : \quad q \rightarrow U_{L,R} q_{L,R}; \quad U(1) : \quad q_{L,R} \rightarrow e^{i\alpha} q_{L,R};$$

$$U_A(1) : \quad q_L \rightarrow e^{i\theta} q_L; \quad q_R \rightarrow e^{-i\theta} q_R$$  \hspace{1cm} (2–17)

where the $U_{L,R}$ are $SU(2)$ matrices. In low energy QCD the chiral symmetry $SU_L(2) \times SU_R(2)$ is spontaneously broken due to the condensation of quark pairs in the vacuum $$<q_L^{ti} q_{Rj}> = \delta_{ij} \mu^3.$$ This however leaves the vector symmetry $SU_{L+R}(2)$ unbroken. So we have:

$$SU_L(2) \times SU_R(2) \rightarrow SU_{L+R}(2).$$  \hspace{1cm} (2–19)

Since three generators are broken we have three Nambu-Goldstone bosons in the chiral limit. In the real world, the light quarks are not massless. Therefore there are three light mesons $\pi^+, \pi^-, \pi^0$. This prediction fits the experiment very well. However, the quark condensate also breaks $U_A(1)$. Thus a fourth Goldstone boson is produced. Weinberg[19] obtained a mass bound of the fourth goldstone boson: $m < \sqrt{3} m_\pi$.

No such fourth pseudoscalaraer is observed. This puzzle is called the $U(1)_A$ problem.
As explained below, it is solved by introducing the effects of instantons. The $U(1)_A$ current has the ABJ anomaly. The ABJ anomaly term is a total divergence. However, for non-abelian strong coupling gauge fields, it contributes to the physics. Therefore, $U(1)_A$ is explicitly broken due to QCD. So the $U(1)_A$ problem is solved. On the other hand, the anomaly creates the strong CP problem.

2.1.3 The $\theta$ Vacuum and Instantons

Let us consider non-abelian gauge field theory. The gauge transformations are $A_\mu \to U A_\mu U^\dagger - i / g U \partial_\mu U^\dagger$, where $g$ is the gauge coupling and $U$ is a unitary matrix. One of the classical field configurations corresponding to the ground state is $A_\mu = 0$. So $A_\mu = i / g U \partial_\mu U^\dagger$ are classical field configurations describing the vacuum. Let us use the gauge $A_\mu = 0$. So we restrict to $U = U(\vec{x})$ which is independent of time. Also we impose a boundary condition $U(\vec{x}) \to \text{Constant}$ when $|\vec{x}| \to \infty$. For $U(\vec{x})$ satisfying above conditions, it turns out that not every $U$ can smoothly deform to the others without passing through field configurations with non-zero energy. Any $2 \times 2$ special unitary matrix may be written $U = a_4 + i \vec{a} \cdot \vec{\sigma}$ where $a_\mu^2 = 1$. We can set $a_0 = (\vec{x}^2 - \lambda^2) / (\vec{x}^2 + \lambda^2)$ and $\vec{a} = 2i \vec{x} / (\vec{x}^2 + \lambda)$ where $\lambda$ is a parameter. The winding number of the above map is one. A gauge field cannot be smoothly deformed into others with different winding number without passing energy barriers. The winding number can be calculated as

$$n = -\frac{1}{24 \pi^2} \int d^3x i^k \text{Tr}[(U \partial_i U^\dagger)(U \partial_j U^\dagger)(U \partial_k U^\dagger)].$$

For a time dependent field $A_\mu(\vec{x}, t)$, the Pontryagin index

$$q = \frac{g^2}{16 \pi^2} \int d^4x \text{Tr}(F \tilde{F}),$$

is the difference in winding numbers between the configuration at $t = -\infty$ and $t = +\infty$. So non-abelian gauge theory has an infinite number of field configurations with zero energy. They are distinguished by winding numbers $n$ and separated by energy barriers. Although they are separated by energy barriers, configuration of different winding
number can tunnel to each other due to instantons. So for the physical vacuum we have to include field configurations with all possible winding number $n$.

Now let us discuss briefly the instantons. For two quantum states separated by an energy barrier, the tunneling amplitude between them can be calculated by the path integral: $< n2 | H | n1 > \sim e^{-S}$, where $S$ is the Euclidean action with boundary $|n1>$ at $t = -\infty$ and $|n2>$ at $t = +\infty$. In most cases, $S$ is infinite since for field theory one integrates over the entire space time. So the tunneling amplitude vanishes and the two states remain exactly degenerate. For strong coupling non-abelian gauge field theory, there are solutions that mediate between states with different winding number ($n1 \neq n2$) and the action $S$ of these solutions is finite even in the infinite volume limit. Solutions that relate $|n1>$ at $t = -\infty$ and $|n2>$ at $t = +\infty$ with $n2 - n1 = 1$ are the instantons. The simplest instanton solution was discovered by Belavin, Polyakov, Schwartz, and Tyupkin (BPST):

$$A_\mu = if(r)g^{-1}(x)\partial_\mu g(x) , \hspace{1cm} (2–22)$$

$$f(r) = \frac{r^2}{r^2 + \rho^2} , \hspace{1cm} (2–23)$$

$$g(x) = -\frac{i}{r} x_\mu \sigma_\mu , \hspace{1cm} (2–24)$$

where $\sigma_\mu = (\vec{\sigma}, i)$ and $\rho$ is the instanton size. The action of this instanton is $S = |n2 - n1| 8\pi^2 / g^2 = 8\pi^2 / g^2$. The instanton describes tunneling between two states with winding number differing by one.

Now let us discuss the physical vacuum $|\omega >$. Let $T$ be a gauge transformation that changes winding number by one: $T |n > = |n + 1 >$. For the physical vacuum, we have $T|\omega > = e^{i\theta}|\omega >$. Since we have seen that $|\omega >$ has to include states with all possible winding numbers $n$, one concludes that the physical vacuum has the form: $|\omega > = \sum_n e^{in\theta}|n >$. Thus the physical vacuum is defined by a parameter $\theta \in [0, 2\pi]$. The $\theta$ vacuum produces an new term in the standard model Lagrangian. Let us consider the
vacuum to vacuum transition amplitude:

\[
<\theta\left| e^{-iHt}\right|\theta > = \sum_{n_1} \sum_n e^{-i(n_1\theta_1-n\theta)} <n_1\left| e^{-iHt}\right|n >
\]

\[
= \sum_{n_1} e^{-i_1(\theta_1-\theta)} \sum_{n_1-n} \int [DA_\mu]_{n_1-n} \exp[-\int d^4xL - i(n_1-n)\theta]
\]

\[
= \delta(\theta_1-\theta) \int [DA_\mu]_q \exp[-\int d^4x(L + \frac{\theta}{32\pi^2} F^a_\mu \pi^a \tilde{F}^{\mu\nu})].
\] (2–25)

where \( \int [DA_\mu]_{n_1-n} \) denotes functional integration with respect to gauge configurations of Pontryagin index \( n_1 - n = q \). We see an additional effective interaction is obtained:

\[
L_\theta = \frac{\theta}{32\pi^2} F^a_\mu \pi^{a\mu}. 
\]

Because \( \tilde{F} \) is CP odd, QCD is not CP invariant if \( \theta \neq 0 \).

Noether’s theorem and the ABJ anomaly imply that the physics of QCD is unchanged under the transformation:

\[
\theta F \tilde{F} \rightarrow (\theta - \sum \alpha_i) F \tilde{F}
\] (2–26)

\[
\bar{q}_i m_i q_i \rightarrow \bar{q}_i e^{(\alpha_i \gamma_5/2)} m_i e^{(\alpha_i \gamma_5/2)} q_i
\] (2–27)

We can choose \( \sum \alpha_i = \theta \) to get rid of the \( \theta \) term at the expense of additional CP violating mass terms:

\[
\sum (m_i \cos \alpha_i \bar{q}_i q_i + m_i \sin \alpha_i \bar{q}_i \gamma_5 q_i).
\] (2–28)

The \( \alpha_i \) are constrained by the requirement that the CP violating Lagrangian should not cause a realignment of vacuum and the QCD vacuum is a flavor singlet, i.e. \( <\bar{u}u> = <\bar{d}d> = <\bar{s}s> \). Baluni\[24\] finds:

\[
\sum \alpha_i m_i = \frac{m_u m_d m_s \theta}{m_u m_s + m_d m_s + m_u m_d}.
\] (2–29)

So the effective interaction is:

\[
L = \frac{m_u m_d m_s \theta}{(m_u + m_d)m_s + m_u m_d} (\bar{u}i \gamma_5 u + \bar{d}i \gamma_5 d + \bar{s}i \gamma_5 s),
\] (2–30)

for small \( \theta \). Here the masses are assumed to be real. If the masses are complex, the parameter controlling CP violation in the strong interaction is \( \bar{\theta} = \theta + \arg det M \). The
CP violation gives a contribution to the neutron electric dipole moment. Baluni used the MIT bag model and found $d_n = 2.7 \times 10^{-16} \bar{q}_{ecm}$; Crewther et al. using current algebra find $d = 5.2 \times 10^{-16} \bar{q}_{ecm}$. The upper bound of the neutron electric dipole moment is $0.6 \times 10^{-24} \text{ecm}$. So $\bar{\theta}$ has an a upper bound $|\bar{\theta}| < 1.2 \times 10^{-9}$. Such a small $\bar{\theta}$ needs an explanation, which is the strong CP problem.

2.1.4 Solutions to the strong CP problem

To solve the strong CP problem, there are three suggestions:

1. The ultraviolet up quark mass is zero, so the RHS of 2–30 vanishes. However, lattice QCD simulations suggest a non-zero up quark mass.

2. CP is spontaneously broken, In that case $\bar{\theta}$ is finite and can be arranged to be sufficiently small.

3. The Peccei-Quinn solution. If $\bar{\theta}$ is a dynamical variable, it will naturally go to the value that minimizes the energy, which is zero as we will see. This extra variable will produce a quasi-Nambu-Goldstone boson which is called the axion.

A simple proof to show that the energy is minimized when $\bar{\theta} = 0$ was given by C. Vafa and E. Witten [20]. Let us consider the path integral in Euclidean space. The QCD lagrangian with the $\bar{\theta}$ term is $L = -1/4 g^2 \text{Tr}(G_{\mu\nu} G_{\mu\nu}) + \sum q_i (D_{\mu} \gamma_{\mu} + m_i) q_i + i\bar{\theta}/32\pi^2 \text{Tr}(G_{\mu\nu} G_{\mu\nu})$. Integrating out the fermions one has:

$$e^{-\mathcal{V}_E} = \int DA_\mu \det(D_{\mu} \gamma_{\mu} + M) e^{\int d^4x \left[ \frac{1}{4} g^2 \text{Tr}(G_{\mu\nu} G_{\mu\nu}) - \frac{i}{32\pi^2} \text{Tr}(G_{\mu\nu} G_{\mu\nu}) \right]}.$$  \hspace{1cm} (2–31)

In QCD $\det(D_{\mu} \gamma_{\mu} + M)$ is positive. This is because $(\gamma_{\mu} iD_{\mu})$ is hermitian in Euclidean space and $\gamma_5$ anti-commutes with $\gamma_{\mu}$. This implies that for every (imaginary) eigenvalue of $(\gamma_{\mu} D_{\mu})$ there is an eigenvalue of opposite sign. Note also that $iG \tilde{G}$ is pure imaginary, so it only reduces the total value of the path integral. Thus when $\bar{\theta}$ is not zero, the energy is increased.
If $\theta$ is dynamical variable in a Lagrangion of the form:

$$
L = -\frac{1}{4} g^2 Tr(G_{\mu\nu} G_{\mu\nu}) + \sum q_i (D_{\mu} \gamma_{\mu} + m_i) q_i 
+ \frac{\theta}{32\pi^2} Tr G_{\mu\nu} \bar{G}_{\mu\nu} + \frac{1}{2} \partial_{\mu} a \partial^{\mu} a + a/(f_a 32\pi^2) Tr G_{\mu\nu} \bar{G}_{\mu\nu},
$$

(2–32)

the shifting of the dynamical field $\theta$ will automatically drive the vacuum energy to its minimum so that $\theta + a/f_a \rightarrow 0$ and the CP violation of QCD no longer exists. One can show that the low energy effective theory has the form 2–32 if there is a $U_{ PQ}(1)$ symmetry in the standard model. $U_{ PQ}(1)$ symmetry must have the following properties:

1) it is a symmetry of the classical action density. 2) it is spontaneously broken. 3) it is explicitly broken by the QCD chiral anomaly.

2.1.5 Axion models

The axion fits into various extensions of the standard model. The Peccei-Quinn-Weinberg-Wilczek (PQWW) axion is the earliest model. This type of model assumes an additional Higgs doublet and a spontaneously broken global symmetry, the Peccei-Quinn symmetry. Appropriate $U(1)_{ PQ}$ charges are assigned to the quarks so that $U_{ PQ}(1)$ is explicitly broken by the QCD chiral anomaly. This type of axion has $f_a \sim 250 GeV$, which is the electrowak scale. Astrophysical and experimental searches have ruled out this type of axion. Therefore we will not discuss such models in detail.
The PQWW axion tied Peccei-Quinn symmetry breaking scale with the electroweak scale and was ruled out. Later, J. Kim et. al. separated these two scales to give rise to a new type of axion, which is often called the "invisible axion". There are two major benchmark models of this type, one is the Kim-Shifman-Vainshtein-Zakharov (KSVZ) axion, and the other is the Dine-Fischler-Srednicki-Zhitnitskii (DFSZ) axion.

Let us first consider the KSVZ axion. The KSVZ model introduces an new complex scalar field $\sigma$ and a new heavy quark $Q$. The $U(1)_{PQ}$ transformation is

$$U(1)_{PQ} : \quad a \rightarrow a + f_a \alpha \quad (2-33)$$

$$\sigma \rightarrow \exp(iq\alpha)\sigma \quad (2-34)$$

$$Q_L \rightarrow \exp(iQ\alpha/2)Q_L \quad (2-35)$$

$$Q_R \rightarrow \exp(-iQ\alpha/2)Q_R \quad (2-36)$$

where $q$ is the PQ charge for the scalar field $\sigma$ and we will assume $q = 1$. The potential for $\sigma$ is $U(1)_{PQ}$ invariant. $U_{PQ}(1)$ is spontaneously broken by the vacuum expectation value $\langle \sigma \rangle = \nu$. The scalar field $\sigma$ may be written as:

$$\sigma = (\nu + \rho)\exp(i\frac{\alpha}{\nu}). \quad (2-37)$$

We see that $f_a = \nu$. To preserve the PQ symmetry the heavy quark cannot have a bare mass. The relevant Yukawa coupling and Higgs potential are

$$L_{Y_L} = -fQ_L^\dagger \sigma Q_R - f^*Q_R^\dagger \sigma^* Q_R \quad (2-38)$$

$$V = -\mu^2_\sigma \sigma^* \sigma + \lambda_\sigma (\sigma^* \sigma)^2 \quad (2-39)$$

where $f$, $\lambda_\sigma$ are parameters. We see that the mass of the heavy quark $Q$ depends on $\nu$.

A drawback of the KSVZ model is it introduces a new heavy quark. The DFSZ model avoids this at the expense of an additional Higgs doublet. In the DFSZ model, $\sigma$ still is a complex scalar. However $\sigma$ couples to the Higgs doublets, which then couple to
the light quarks. The relevant Yukawa interactions and scalar potential are:

\[
L_{\gamma q} = -f_{y}^{(u)} q_{Lj} \phi_{2} u_{Ri} - f_{y}^{(d)} q_{Lj} \phi_{1} d_{Ri} + c.c. \tag{2–40}
\]

\[
V = \left( a \phi_{1}^{\dagger} \phi_{1} + b \phi_{2}^{\dagger} \phi_{2} \right) \sigma \sigma^{\ast} + c(\phi_{1}^{\dagger} \phi_{2} \sigma^{2} + h.c.)
\]

\[
+ d | \phi_{1}^{T} \phi_{2} |^{2} + e | \phi_{1} \phi_{2}^{T} |^{2} + \lambda_{1}(\phi_{1}^{\dagger} \phi_{1} - \nu_{1}^{2}/2)^{2}
\]

\[
+ \lambda_{2}(\phi_{2}^{\dagger} \phi_{2} - \nu_{2}^{2}/2)^{2} + \lambda(\sigma^{\ast} \sigma - \nu^{2}/2)^{2}, \tag{2–41}
\]

where \( T \) denotes transpose, \( a, b, c, d \) are parameters, \( \nu_{1,2} \) are the vacuum expectation values of the two Higgs doublets respectively and \( i, j \) are the family indices. Under the Peccei-Quinn symmetry:

\[
U(1)_{PQ} : \quad \phi_{1} \rightarrow \exp(-i \beta Q) \phi_{1} \tag{2–42}
\]

\[
\phi_{2} \rightarrow \exp(-i \gamma Q) \phi_{2} \tag{2–43}
\]

\[
u_{R} \rightarrow \exp(i \gamma Q) \nu_{R} \tag{2–44}
\]

\[
d_{R} \rightarrow \exp(i \beta Q) d_{R} \tag{2–45}
\]

\[
\sigma \rightarrow \exp(i \alpha) \sigma \tag{2–46}
\]

We have

\[
2 \alpha = \beta + \gamma. \tag{2–47}
\]

where

\[
\beta = \frac{2x}{x + x^{-1}} \alpha \tag{2–48}
\]

\[
\gamma = \frac{2x^{-1}}{x + x^{-1}} \alpha, \tag{2–49}
\]

with \( x = \nu_{2}/\nu_{1} \).
2.2 Axion properties

2.2.1 The axion mass

Axions are quasi-Goldstone bosons and they obtain a mass due to the potential generated by the instantons. One can remove the $aG\tilde{G}$ term by introducing terms that mix pseudo scalar mesons and axions. Therefore the mass of axions is related to the mass of the pions. Going to the mass eigenstates, one finds:

$$m_a \sim \frac{m_{\pi^0} f_\pi}{f_a}$$  \hspace{1cm} (2–50)

where $f_\pi$ is the pion decay constant and we have neglected the factors depending on the quark mass ratios.
2.2.2 The axion couplings

The axion-photon coupling is the sum of two pieces: one is due to the ABJ anomaly and the other is due to axion mixing with mesons and the coupling of mesons to photons. Both pieces are suppressed by the symmetry breaking scale. We have:

$$ L_{a\gamma\gamma} \sim \frac{\alpha}{\pi f_a} a F \tilde{F}. $$

(2–51)

The axion couples to the fermions too

$$ L_{aq_i} \sim i \frac{m_i}{f_a} a \bar{q}_i \gamma_5 q_i, $$

(2–52)

where $i$ denote different fermions. These couplings are also suppressed by the symmetry breaking scale $f_a$.

2.3 Axion astrophysics and cosmology

2.3.1 Constraint due to cosmology

One of the most notable features of axions is that they are an excellent candidate for cold dark matter. Dark means negligibly interacting except for gravitational interactions. The dark matter particles should also be stable compared to the age of Universe. The axions couple to other particles very weakly so they naturally satisfy these two
conditions. For example, the lifetime $T$ of axons is approximately

$$T \approx 10^{50} \text{s} \left( \frac{f_a}{10^{12} \text{ GeV}} \right)^5.$$

(2–53)

Axions can be hot dark matter or cold dark matter depending on their origin. Hot dark matter axions are created from thermal processes in the hot plasma in the early universe, and their present number density is

$$n_a(t_0) = \frac{\zeta(3) T_D^3}{\pi^2} \left( \frac{a_D}{a_0} \right)^3,$$

(2–54)

where $T_D$ is the axion decoupling temperature. $a_D$ and $a_0$ are the scale factors at decoupling and today, respectively.

The cold axions are produced by the misalignment mechanism [21] and their number density today is

$$n_a \approx \frac{f_a^2}{2t_1} \left( \frac{a_1}{a_0} \right)^3,$$

(2–55)

where $t_1$ is defined by $m_a(t_1) t_1 = 1$ and is order of $10^{-7} \text{sec}(f_a/10^{12} \text{ GeV})^{1/3}$. If all the dark matter is composed of cold axions, the mass of axions is of order $m_a \approx 10^{-6} \text{eV}$.

### 2.3.2 Constraint due to astrophysics

The axions change the evolution of stars because stars emit axions from their bulk while they emit photons only from their surfaces [22]. A detailed study of our star, the Sun, shows that axion emission due to the Primakoff process modifies the temperature distribution profile of the Sun and therefore changes the neutrino flux. The measured neutrino flux from the Sun gives the constraint $g_{a\gamma} \lesssim 5 \cdot 10^{-10} \text{ GeV}^{-1}$ [23].

Studies of red giants and horizontal branch stars in globular clusters also give very important constraints on axion couplings. A giant star is a star with larger radius and luminosity than a main sequence star of the same surface temperature, and a red giant is a giant star with 0.5-10 solar masses in its late phase of evolution. The horizontal branch stars are stars in the evolutionary phase immediately after the red giant phase for stars with a mass similar to the mass of Sun. Axions are produced more efficiently in
Figure 2-5. The cosmology and astrophysical constraint on ALPs/axions

Figure 2-6. The Primakoff Process convert photons to axions
the horizontal branch stars than in the red giants. The additional axion flux in horizontal branch stars exhausts their nuclear fuel and therefore decreases their lifetime. This would imply that the red giants are more abundant in globular clusters compared to the prediction of standard star evolution. The observations agree within 10% of the standard star evolution prediction, and therefore constrain the additional energy-loss channel. Consequently, the coupling is limited to [26]

\[ g_{\gamma\gamma} < 10^{-10} \text{ GeV}^{-1}. \]  \hspace{1cm} (2–56)

A white dwarf is the final evolution phase of a star with a mass not high enough to become a neutron star. The energy losses of a white dwarf in the standard star evolution model are due to neutrino emission from the core and photon emission from the surface. If axions exist, there will be an additional channel of energy loss due to axion emission due to Compton-like scattering involving the couplings of axions to electrons. The studies of the cooling rate of white dwarfs conclude [27]:

\[ g_{\text{ae}} < 10^{-13} \]  \hspace{1cm} (2–57)

A supernova is the explosion of a star that forms a neutron star or a black hole. The explosion expels the matter of the star into an expanding shell with a speed of 10% that of the speed of light. The core density is high enough to diffuse the propagation of neutrinos. If axions exist, additional energy loss channels, dominated by axion nucleon bremsstrahlung, will be present besides the channel due to neutrino diffusive transportation, so that the duration of the neutrino burst from supernovas can be shortened. If the coupling is too weak the duration is not changed since too few axions are produced. Too strong a coupling also does not change the duration because axions cannot free steam from the core. The study of supernova 1987A constrains the axion
nucleon coupling [28]:

\[
10^{-10} < g_{\text{ann}} \quad (2-58)
\]

or

\[
10^{-7} > g_{\text{ann}}. \quad (2-59)
\]
CHAPTER 3
AXION SEARCHES

3.1 Axion dark matter search

In chapter 1 we saw that 23% of the universe’s total energy density today is contributed by dark matter. In chapter 2 we saw that QCD axions are one of the leading candidates for cold dark matter[21] because the QCD axions are effectively collisionless and the misalignment mechanism produces a very cold population of axions which have the required energy density.

The Axion Dark Matter eXperiment (ADMX)[29][30] is a realization of axion haloscope in which axions in the halo of our galaxy are induced to convert in a cavity to microwave photons that are then picked up by an antenna. The power of the axion signal is proportional to the local axion density, and the signal width is proportional to the energy dispersion of the dark-matter axions. Therefore, the signal properties observed by ADMX depend on the structure of the galactic halo.

3.1.1 Theory and experiment

The axions couple to photons via:

\[ \mathcal{L}_{a\gamma\gamma} = g_\gamma \frac{\alpha}{\pi} \frac{a(\chi)}{f_a} F^{\mu\nu} \tilde{F}_{\mu\nu}, \]  

(3–1)

where \( \alpha \) is the fine structure constant, \( f_a \) is the axion decay constant, and \( g_\gamma \) is a model-dependent coupling of order one (In the KSVZ model \( g_\gamma \approx -0.97 \), and in the DFSZ mode \( g_\gamma \approx 0.36 \)).

Cold axions in the galactic halo are non-relativistic so that the axion-photon conversion in a magnetic field creates photons whose energy approximately equals the mass of the axions. If the mass is within the bandwidth of the cavity, the conversion is resonant and the power \( P \) of the photons is

\[ P = \left( \frac{\alpha g_\gamma}{\pi f_a} \right)^2 \frac{V B_0^2 \rho_a C}{m_a} \min(Q, Q_a), \]  

(3–2)
where $V$ is the cavity volume, $B_0$ is the magnetic field inside the cavity, $\rho_a$ is the local density of halo axions, $Q$ is the loaded quality factor of the cavity (defined as center frequency divided by the frequency bandwidth), $Q_a$ is the ratio of the energy of axions to their energy spread, and $C$ is a form factor given by:

$$C = \frac{\left| \int_V d^3x \mathbf{E}_\omega \cdot \mathbf{B}_0 \right|^2}{B_0^2 V \int_V d^3x \epsilon|\mathbf{E}_\omega|^2},$$

where $\mathbf{E}_\omega(x)$ is the electric field of the cavity mode the axions convert into. $\epsilon$ in (3–3) is the dielectric constant of the medium inside the cavity. A typical $C$ is order of unity. (3–2) can be rewritten as:

$$P = 0.5 \times 10^{-21} \, \text{W} \left( \frac{V}{500 \, \text{L}} \right) \left( \frac{B_0}{7 \, \text{T}} \right)^2 C \left( \frac{\rho_a}{0.35} \right)^2 \times \left( \frac{0.5 \times 10^{-24} \, \text{g cm}^{-3}}{1 \, \text{GHz}} \right) \left( \frac{\nu}{1 \, \text{GHz}} \right) \left( \frac{\min(Q, Q_a)}{10^6} \right).$$

The QCD axion mass is the range $10^{-6} \, \text{eV} < m_a < 10^{-3} \, \text{eV}$, which corresponds to a frequency range $240 \, \text{MHz}$ to $240 \, \text{GHz}$. Because of the high $Q$, only a few kHz of bandwidth can be observed at any given time. Since the power of thermal noise $\bar{P} = k \Delta \omega T$, where $k$ is the Boltzmann constant and $\Delta \omega$ is the bandwidth, the signal to noise ratio is

$$SNR = \frac{P}{kT \sqrt{t/\Delta \omega}},$$

where $t$ is the measurement integration time. Therefore it will take a long time to scan the entire mass range with a feasible laboratory setup. The signal frequency will be modulated due to the axions velocity relative to the Earth. The shift in frequency $\Delta f$ due to a shift $\Delta \nu$ in the velocity $\nu$ of the axions relative to the Earth is

$$\Delta f = f \nu \Delta \nu.$$

Effects due to the rotation and orbital motion of the Earth are observable.
An illustration of the setup of ADMX is shown in Figure 3-1. The frequency of the resonant mode is tuned by moving the metal rods inside the cavity. The cavity is cooled to liquid He temperature to reduce thermal noise.

### 3.2 Solar axion searches

Stars produce abundant numbers of axions in their core by the Primakoff effect, and the Sun is the dominant source of this kind of axions in the sky due to its relatively short distance to the Earth. After production, the axions propagate freely with a speed very close to the speed of light and reach the Earth after $\sim 500$s. One can detect the solar axions by using an axion helioscope[30]. Naturally, the sensitivity of the axion helioscope depends on the axion flux from the Sun and the probability of axion photon conversion in the detector.
3.2.1 Solar axion production

In the core of stars thermal photons with energies of about a keV are converted into axions in the electric fields of charged particles: $\gamma + Ze \leftrightarrow Ze + a$. This process is known as the Primakoff effect. The temperature of electrons and nuclei is about a keV, which is much smaller than their mass, so that the differential cross-section for this process is given by:

$$\frac{d\sigma}{d\Omega} = \frac{g_{\gamma\gamma}^2 Z^2 \alpha |\vec{k}_\gamma \times \vec{k}_a|^2}{8\pi |\vec{q}|^4},$$

where $\vec{q} = \vec{k}_\gamma - \vec{k}_a$ is the momentum transfer, $Z$ is the charge of a nucleus, and $\alpha$ is the fine structure constant. In a plasma [34], the differential cross section is modified to:

$$\frac{d\sigma}{d\Omega} = \frac{g_{\gamma\gamma}^2 Z^2 \alpha |\vec{k}_\gamma \times \vec{k}_a|^2}{8\pi |\vec{q}|^4} \frac{\kappa^2}{\kappa + \vec{q}^2},$$

where $\kappa$ is the Debye-Huckel scale:

$$\kappa^2 = \frac{4\pi \alpha \eta_B}{T} \left( Y_e + \sum_j Z_j^2 Y_j \right),$$

Figure 3-2. Axion-photon coupling excluded at the 90% confidence level[32].

[32] Reference for the 90% confidence level.

[34] Reference for the plasma modification of the differential cross section.
where \( n_B \) is the baryon number density, \( Y_e \) and \( Y_j \) are the percentages of electrons and nucleons in the plasma and \( T \) is the plasma temperature. The total cross section can be obtained by integrating (3–8) and summing over all target species:

\[
\sigma = n_B (Y_e + \sum Z_j^2 Y_j) \int d\Omega \frac{g_{\gamma \gamma}^2 Z^2 \alpha}{8\pi} \left| \vec{k} \times \vec{k}_3 \right|^2 \frac{q^2}{\kappa + q^2}.
\] (3–10)

For small axion mass, \( m_a \ll k_\gamma \), one finds:

\[
\sigma = \frac{g_{\gamma \gamma}^2 T k_s^2}{32\pi} \left[ \left( 1 + \frac{k_s^2}{4E^2} \right) \log \left( 1 + \frac{4E^2}{k_s^2} \right) - 1 \right].
\] (3–11)

The energy spectrum of axions at Earth is then:

\[
\frac{d\phi}{dE} = \frac{\sigma n_\gamma}{4\pi d^2} = \frac{1}{4\pi d^2} \int_{0}^{R} d^3 r \frac{\sigma E}{e^{E/T(r)} - 1},
\] (3–12)

where \( d \) is the distance between the Sun and the Earth, \( R \) is the diameter of the Sun and \( T(r) \) is the temperature inside the Sun. We see that the solar axion flux density depends on \( T(r) \) which can be obtained from solar models. For the well established Solar model one finds the axion flux spectrum [22] shown in figure 3-3.
3.2.2 Detector using Bragg scattering

To detect solar axions, one may use the Coulomb field of nuclei in a crystal to convert axions to photons by the inverse Primakoff process. The energy of solar axions is $\sim 4\text{keV}$, so the wavelength of converted photons is comparable to the lattice spacing of the crystal. Thus the photons converted in a crystal will produce a Bragg pattern. The constructive interference in the Bragg pattern can enhance the signal by order $10^4$. There are several experiments using this technique on crystals such as Germanium (SOLAX) and Sodium-iodide (DAMA). The resulting bounds on the axion-photon coupling are of order $10^{-9}\text{GeV}^{-1}[38]$.

3.2.3 Axion helioscope

The axion helioscope[30] employs a laboratory magnetic field to convert solar axions to low-energy X-rays. In a region of length $L$, magnetic field $B$, and buffer gas whose absorption rate is $\Gamma$, the conversion probability of axions is [33]

$$P = \frac{(g_{a\gamma}BL/2)^2}{L^2 \left(q^2 + \Gamma^2/4\right)} \left[1 + e^{-\Gamma L} - 2e^{-\Gamma L/2} \cos(qL)\right], \quad (3-13)$$

where $q$ is the momentum transfer:

$$q = \left| \frac{m_a^2 - m_\gamma^2}{2E_a} \right|, \quad (3-14)$$

and $m_\gamma$ is the effective photon mass in the buffer gas:

$$m_\gamma = \omega_p = \sqrt{\frac{4\pi\alpha n_e}{m_e}}. \quad (3-15)$$

$n_e$ is the electron density of the buffer gas. The purpose of the buffer gas is to restore the coherence of axion-photon conversion ($q = 0$) in case the axions are heavy.

The Tokyo Axion Helioscope provides the limit: $g_{a\gamma\gamma} \sim 6.0 \cdot 10^{-10}\text{GeV}^{-1}$ for $m_a \lesssim 0.03\text{eV}[39]$. The CERN Axion Solar Telescope (CAST) is the most sensitive axion-helioscope experiment today. The limit can be seen in Figure 3-5.
experimental setup.

Figure 3-4. The CAST Helioscope at CERN [40].

Figure 3-5. Axion-photon coupling limit as a function of the axion mass provided by CAST experiment[41].
3.3 Laser experiments

Both photon regeneration experiments and polarization experiments employ laser technology. They do not rely on physical processes inside the Sun or the hypothesis that axions are the dark matter. So everything is under laboratory control, but they are generally less sensitive than the helioscope and haloscope experiments.

3.3.1 Photon regeneration

In photon regeneration experiments a small fraction of photons in a laser beam traverses a region permeated by a magnetic field, where it is converted to axions. Because of their weak coupling to ordinary matter, the axions then travel essentially unimpeded through a wall, on the other side of which is an identical arrangement of magnets, where some of the axions are induced to convert back to photons, which can be detected. The major drawback of this kind of experiment is that its signal is very weak since two stages of conversion are required. Resonantly enhanced photon regeneration experiments address this problem by adding resonant cavities on both sides of the wall so that the signal is enhanced by a factor of \( \mathcal{F}^2 \) where \( \mathcal{F} \) is the finesse of the Fabry-Perot cavities which could be order of \( 10^5 \). The improvement in sensitivity to \( g_{a\gamma\gamma} \) is a factor \( \mathcal{F}^{1/2} \) which is 300. The constraints from non-resonant photon regeneration experiments, such as that by the Brookhaven-Fermilab-Rutherford-Trieste (BFRT) collaboration, are of order[42]:

\[
g_{a\gamma\gamma} < 10^{-7} \text{ GeV} \quad \text{with} \quad m_a < 10^{-3} \text{ eV}. \tag{3-16}
\]

3.3.2 Polarization experiments

The polarization experiments look for birefringence and dichroism created by axion-photon mixing. Starting with a beam linearly polarized at 45 degrees with respect to the direction of a transverse external magnetic field, the beam can be viewed as a superposition of two beams, one parallel and one normal to the magnetic field. The
photon polarization vector of the beam emerging from the magnetic region rotates due to the amplitude reduction of the parallel component. Initially linearly polarized light also becomes elliptically polarized due to the phase lag of the parallel component:

\[ \delta \theta \approx \frac{g_{\gamma\gamma}^2 B^2 \omega_{\gamma}}{m_a^4} \left[ \frac{m_a^2 L}{2 \omega_{\gamma}} - \sin \left( \frac{m_a^2 L}{2 \omega_{\gamma}} \right) \right]. \]  

(3–17)

However, QED effects can also produce a phase lag due to the effective interaction

\[ \frac{\alpha^2}{90 \pi^2} \left[ (F_{\mu\nu} F^{\mu\nu})^2 + \frac{7}{4} (F_{\mu\nu} F^{\mu\nu})^2 \right]. \]

The main difficulty of the polarimetry experiments is the unavoidable intrinsic birefringence of the optical elements. In practice, the sensitivity of the polarization measurements is severely limited by this fact. The bound obtained by the BFRT collaboration is

\[ g_{\gamma\gamma} < 3.6 \times 10^{-7} \text{GeV}^{-1} \]  

with \( m_a < 5 \times 10^{-4} \text{eV} \).
In this chapter we discuss the axions Bose-Einstein condensation and the possible thermalization of axions with other particle species. This discussion is based on [46].

4.1 Review of cold axion properties

Cold axions (with mass $\sim 10^{-5}\text{eV}$) are one of the leading candidates for cold dark matter (CDM) along with WIMPs (Weakly Interacting Massive Particles with mass $\sim 100\text{GeV}$) and sterile neutrinos (with mass $\sim \text{KeV}$). Cold axions are created when the axion mass turns on at QCD time: $t_1 \sim 2 \cdot 10^{-7}\text{sec}$. The cold axions have small velocity dispersion, of order:

$$\delta v(t) \sim \frac{1}{m a(t)} a(t_1),$$

where $a(t)$ is the scale factor. The velocity dispersion determines the effective temperature of cold axions. In case inflation occurs after the Peccei-Quinn phase transition, $\delta v$ is even smaller because the axion field gets homogenized during inflation.

The number density of cold axions is:

$$n(t) \sim \frac{4 \cdot 10^{47}}{\text{cm}^3} \left( \frac{f}{10^{12}\text{ GeV}} \right)^{\frac{6}{3}} \left( \frac{a(t_1)}{a(t)} \right)^{3}.$$  \hspace{1cm} (4–2)

Clearly, the average phase space density $\mathcal{N}$ is very high due to the combination of high number density and small velocity dispersion. We have:

$$\mathcal{N} \sim n \frac{(2\pi)^3}{3 (m \delta v)^3} \sim 10^{61} \left( \frac{f}{10^{12} \text{ GeV}} \right)^{\frac{8}{3}}.$$  \hspace{1cm} (4–3)

In addition, axions interact very weakly. For example, the interactions $\frac{\lambda}{4f} \varphi^4$ and $g_{a\gamma\gamma} \vec{E} \cdot \vec{B}$ have $\lambda \sim 10^{-48}$ and $g_{a\gamma\gamma} \sim 10^{-22} \text{ eV}^{-1}$. 


4.2 Axion thermalization in the particle kinetic and condensed regimes

4.2.1 Evolution equations for nonrelativistic axions

The action density of axions is

\[ \mathcal{L}_a = -\frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 + \frac{\lambda}{4!} \phi^4 + \ldots \]  \hspace{1cm} (4–4)\]

The dots represent interactions of the axion with other particles and axion self-interactions which are higher order in an expansion in powers of \( \phi \). For the axion that solves the strong CP problem [1–3], the mass \( m \) and self-coupling \( \lambda \) are given by

\[
m = \frac{m_\pi f_\pi \sqrt{m_u m_d}}{f_a} \approx 6 \cdot 10^{-6} \text{eV} \frac{10^{12} \text{GeV}}{f_a} \]
\[
\lambda = \frac{m^2}{f_a^2} \left[ \frac{m_u^3 + m_d^3}{(m_u + m_d)^3} \right] \approx 0.35 \frac{m^2}{f_a^2} \]  \hspace{1cm} (4–5)\]

where \( m_\pi \) is the pion mass, \( f_\pi \approx 93 \text{ MeV} \) the pion decay constant, and \( m_u \) and \( m_d \) are the up and down quark masses. The formula for the axion mass [2] is obtained by expanding the effective potential for pions and axions to second order in the physical axion field. To obtain \( \lambda \), simply expand to fourth order.

We introduce a cubic box of volume \( V = L^3 \) with periodic boundary conditions at its surface. Inside the box, the axion field \( \phi(\vec{x}, t) \) and its canonical conjugate field \( \pi(\vec{x}, t) \) are expanded into Fourier components

\[
\phi(\vec{x}, t) = \sum_{\vec{\kappa}} \left( a_{\vec{\kappa}}(t) \Phi_{\vec{\kappa}}(\vec{x}) + a_{\vec{\kappa}}^\dagger(t) \Phi^*_{\vec{\kappa}}(\vec{x}) \right)
\]
\[
\pi(\vec{x}, t) = \sum_{\vec{\kappa}} (-i \omega_{\vec{\kappa}}) \left( a_{\vec{\kappa}}(t) \Phi_{\vec{\kappa}}(\vec{x}) - a_{\vec{\kappa}}^\dagger(t) \Phi^*_{\vec{\kappa}}(\vec{x}) \right) \]  \hspace{1cm} (4–6)\]

where

\[
\Phi_{\vec{\kappa}}(\vec{x}) = \frac{1}{\sqrt{2 \omega_{\vec{\kappa}}} V} e^{i \vec{\kappa} \vec{x}} \]  \hspace{1cm} (4–7)\]
\( \vec{n} = (n_1, n_2, n_3) \) with \( n_k \) \( (k = 1, 2, 3) \) integers, \( \vec{p}_n = \frac{2\pi}{L} \vec{n} \), and \( \omega = \sqrt{\vec{p} \cdot \vec{p} + m^2} \). The \( a_{\vec{n}} \) and \( a_{\vec{n}}^\dagger \) satisfy canonical equal-time commutation relations:

\[
[a_{\vec{n}}(t), a_{\vec{n}}^\dagger(t)] = \delta_{\vec{n},\vec{n}^\prime}, \quad [a_{\vec{n}}(t), a_{\vec{n}}(t)] = 0.
\] (4–8)

Note that we are quantizing in the Heisenberg picture, not the interacting picture.

Provided the axions are non-relativistic, the Hamiltonian is

\[
H = \sum_{\vec{n}} \omega_{\vec{n}} a_{\vec{n}}^\dagger a_{\vec{n}} + \frac{1}{4} \sum_{\vec{n}_1, \vec{n}_2, \vec{n}_3, \vec{n}_4} \Lambda_{\vec{n}_1, \vec{n}_2} a_{\vec{n}_1}^\dagger a_{\vec{n}_2}^\dagger a_{\vec{n}_3} a_{\vec{n}_4}
\] (4–9)

where

\[
\Lambda_{\vec{n}_1, \vec{n}_2} = -\frac{\lambda}{4m^2 \sqrt{\vec{n}_1 + \vec{n}_2 + \vec{n}_3 + \vec{n}_4}}
\] (4–10)

The presence of the Kronecker symbol \( \delta_{\vec{n}_1 + \vec{n}_2, \vec{n}_3 + \vec{n}_4} \) expresses momentum conservation for each individual interaction. In Eq. (4–9) we dropped all terms of the form \( a_{\vec{n}}^\dagger a_{\vec{n}}^\dagger a_{\vec{n}}^\dagger \), \( a_{\vec{n}} a_{\vec{n}}^\dagger a_{\vec{n}}^\dagger \), \( a_{\vec{n}} a_{\vec{n}}^\dagger a_{\vec{n}} a_{\vec{n}} \), and \( a_{\vec{n}} a_{\vec{n}} a_{\vec{n}}^\dagger \). We are justified in doing so because energy conservation allows only axion number conserving processes at tree level. Axion number violating processes occur in loop diagrams but can be safely ignored because they are higher order in an expansion in powers of \( \frac{1}{\lambda} \). In fact, all axion number violating processes, including the axion decay to two photons, occur on time scales much longer than the age of the universe in the axion mass range \((10^{-5} \text{ eV})\) of interest.

In the Newtonian limit, the gravitational self-interactions of the axion fluid are described by

\[
H_g = -\frac{G}{2} \int d^3x \ d^3x' \frac{\rho(\vec{x}, t)\rho(\vec{x}', t)}{|\vec{x} - \vec{x}'|}
\] (4–11)

where \( \rho = \frac{1}{2}(\pi^2 + m^2 \phi^2) \) is the axion energy density. Because we neglect general relativistic corrections, our conclusions are applicable only for processes well within the horizon. Substituting \( \phi \) and \( \pi \) by their expansions in terms of creation and annihilation operators, Eqs. (4–6), and dropping again all axion number violating terms, Eq. (4–11)
becomes

\[
H_g = \sum_{\vec{n}_1, \vec{n}_2, \vec{n}_3, \vec{n}_4} \frac{1}{4} \Lambda_{g \vec{n}_1, \vec{n}_2}^{\vec{n}_3, \vec{n}_4} a_{\vec{n}_1}^\dagger a_{\vec{n}_2} a_{\vec{n}_3} a_{\vec{n}_4} \tag{4-12}
\]

where

\[
\Lambda_{g \vec{n}_1, \vec{n}_2}^{\vec{n}_3, \vec{n}_4} = -\frac{4\pi GM^2}{V} \delta_{\vec{n}_1, \vec{n}_2, \vec{n}_3, \vec{n}_4} \left( \frac{1}{|\vec{p}_{\vec{n}_1} - \vec{p}_{\vec{n}_3}|^2} + \frac{1}{|\vec{p}_{\vec{n}_1} - \vec{p}_{\vec{n}_4}|^2} \right) . \tag{4-13}
\]

In summary so far, the axion field is equivalent to a large number of coupled oscillators with Hamiltonian of the form

\[
H = \sum_{j=1}^{M} \omega_j a_j^\dagger a_j + \sum_{i,j,k,l} \frac{1}{4} \Lambda_{kl}^{ij} a_i^\dagger a_j^\dagger a_j a_i \tag{4-14}
\]

In particular, the total number of quanta

\[
N = \sum_{j=1}^{M} a_j^\dagger a_j \tag{4-15}
\]

is conserved. In Eq. (4–14), \( \Lambda_{kl}^{ij} = \Lambda_{ij}^{kl} = \Lambda_{ij}^{kl} * \). The question of interest now is the following: starting with an arbitrary initial state, how quickly will the averages \( \langle N_k \rangle \) of the oscillator occupation numbers \( N_k = a_k^\dagger a_k \) approach a thermal distribution? The usual approach to this question uses the Boltzmann equation. However, we will see that the assumptions underlying the Boltzmann equation are not valid for the cold axion fluid. So we need a more general approach.

It is instructive to start with a system of just four oscillators \( M = 4 \) and one interaction between them:

\[
H = \sum_{j=1}^{4} \omega_j a_j^\dagger a_j + \Lambda (a_1^\dagger a_2^\dagger a_3 a_4 + a_3^\dagger a_4^\dagger a_1 a_2) \tag{4-16}
\]

We have in this case

\[
\dot{a}_1 = i[H, a_1] = i(-\omega_1 a_1 - \Lambda a_2^\dagger a_3 a_4) \tag{4-17}
\]

and therefore

\[
\dot{N}_1 = i\Lambda (a_1 a_2^\dagger a_3 a_4^\dagger - a_1^\dagger a_2 a_3 a_4) \tag{4-18}
\]
and similar equations for the other $\dot{a}_j$ and $\dot{N}_j$. We solve the equations perturbatively up to $O(\Lambda^2)$. Let us define
\[
\dot{a}_j(t) = (A_j + B_j(t))e^{-i\omega_j t} + O(\Lambda^2) \tag{4–19}
\]
where $A_j \equiv a_j(0)$ and $B_j(t)$ are respectively zeroth and first order, and $B_j(0) = 0$.

Eq. (4–17) implies
\[
\dot{B}_1 = -i\Lambda A_2^\dagger A_3 A_4 e^{+i\Omega t} + O(\Lambda^2) \tag{4–20}
\]
with $\Omega \equiv \omega_1 + \omega_2 - \omega_3 - \omega_4$, and therefore
\[
\dot{B}_1(t) = -i\Lambda A_2^\dagger A_3 A_4 e^{+\kappa t/2} \frac{2}{\Omega} \sin\left(\frac{\Omega t}{2}\right) + O(\Lambda^2) \tag{4–21}
\]
Substituting this into Eq. (4–18), we have
\[
\dot{N}_1 = i\Lambda (A_1 A_2^\dagger A_3^\dagger A_4^\dagger e^{-i\kappa t} - h.c.) + \Lambda^2 [A_2^\dagger A_3 A_3^\dagger A_2^\dagger + A_1 A_2^\dagger A_3^\dagger A_4^\dagger + A_2 A_3^\dagger A_4^\dagger A_1^\dagger - A_1 A_2^\dagger A_4^\dagger A_3^\dagger - A_1 A_2^\dagger A_3^\dagger A_3^\dagger A_2^\dagger] e^{+\kappa t/2} \frac{2}{\Omega} \sin\left(\frac{\Omega t}{2}\right) + h.c. + O(\Lambda^3) \tag{4–22}
\]
Eq. (4–22) may be recast in the form
\[
\dot{N}_1 = i\Lambda (A_1 A_2 A_3^\dagger A_4^\dagger e^{-i\kappa t} - h.c.) + \Lambda^2 [N_3(N_1 + 1)(N_2 + 1) - N_1 N_2 (N_3 + 1)(N_4 + 1)] \frac{2}{\Omega} \sin(\Omega t) + O(\Lambda^3) \tag{4–23}
\]
by rewriting the second order terms.

We now generalize to a system with an arbitrarily large number $M$ of coupled oscillators, Eqs. (4–14). The calculation is essentially the same as for the $M = 4$ toy model, except that there is a multiplicity of interaction terms to keep track of. One finds
(\ell = 1 \ldots M)

\[ \mathcal{N}_i = i \sum_{i,j,k=1}^{M} 1/2 \left( \Lambda^k_{ij} A_i^\dagger A_j^\dagger A_k A_j e^{-i\Omega_{ij}^{kl} t} - h.c. \right) \]

+ \sum_{k,i,j=1}^{M} 1/2 \left| \Lambda^k_{ij} \right|^2 \left[ \mathcal{N}_i \mathcal{N}_j (\mathcal{N}_i + 1)(\mathcal{N}_k + 1) - \mathcal{N}_j \mathcal{N}_k (\mathcal{N}_i + 1)(\mathcal{N}_j + 1) \right] \frac{2}{\Omega_{ij}^{kl}} \sin(\Omega_{ij}^{kl} t)

+ \sum_{k,i,j=1}^{M} \sum_{p,mn=1}^{M} \left[ \frac{1}{2} \Lambda^k_{ij} \Lambda^p_{mn} A_i^\dagger A_j^\dagger A_k^\dagger A_p A_i A_j e^{i(\Omega_{ij}^{kl} + \Omega_{mn}^{kl} / 2) t} \right] \frac{1}{\Omega_{ip}^{kl}} \sin(\frac{\Omega_{mn}^{kl}}{2} t) + h.c.

+ \sum_{k,i,j=1}^{M} \sum_{p,mn=1}^{M} \left[ \frac{1}{2} \Lambda^k_{ij} \Lambda^p_{mn} A_i^\dagger A_j^\dagger A_m A_n A_p A_i A_j e^{i(\Omega_{ij}^{kl} + \Omega_{mn}^{kl} / 2) t} \right] \frac{1}{\Omega_{mn}^{kl}} \sin(\frac{\Omega_{mn}^{kl}}{2} t) + h.c.

- \sum_{k,i,j=1}^{M} \sum_{p,mn=1}^{M} \left[ \frac{1}{2} \Lambda^k_{ij} \Lambda^p_{mn} A_i^\dagger A_j^\dagger A_p A_m A_n A_j e^{i(\Omega_{ij}^{kl} + \Omega_{mn}^{kl} / 2) t} \right] \frac{1}{\Omega_{mn}^{kl}} \sin(\frac{\Omega_{mn}^{kl}}{2} t) + h.c.

- \sum_{k,i,j=1}^{M} \sum_{p,mn=1}^{M} \left[ \frac{1}{2} \Lambda^k_{ij} \Lambda^p_{mn} A_i^\dagger A_j^\dagger A_m A_n A_p A_i A_j e^{i(\Omega_{ij}^{kl} + \Omega_{mn}^{kl} / 2) t} \right] \frac{1}{\Omega_{mn}^{kl}} \sin(\frac{\Omega_{mn}^{kl}}{2} t) + h.c.

+ \mathcal{O}(\Lambda^3) \quad . \quad \quad (4-24)

where \( \Omega_{ij}^{kl} \equiv \omega_k + \omega_j - \omega_i - \omega_j \). The double sums are absent in the toy model because there is only one interaction in that case. At any rate, the double sums will not play an important role in the discussion that follows.

4.2.2 The particle kinetic regime

In most physical systems, the rate of change of the occupation number for a particular state is small compared to the energy exchanged in the transition, so

\( \Omega_{ij}^{kl} t \gg 1 \). Let us call this the ‘particle kinetic regime’. In the particle kinetic regime, the first order terms and off diagonal second terms average to zero in time. Energy is conserved in each transition because

\[ \frac{2}{\Omega_{ij}^{kl}} \sin(\Omega_{ij}^{kl} t) \to 2 \pi \delta(\Omega_{ij}^{kl}) \]  \quad (4-25)
for $\Omega_{ij}^{kl} t \to \infty$. We have then

$$< \tilde{N}_i > = \frac{1}{2} \sum_{i,j,k=1}^{M} |\Delta_{ij}^{kl}|^2 [\hat{N}_i \hat{N}_j (\hat{N}_i + 1)(\hat{N}_j + 1) - \hat{N}_i \hat{N}_k (\hat{N}_i + 1)(\hat{N}_k + 1)] 2\pi \delta (\Omega_{ij}^{kl})$$

(4–26)

where the average on the LHS of this equation is a time average. 4–26 is an operator equation.

After substituting the interactions and taking the infinite volume limit, we recover the Boltzmann equation [17]. For example, the $\phi^4$ self-interaction leads to:

$$< \tilde{N}_1 > = \frac{1}{2 \omega_1} \int \frac{d^3 p_2}{(2\pi)^3 2\omega_2} \int \frac{d^3 p_3}{(2\pi)^3 2\omega_3} \int \frac{d^3 p_4}{(2\pi)^3 2\omega_4} \lambda^2 (2\pi)^4 \delta^4 (p_1 + p_2 - p_3 - p_4) \frac{1}{2}$$

$$[ (\hat{N}_1 + 1)(\hat{N}_2 + 1)\hat{N}_3 \hat{N}_4 - \hat{N}_1 \hat{N}_2 (\hat{N}_3 + 1)(\hat{N}_4 + 1) ]$$

(4–27)

In the non-relativistic limit, the $a + a \to a + a$ cross-section due to $\lambda \phi^4$ interaction is

$$\sigma_\lambda = \frac{1}{|v_1 - v_2|} \frac{1}{2 \omega_1} \frac{1}{2 \omega_2} \int \frac{d^3 p_3}{(2\pi)^3 2\omega_3} \int \frac{d^3 p_4}{(2\pi)^3 2\omega_4} \lambda^2 (2\pi)^4 \delta^4 (p_1 + p_2 - p_3 - p_4)$$

$$\simeq \frac{\lambda^2}{64 \pi m^2}$$

(4–28)

The particle density in physical space is

$$n = \int \frac{d^3 p}{(2\pi)^3} \hat{N}_\rho$$

(4–29)

If most states are not occupied, 4–27,4–28 and 4–29 imply the usual expression for the relaxation rate

$$\Gamma \sim \frac{\hat{N}}{\hat{N}} \sim n \sigma \delta \nu$$

(4–30)

where $\delta \nu$ is a measure of the velocity dispersion of the system.

If the momentum states are highly occupied, such as in the case of cold axions ($\hat{N}_\rho \sim 10^{61}$), the relaxation rate is multiplied by one factor of $\hat{N}$:

$$\Gamma \sim \frac{\hat{N}}{\hat{N}} \sim n \sigma \delta \nu \hat{N}$$

(4–31)
4.2.3 The condensed regime

The condensed regime refers to the case where the energy spread of the highly occupied states is small compared to the evolution rate of the system: \( \delta \omega < \Gamma \). For transitions between such closely spaced states,

\[
e^{-i \Omega_{ij} t} = 1. \tag{4–32}
\]

The first order terms in that equation no longer average to zero. Let us define \( c_i(t) \equiv a_i(t) e^{i \omega_i t} \). The Hamiltonian implies

\[
\dot{c}_i(t) = -i \sum_{k,i,j=1}^{M} \frac{1}{2} \Lambda_{ij}^{k} C_k^{*} C_j e^{i \Omega_{ij} t}. \tag{4–33}
\]

Let us define further

\[
c_i(t) \equiv C_i(t) + d_i(t) \tag{4–34}
\]

where the \( d_i(t) \), like the \( c_i(t) \), are annihilation operators satisfying canonical equal time commutation relations and the \( C_i(t) \) are complex c-number functions which satisfy the classical equations of motion

\[
\ddot{C}_i(t) = -i \sum_{k,i,j=1}^{M} \frac{1}{2} \Lambda_{ij}^{k} C_k^{*} C_j e^{i \Omega_{ij} t}. \tag{4–35}
\]

For the highly occupied cold axion states, the \( C_i \) have magnitude of order \( \sqrt{N} \). The relaxation rate of the highly condensed cold axions is the inverse of the time scale over which those \( C_i(t) \) change by order \( \sqrt{N} \).

Since the sum in Eq. (4–35) is dominated by terms for which \( k, i \) and \( j \) label highly occupied axion states,

\[
\dot{C}_i(t) \sim -i \sum_{k,i,j=1}^{K} \frac{1}{2} \Lambda_{ij}^{k} C_k^{*} C_j. \tag{4–36}
\]

For \( \lambda \phi^4 \) self-interactions, we substitute Eq. (4–10). This yields

\[
\dot{C}_{\lambda}(t) \sim +i \frac{\lambda}{4 m^2 V} \sum_{\rho_2, \rho_3} \frac{1}{2} C_{\rho_2}^{*} C_{\rho_3} C_{\rho_4}. \tag{4–37}
\]
where $\bar{\rho}_4 = \bar{\rho}_1 + \bar{\rho}_2 - \bar{\rho}_3$, and the sum is restricted to the $K$ highly occupied states for which $\rho \lesssim \rho_{\text{max}}$. We may think of the terms on the RHS of Eq. (4–37) as steps in a random walk in complex space. The magnitude of each step is of order $N^{3/2}$ and the number of steps is of order $K^2$. Hence

$$\mathcal{C}_\beta \sim \frac{\lambda}{4m^2V} K N^{3/2} \sim \frac{\lambda}{4m^2V} N N^{3/2}. \quad (4–38)$$

Hence our estimate for the relaxation rate due to $\lambda\phi^4$ self-interactions in the condensed regime:

$$\Gamma_\lambda \sim \frac{1}{4} n \lambda m^{-2}. \quad (4–39)$$

where $n = N/V$ is the density of particles in the highly occupied closely spaced states. Likewise, using Eq. (4–13), we find that through gravitational self-interactions

$$\mathcal{C}_\phi(t) \sim +i \frac{4\pi Gm^2}{V} \sum_{\bar{\rho}_2,\bar{\rho}_3} \frac{1}{2} C_{\bar{\rho}_2} C_{\bar{\rho}_3} C_{\bar{\rho}_4} \left( \frac{1}{|\bar{p}_2 - \bar{p}_3|^2} + \frac{1}{|\bar{p}_1 - \bar{p}_4|^2} \right). \quad (4–40)$$

The corresponding relaxation rate is

$$\Gamma_g \sim 4\pi G m^2 \ell^2. \quad (4–41)$$

where $\ell \sim \frac{1}{\rho_{\text{max}}}$ is the correlation length of the particles.

We note that at the boundary between the particle kinetic and condensed regimes, where $\delta\omega \sim \Gamma$, the two estimates of the relaxation rate agree with one another. Indeed at that boundary, up to factors of order two or so,

$$\delta \nu N \sim \delta \nu \frac{n}{(\delta p)^3} \sim \frac{n}{m^2 \delta \omega} \sim \frac{n}{m^2 \Gamma}. \quad (4–42)$$

Substituting this into Eq. (4–31) for the relaxation rate due to $\lambda\phi^4$ self-interactions in the particle kinetic regime yields Eq. (4–39) which is the corresponding estimate in the condensed regime. The same holds true for the gravitational self-interactions.

Let us also note that Eqs. (4–39) and (4–41) are not valid when almost all axions are in a single state ($K = 1$), as when the Bose-Einstein condensation has been
completed. Indeed, if $K = 1$, there is only one term in the sum on the RHS of Eqs. (4–37) and (4–40) that is enhanced by large occupation numbers, i.e. the term for which both $\vec{p}_2$ and $\vec{p}_3$ equal the momentum of the single highly occupied state, and it describes an interaction with zero momentum transfer. Thus, once the Bose-Einstein condensation is complete and all axions are in the lowest energy state, any further thermalization is suppressed.

Finally consider transitions $a(\vec{p}_1) + a(\vec{p}_2) \leftrightarrow a(\vec{p}_3) + a(\vec{p}_4)$ where $\vec{p}_2$ and $\vec{p}_4$ are momenta of highly occupied states but $\vec{p}_1$ and $\vec{p}_3$ are not. Such transitions are in the condensed regime because the momentum transfer, and hence the energy transfer, is small. Eqs. (4–37) and (4–40) apply to such transitions and imply that the rate at which states with $\rho > \rho_{\text{max}}$ modify their occupation numbers is also given by Eqs. (4–39) and (4–41) with the proviso that the quanta can only move between states differing in momentum by less than $\rho_{\text{max}}$.

4–41 has a simple interpretation. The axions, having energy density $\rho = \rho_{\text{max}}$ and correlation length $\ell$, produce gravitational fields $g \sim 4\pi G \rho \ell$. The gravitational force on a particle is of order $g \omega$, where $\omega$ is the energy of the particle. Since the force is the rate of change of the particle’s momentum, the relaxation rate is of order

$$\Gamma_g \sim \frac{g\omega}{\Delta \rho} \sim 4\pi G \rho m \ell \frac{\omega}{\Delta \rho} \quad (4–43)$$

where $\Delta \rho$ is the momentum dispersion of the particles. For the axions themselves, we obtain the relaxation rate Eq. (4–41) by substituting $\omega = m$ and $\Delta \rho \sim \ell^{-1}$. Eq. (5–20) shows that the momentum distribution of any particle species is modified by the gravitational fields of the cold axion fluid and therefore that gravitational interactions may produce thermal contact between the cold axions and other particle species.
4.2.4 Cold axions form a BEC

Using 4–41 one finds that the thermalization condition: $\Gamma_g / H > 1$ is satisfied at a time $t_{\text{BEC}}$ when the photon temperature is of order

$$T_{\text{BEC}} \sim 500 \text{ eV} \times \left( \frac{f_a}{10^{12} \text{ GeV}} \right)^{\frac{1}{2}}. \quad (4–44)$$

The axions thermalize then and form a BEC as a result of their gravitational self-interactions. The whole idea may seem far-fetched because we are used to think that gravitational interactions among particles are negligible. The axion case is special, however, because almost all particles are in a small number of states with very long de Broglie wavelength, and gravity is long range. By gravitational self-interactions the axions modify their momentum distribution till their entropy is maximized for the available energy, which in this case means that they form a BEC.

Axion BEC causes the correlation length to increase. Indeed in an infinite volume, when all particles are in the lowest energy state, the momentum dispersion is theoretically zero and the correlation length infinite. This ideal state never occurs because thermalization and hence BEC formation are constrained by causality. The axions in one horizon are unaware of the doings of axions in the next horizon. Hence we expect the correlation length $\ell$, which may now be thought of as the size of condensate patches, to become of order but less than the horizon. The growth in the correlation length causes the thermalization to accelerate. Once $\ell$ is some fraction of $t$, $\Gamma_g(t) / H(t) \propto a(t)^{-3} t^3$, implying that thermalization occurs on ever shorter time scales compared to the Hubble time.

4.3 Thermal contact with other species

4.3.1 Evolution equations for other species

Our purpose in this subsection is to estimate the gravitational interaction rates of other species - baryons, relativistic axions and photons - with the cold axion fluid.
The Hamiltonian describing gravitational interactions between the cold axions and any other species has the general form:

\[ H = \sum_{j=1}^{M} \omega_j a_j^\dagger a_j + \sum_{r=1}^{S} \omega_r b_r^\dagger b_r + \sum_{i,j,k,l} \frac{1}{4} \Lambda_{ijkl}^{rs} a_k^\dagger a_l^\dagger a_i a_j + \sum_{j,k,r,s} \Lambda_{b,ks}^r a_k^\dagger b_s^\dagger a_j b_r \]  

(4–45)

where \( \Lambda_{b,ks}^r = (\Lambda_{b,sr}^{ks})^* \). The \( b_r \) are the annihilation operators for quanta of the new species. They satisfy canonical commutation or anti-commutation relations. The \( \omega_r \) are the energies of those quanta. As before, we quantize in a box of volume \( V = L^3 \) with periodic boundary conditions. The labels of the new particle states are then \( r = (\vec{n}, \sigma) \), giving their momenta \( \vec{p} = \frac{2\pi}{L} \vec{n} \) and their spin \( \sigma \). Their energy is \( \omega = \sqrt{\vec{p} \cdot \vec{p} + m_b^2} \) where \( m_b \) is the mass of the new species.

We define \( c_j(t) \equiv a_j(t)e^{i\omega_j t} \) as before, and \( c'_r(t) \equiv b_r(t)e^{i\omega_r t} \). The Heisenberg equations of motion for the \( c'_r(t) \) are then:

\[ \dot{c}'_s = -i \sum_{j,k,r} \Lambda_{b,ks}^r c_j^\dagger c'_r e^{i\Omega_{r,s}' t} \]  

(4–46)

where \( \Omega_{r,s}' \equiv \omega_k + \omega_s - \omega_j - \omega_r \). Because 3-momentum is conserved in each interaction, the \( \Lambda_{b,ks}^r \) have the form:

\[ \Lambda_{b,ks}^r = -\lambda_{b,ks}^r \delta_{\vec{p}_j + \vec{p}_k - \vec{p}_r} . \]  

(4–47)

The important contributions in the sum on the RHS of Eq. (4–46) are from terms in which both \( j \) and \( k \) label highly occupied cold axion states. Therefore

\[ \dot{c}'_s \sim +i \sum_{k,j=1}^{K} \lambda_{b,ks}^r C_k^* C'_r e^{i(\omega_k - \omega_r)t} \]  

(4–48)

with \( \vec{p}_r = \vec{p}_s + \vec{p}_k - \vec{p}_j \). As before, \( K \) is the number of highly occupied cold axion states and the \( C_k \) are defined by Eq. (4–34). Again, the sum on the RHS of Eq. (4–48) represents a random walk in complex space. The number of steps is \( K^2 \) and the typical step size is \( \lambda_b N c' \), where \( \lambda_b \) is the typical value of \( \lambda_{b,ks}^r \). The rate at which all quanta of the new species may move to neighboring states separated in momentum space by less...
than $\delta \rho \sim \frac{1}{\ell}$ is therefore
\[ \Gamma_{b,\delta \rho} \sim K \mathcal{N} \lambda_{b} = \lambda_{b} \mathcal{N} \] (4–49)
where $\mathcal{N} = K \mathcal{N}$ is the number of cold axions in volume $V$. The relaxation rate of the new species is then
\[ \Gamma_{b} \sim \lambda_{b} \mathcal{N} \frac{\delta \rho}{\Delta \rho} \sim \lambda_{b} \mathcal{N} \frac{1}{\ell \Delta \rho} \] (4–50)
where $\Delta \rho$ is the momentum dispersion of the new species population. (If the momentum dispersion is very different in the initial state than in the final state, $\Delta \rho$ is the larger of the two.) Eq. (4–50) assumes that the $b$ particles are bosons or non-degenerate fermions. If they are degenerate fermions, their relaxation rate is suppressed, relative to Eq. (4–50), by Pauli blocking.

Also, let us reiterate that when most cold axions are in the lowest energy state, implying $K = 1$, thermalization is suppressed compared to the estimate in Eq. (4–50), because there is only one term in the sum of Eq. (4–48) in that case and the momentum transfer vanishes for that term.

### 4.3.1.1 Baryons

For non-relativistic species, such as baryons and WIMPs, the term in the Hamiltonian that describes gravitational interactions with the cold axions is
\[ H_{B} = -G \int d^{3}x \, d^{3}x' \frac{\rho(\vec{x}, t) \rho_{B}(\vec{x}', t)}{|\vec{x} - \vec{x}'|} \] (4–51)
where
\[ \rho_{B}(\vec{x}, t) = \frac{m_{B}}{V} \sum_{\vec{r}, \vec{r}', \sigma} b_{\vec{r}, \sigma}^\dagger b_{\vec{r}', \sigma} e^{i(\vec{p}' - \vec{p}) \cdot \vec{x}} \] (4–52)
and $m_{B}$ is the mass of the non-relativistic particle. This yields
\[ \chi_{B}^{\vec{r}, \sigma}_{\vec{r}', \sigma'} = \frac{4 \pi G m_{B}}{V q^{2}} \delta_{\sigma, \sigma'} \] (4–53)
where \( \vec{q} = \vec{p}_1 - \vec{p}_3 \) is the momentum transfer. Since \( q \sim \ell^{-1} \), the \( B \) particles have relaxation rate
\[
\Gamma_B \sim 4\pi Gm\bar{m}_B \frac{\ell}{\Delta \rho_B} , \tag{4–54}
\]
where \( \Delta \rho_B \) is their momentum dispersion. Eq. (4–54) assumes that the \( B \) particles are bosons or non-degenerate fermions, as is the case for baryons and WIMPs.

### 4.3.1.2 Hot axions

For relativistic species, the term that describes gravitational interactions with the cold axion fluid is
\[
H_r = - \int d^3x \frac{1}{2} h_{\alpha\beta} T^{\alpha\beta} \tag{4–55}
\]
where \( T^{\alpha\beta}(\vec{x}, t) \) is the stress-energy-momentum tensor of this species and \( h_{\alpha\beta} \) is the perturbation of the space-time metric caused by the cold axions:
\[
\begin{align*}
  h_{00}(\vec{x}, t) &= 2G \int d^3x' \frac{\rho(\vec{x}', t)}{|\vec{x} - \vec{x}'|} \\
h_{0k}(\vec{x}, t) &= 0 \\
h_{kk}(\vec{x}, t) &= 2G \int d^3x' \frac{\rho(\vec{x}', t)}{|\vec{x} - \vec{x}'|^3} (x_k - x_k')(x_l - x_l') . \tag{4–56}
\end{align*}
\]
Note that \( h^0_\alpha = 0 \). For a scalar field \( \phi(x) \)
\[
H_r = - \int d^3x \frac{1}{2} h_{\alpha\beta} \partial^\alpha \phi \partial^\beta \phi . \tag{4–57}
\]
After some algebra, Eq. (4–57) yields
\[
\chi_{\tilde{r}_{\tilde{r}_1, \tilde{r}_2}} = \frac{4\pi Gm}{\sqrt{q^2 \omega_2 \omega_4}} \left[ \omega_2 \omega_4 + \tilde{p}_2 \cdot \tilde{p}_4 - 2 \frac{\tilde{q} \cdot \tilde{p}_2 \tilde{q} \cdot \tilde{p}_4}{q^2} \right] , \tag{4–58}
\]
where \( \tilde{q} = \vec{p}_1 - \vec{p}_3 \). The relaxation rate for relativistic scalars through gravitational interactions with the highly occupied low momentum axion modes is thus of order
\[
\Gamma_r \sim 4\pi Gm\ell , \tag{4–59}
\]
since \( q \sim \ell^{-1} \) and \( \Delta \rho \sim \omega \).
4.3.1.3 Photons

The term that describes the gravitational interactions of photons with the cold axion fluid is

\[ H_\gamma = - \int d^3x \frac{1}{2} h_{\alpha\beta} F^{\alpha\mu} F^{\beta\mu}_\mu \] (4–60)

where \( h_{\alpha\beta} \) is the electromagnetic field strength tensor, and the \( h_{\alpha\beta} \) are given by Eqs. (4–56) as before. This yields

\[ \lambda_\gamma = \frac{8\pi G m}{\sqrt{q^4 \sqrt{\omega_2 \omega_4}}} \left[ \omega_2 \omega_4 (\vec{e}_2 \cdot \vec{q})(\vec{e}_4^* \cdot \vec{q}) + \left( \vec{p}_2 \times \vec{e}_2 \right) \cdot \vec{q} \left( \vec{p}_4 \times \vec{e}_4^* \right) \cdot \vec{q} \right] , \] (4–61)

where \( \vec{e}_2 \) and \( \vec{e}_4 \) are the polarization vectors of the initial and final state photons. We find therefore

\[ \Gamma_\gamma \sim 4\pi G m \ell \] (4–62)

for the relaxation rate of photons. It is the same as for relativistic axions, Eq. (4–59), in order of magnitude.

4.3.2 Possible outcomes

The rate at which non-relativistic species such as baryons change their momentum distribution through gravitational interactions with the cold axion fluid is given by Eq. (4–54). The momentum dispersion of baryons is of order \( \Delta p \sim \sqrt{3m_B T} \) where \( T \) is the photon temperature. We will assume here that cold axions are the bulk of the dark matter. The Friedmann equation implies then

\[ 4\pi G m \sim \frac{3}{8} t^3 \left( \frac{t}{t_{eq}} \right)^3 \] (4–63)

for \( t < t_{eq} \), where \( t_{eq} \) is the time of equality between matter and radiation. Hence

\[ \Gamma_B/H \sim \frac{1}{4} \frac{\ell}{4\sqrt{tt_{eq}}} \sqrt{\frac{3m_B}{T}} \sim \frac{\ell}{t} \left( \frac{t}{t_{BEC}} \right)^\frac{3}{2} \left( \frac{500 \text{ eV}}{T_{BEC}} \right)^\frac{3}{2} . \] (4–64)

Eq. (4–64) shows that baryons reach thermal contact with the axion BEC when \( \ell \) becomes of order \( t \) or soon after that.
Photons are in thermal contact with the baryons, but the nature and degree of this thermal contact are changing at the time of axion BEC formation [51]. Baryons interact with electrons by Coulomb scattering. Electrons interact with photons by Compton scattering, double Compton scattering and bremsstrahlung. Above approximately 1 keV photon temperature, double Compton scattering and bremsstrahlung assure chemical equilibrium between baryons and photons (the number of photons is not conserved in these processes). Below approximately 1 keV photon temperature, Compton scattering is the only important interaction remaining. It maintains kinetic, but not chemical, equilibrium between baryons and photons till approximately 100 eV photon temperature. Below 100 eV, the degree of kinetic equilibrium progressively diminishes till approximately 2 eV, when it disappears altogether.

In any case, as long as there is only thermal contact between baryons and a few low momentum modes of the axion field, only a very small amount of energy can be exchanged between the axion field and the other species. However, as time goes on, higher and higher momentum modes of the axion field reach thermal contact with its highly occupied low momentum modes. Eq. (4–59) gives the relaxation rate \( \Gamma_r \) of the axion field as a whole, including relativistic states. The relaxation rate of photons \( \Gamma_\gamma \), Eq. (4–62), is of the same order of magnitude. Combining Eqs. (4–59) and (4–62) with the Friedmann equation, we have

\[
\frac{\Gamma_\gamma}{H} \sim \frac{\Gamma_r}{H} \sim \frac{3}{2} H \ell \frac{\rho_a}{\rho_{\text{tot}}} ,
\]

where \( \rho_a \) is the cold axion density and \( \rho_{\text{tot}} \) the total density. Since \( \ell \propto t \), Eq. (4–65) implies that \( \Gamma_\gamma/H \) and \( \Gamma_r/H \) grow proportionally to \( a(t) \) till equality and remain constant after that. The critical parameter is their value at equality

\[
\left. \frac{\Gamma_\gamma}{H} \right|_{\text{eq}} \sim \left. \frac{\Gamma_r}{H} \right|_{\text{eq}} \sim \frac{\ell(t_{\text{eq}})}{t_{\text{eq}}} .
\]

(4–66)
Thermal contact between axions and photons is established if the RHS of Eq. (4–66) is of order one, i.e. if the axion correlation length is horizon size at equality. To show that the axion BEC correlation length becomes truly as large as the horizon is a problem, involving both out-of-equilibrium statistical mechanics and general relativity, which we do not know how to solve at present.

Hence we consider at this stage two possibilities, which we call cases A and B. In case A, $\Gamma_{r,\gamma}/H$ do not reach one before equality (because $\ell$, although proportional to $t$, may be much less than $t$, e.g. $\ell = t/100$), and hence thermal contact gets established only between baryons and low momentum modes of the axion field. In case B, $\Gamma_{r,\gamma}/H$ do reach one before equality and thermal equilibrium is reached between baryons, axions and photons. This equilibrium is kinetic only since gravitational interactions conserve particle number for all the species involved.

We should ask whether neutrinos may also reach thermal contact with the highly occupied low momentum axion modes, in which case neutrinos, axions, baryons and photons would all reach the same temperature. We believe this possibility, which we call case C, unlikely for the following reason. Eq. (5–20) does not apply to degenerate fermions because of Pauli blocking. Cosmic neutrinos are semi-degenerate since they have a thermal distribution with zero chemical potential. Because of partial Pauli blocking, their thermalization is slower than that of relativistic axions, Eq. (4–59). Since relativistic axions only barely reach thermal contact with the highly occupied low momentum modes of the axion field if they do so at all, and thermal contact between those low momentum modes of the axion field and neutrinos is delayed relative to relativistic axions, it appears most likely that neutrinos remain decoupled from the axions at all times.
CHAPTER 5
IMPLICATIONS FOR OBSERVATION

5.1 A non-rethermalizing axion BEC behaves as ordinary CDM

In this section, we show that axion BEC behaves as ordinary CDM on all scale of observational interest as long as the axions remain in the same state, i.e. as long as they do not rethermalize [52].

The axion field satisfies the Heisenberg equation of motion:

$$D^\mu D_\mu \varphi(x) = g^{\mu\nu} [\partial_\mu \partial_\nu - \Gamma^\lambda_{\mu\nu} \partial_\lambda] \varphi(x) = m^2 \varphi(x) \ .$$  (5–1)

where $\varphi(x)$ is the axion field operator. The self interaction term $-\frac{1}{6} \lambda \varphi^3$ is only important for a very short period after QCD time. It is neglected here.

The axion field can be expanded in particle modes as

$$\varphi(x) = \sum_\alpha [a^\dagger_\alpha \Phi_\alpha(x) + a_\alpha \Phi^*_\alpha] \ .$$  (5–2)

Except for a tiny fraction, most axions go to a single state that we label as $\alpha = 0$. $\phi_0(x)$ is the corresponding wave function. The state of the axion field is $| N > = (1/\sqrt{N!})(a^\dagger_0)^N|0 >$ where $|0 >$ is the vacuum and $N$ is the particle number. In the spatially flat homogeneous and isotropic Robertson-Walker space-time,

$$\Phi_0 = \frac{A}{a(t)^{\frac{n}{2}}} e^{-imt} \ .$$  (5–3)

where $A$ is a constant. The stress-energy-momentum tensor has expectation value:

$$< N | T_{\mu\nu} | N > = N [\partial_\mu \Phi^*_0 \partial_\nu \Phi_0 \\
+ \partial_\nu \Phi^*_0 \partial_\mu \Phi_0 + g_{\mu\nu} (\partial_\lambda \Phi^*_0 \partial^\lambda \Phi_0 - m^2 \Phi^*_0 \Phi_0)] \ .$$  (5–4)

In a flat Minkowski space-time and in the non-relativistic limit, neglecting terms of order $(1/m) \partial_t$ compared to terms of order one, (5–1) implies the Schrödinger equation:

$$i \partial_t \psi = -\frac{\nabla^2}{2m} \psi \ .$$  (5–5)
The wave function can be written as:

$$\Psi(\vec{x}, t) = \frac{1}{\sqrt{2\pi N}} B(\vec{x}, t) e^{i\beta(\vec{x}, t)} .$$  \hspace{1cm} (5–6)

In terms of $B(\vec{x}, t)$ and $\beta(\vec{x}, t)$ the density and velocity fields of the axion fluids are:

$$\rho = mB(\vec{x}, t), \text{ and } \vec{v}(\vec{x}, t) = \frac{1}{m} \vec{\nabla} \beta(\vec{x}, t).$$ Then (5–5) leads to the continuity equation, and the equation of motion:

$$\partial_t v^k + v^j \partial_j v^k = -\vec{\nabla} q$$ \hspace{1cm} (5–7)

where

$$q(\vec{x}, t) = -\frac{\nabla^2 \sqrt{\rho}}{2m^2 \sqrt{\rho}} .$$ \hspace{1cm} (5–8)

Following the motion, the stress tensor is

$$T_{jk} = \rho \nu_j \nu_k + \frac{1}{4m^2} \left( \frac{1}{\rho} \partial_j \rho \partial_k \rho - \delta_{jk} \nabla^2 \rho \right) .$$ \hspace{1cm} (5–9)

For ordinary CDM, the last terms on the RHS of (5–7) and (5–9) are absent. In the linear regime of evolution within the horizon, neglecting second order terms, (5–9) becomes

$$\delta T_{jk} = -\delta_{jk} \frac{\rho_0(t)}{4m^2} \nabla^2 \delta(\vec{x}, t)$$ \hspace{1cm} (5–10)

where $\rho_0(t)$ is unperturbed density and $\delta(\vec{x}, t) \equiv \frac{\delta \rho(\vec{x}, t)}{\rho_0(t)}$. Because the RHS of (5–7) is a gradient and RHS of (5–10) is proportional to the Kronecker symbol, the vector and tensor perturbations are not affected by the additional forces associated with the axion BEC in the linear regime. Only the scalar perturbations are affected. The scalar perturbations are conveniently described in conformal Newtonian gauge where the metric is

$$ds^2 = -(1 + 2\psi(\vec{x}, t)) dt^2 + a(t)^2 (1 + 2\phi(\vec{x}, t)) d\vec{x} \cdot d\vec{x} .$$ \hspace{1cm} (5–11)
Conservation of energy and momentum in this background implies the first order equations

\[ \partial_t \delta + \frac{1}{a} \vec{\nabla} \cdot \vec{\nabla} = -3 \partial_t \phi + \frac{3H}{4m^2 a^2} \nabla^2 \delta \]
\[ \partial_t \vec{\nabla} + H \vec{\nabla} = -\frac{1}{a} \vec{\nabla} \psi + \frac{1}{4m^2 a^3} \vec{\nabla} \nabla^2 \delta \]  

(5–12)

where \( H = \frac{1}{3} \frac{da}{dt} \). The equations for CDM are recovered by letting \( m \to \infty \). The RHS of Einstein’s equations are modified by the addition of \( \delta T_{jk} \) to the stress tensor, but this modification does not play a role in our discussion because it is suppressed, relative to the leading terms, by the factor \( \left( \frac{k_{ph}}{m} \right)^2 \), where \( k_{ph} \) is the physical wavevector of the perturbation. For scales within the horizon, one obtains from (5–12):

\[ \partial_t^2 \delta + 2H \partial_t \delta - \left( 4\pi G \rho_0 - \frac{k^4}{4m^2 a^4} \right) \delta = 0 \]  

(5–13)

(5–13) implies that the axion BEC has a Jeans length:

\[ k_j^{-1} = (16\pi G \rho m^2)^{-\frac{1}{2}} \]
\[ = 1.02 \cdot 10^{14} \text{ cm} \left( \frac{10^{-5} \text{ eV}}{m} \right)^{\frac{1}{2}} \left( \frac{10^{-20} \text{ g/cm}^3}{\rho} \right)^{\frac{1}{2}} \]  

(5–14)

The Jeans length is small compared to the smallest observable scales (100 kpc), thus the axion BEC and CDM are indistinguishable in the linear regime. In the nonlinear regime of structure formation, the relevant equations of motion are:

\[ \partial_t \rho + \vec{\nabla} \cdot (\rho \vec{v}) = 0 \quad , \quad \vec{\nabla} \times \vec{v} = 0 \]
\[ \partial_t \vec{\nabla} + (\vec{v} \cdot \vec{\nabla}) \vec{v} = -\vec{\nabla} \psi - \vec{\nabla} q \]  

(5–15)

(5–15) is equivalent to the Schrödinger equation for particles in a Newtonian gravitational field. Axion BEC and CDM differ by the \( \vec{\nabla} q \) term, which indicates a local quantum effect of axion BEC. However, as was shown by numerical simulation and is expected from the WKB approximation, the differences occur only on length scales smaller than the de-Broglie wavelength[4], which is of order \( 1/(m \delta v) \approx 1/(m \cdot 10^{-3} \text{ c}) \approx 10 \text{ m} \).
5.2 Tidal torquing, inner caustics and axion BEC

In this section, we show that axion BEC behaves differently from ordinary CDM when falling into galactic halos because axion BEC rethermalizes as it falls in [46].

Let us consider a region of size $L$ inside of which the axion state stops being the lowest energy available state because the background is time dependent. We expect that the axion BEC rethermalizes provided the gravitational forces produced by the BEC are larger than the typical rate $\dot{\rho}$ of change of axion momenta required for the axions to remain in the lowest energy state. The gravitational forces are of order $4\pi G m^2 \ell$. In this case, the correlation length $\ell$ must be taken to be of order the size $L$ of the region of interest since the gravitational fields due to axion BEC outside the region do not help the thermalization of the axions within the region. Hence the condition is

$$4\pi G m^2 L \gtrsim \dot{\rho} \quad (5–16)$$

We now apply this criterion to the question whether axions rethermalize sufficiently quickly that they share angular momentum when they are about to fall into a galactic gravitational potential well.

We use the self-similar infall model of galactic halo formation to estimate $L$ and $\dot{\rho}$. $L$ is of order a few times the turnaround radius $R(t)$, say $L(t) \sim 3 R(t)$, whereas $\dot{\rho}(t) \sim m v_{\text{rot}}(t) j_{\text{max}}$ where $v_{\text{rot}}$ is the rotation velocity and $j_{\text{max}}$ is the dimensionless number characterizing the amount of angular momentum of the halo. In the self-similar model, $v_{\text{rot}}(t) \sim R(t)/t$ and $R(t) \propto t^{2+\frac{2}{3}} / t^{\frac{2}{3}+\frac{2}{3}}$ where $\epsilon$ is in the range 0.25 to 0.35 [53]. Assuming that most of the dark matter is axions, the Friedmann equation implies

$$4\pi G m = \frac{3}{2} H(t)^2 = \frac{2}{3 t^2} \quad (5–17)$$

for $t > t_{\text{eq}}$. The LHS of Eq. $(5–16)$ is therefore of order

$$2m \frac{R(t)}{t^2} \sim 2mv_{\text{rot}}(t) \frac{1}{t} \quad (5–18)$$
whereas its RHS is of order

\[
\frac{d}{dt} \left[ m j_{\text{max}} v_{\text{rot}}(t_0) \left( \frac{t}{t_0} \right)^{-\frac{1}{2} + \frac{2}{9}} \right] = m j_{\text{max}} v_{\text{rot}}(t) \frac{1}{t} \left( \frac{2}{9} \epsilon - \frac{1}{3} \right).
\]

The typical value of \( j_{\text{max}} \) is 0.18. Hence Eq. (5–16) is satisfied at all times from equality till today by a margin of order \( \frac{2}{j_{\text{max}}(\frac{2}{9} - \frac{1}{3})} \sim 30 \).

We conclude that the axion BEC does rethermalize before falling into the gravitational potential well of a galaxy. Most axions go to the lowest energy state consistent with the total angular momentum acquired from neighboring inhomogeneities through tidal torquing [54]. That state is a state of rigid rotation on the turnaround sphere, implying \( \vec{\nabla} \times \vec{v} \neq 0 \) where \( \vec{v} \) is the velocity field of the infalling axions. In contrast, the velocity field of WIMP dark matter is irrotational. The inner caustics of galactic halos are different in the two cases. Axions produce caustic rings whereas WIMPs produce the ‘tent-like’ caustics described in ref. [71]. There is evidence for the existence of caustic rings in various galaxies at the radii predicted by the self-similar infall model. For a review of this evidence see ref. [55]. It is shown in ref. [70] that the phase space structure of galactic halos implied by the evidence for caustic rings is precisely and in all respects that predicted by the assumption that the dark matter is a rethermalizing BEC.

### 5.3 Axion BEC and cosmological parameters

In this section we show that axion BEC may provide an explanation to the "Lithium Problem" and may change cosmology parameters such as the effective number of neutrino species comparing to the standard cosmology values [56].

The agreement between observations and the BBN predictions for the primordial abundances of light elements is often touted as a triumph of the standard ΛCDM cosmological model. Under the assumption that there are three neutrino species, BBN as a theory requires essentially a single input: the baryon-to-photon ratio, usually given by the parameter \( \eta_0 = 10^{30} n_B/n_\gamma \) [57]. If one takes \( \eta_0 \) to be \( 6.190 \pm 0.145 \), in accordance with the latest Wilkinson Microwave Anisotropy Probe (WMAP) results,
the inferred primordial abundances of the majority of the light elements (D, \(^4\)He, \(^3\)He) are remarkably consistent with BBN predictions, save one exception: that of \(^7\)Li is approximately two to three times less than what the theory predicts. The discrepancy is deemed statistically significant, and there is so far no widely accepted explanation for the anomaly. In the literature, this is referred to as the "Lithium Problem".

One of the most difficult issues involved in testing BBN is how reliably to infer the primordial abundances of light elements from measurements that are available to us. Subsequent to BBN, the original relic abundances are all subject to further modification by complicated stellar processes. \(^7\)Li, for example, can be both depleted and synthesized in stars, as well as produced by cosmic-ray nucleosynthesis. As such, the abundance of \(^7\)Li is inferred primarily from absorption lines in the atmosphere of galactic halo stars with low metallicity, since these stars are very old and have experienced very little nuclear processing.

Although these post-BBN effects lead to considerable complication, they also open up many different avenues to explain the \(^7\)Li anomaly. For many years, it has been hoped that better determination of nuclear parameters will gradually narrow the discrepancy, though it was eventually realized that does not seem achievable [60]. Quite the contrary, it was found in [57] that improved data on the neutron life-time and the cross sections \(p(n, \gamma)d\) and \(^3\)He\((\alpha, \gamma)^7\)Be increases the predicted abundance of \(^7\)Li, worsening the disagreement. Revisions to stellar evolution, as a consequence of systematic errors in the effective temperature of the metal-poor stars [61, 62], and surface \(^7\)Li depletion in the interior of stars due to some mixing or diffusive processes [63], have also been investigated as possible solutions, but are still considered controversial [57].

The fact that the nuclear reactions relevant to the production of both primordial and post-BBN \(^7\)Li are now quite well understood has led to speculations that the anomaly might instead be caused by new physics. Many explanations have been proposed, such
as the variation in time of the deuteron binding energy and of fundamental couplings [64, 65], and the decay of a relatively long-lived particle in the context of supersymmetry [66]. At this point, none of these explanations have won general acceptance in the cosmology community.

One way to remove the conflict between data and theory can be the cooling of photons between the end of BBN and decoupling. Processes that do this are difficult to come by. Indeed, typical processes arising from new physics tend to heat up the photons, modifying $\eta_{10}$ in the wrong direction [57]. However, the recent realization that dark matter axions form a BEC at approximately 500 eV photon temperature [67] provides a possible mechanism [68]. Essentially, the high occupation of axion modes with very low momenta greatly enhances the strength of their gravitational interactions, such that an exchange of energy between the photons and the much colder axions becomes possible. Photon cooling implies that $\eta_{10,\text{BBN}} < \eta_{10,\text{WMAP}}$, which has the effect of reducing the production of $^7\text{Li}$ [59]. If thermal equilibrium between the photons and axions is achieved, the $^7\text{Li}$ abundance is reduced by approximately a factor 2 (see below), alleviating the discrepancy and perhaps removing it altogether. However, our proposal predicts a higher abundance for D than present observations indicate and predicts that the effective number of thermally excited neutrino degrees of freedom is high: $N_{\text{eff}} = 6.77$.

Photon cooling by kinetic mixing with hidden photons was proposed in ref. [69].

5.3.1 Possibility of photon cooling

The gravitational fields of the cold axion fluid cause transitions between momentum states of other particle species present. For particles which are bosons or non-degenerate fermions, the relaxation rate through gravitational interactions with the cold axions is of order [68]

$$\Gamma \sim 4\pi G m_a \frac{\omega}{\Delta p}$$

(5–20)
where $\omega$ is the typical energy of the particles and $\Delta p$ their momentum dispersion. For photons to cool substantially it is necessary that energy is transferred from the photons to the low momentum highly occupied axion states and from those to the relativistic axion states. For both relativistic axion states and for photons, $\Delta p \sim \omega$ and hence their relaxation rate $\Gamma_r$ through gravitational interactions with cold axions is of order $4\pi G n m \ell$.

Using the Friedmann equation, one finds that $\Gamma_r / H \propto a(t)$ before equality between matter and radiation and remains constant after that. At equality, $\Gamma_r / H|_{t_\text{eq}} \sim \ell(t_\text{eq}) / t_\text{eq}$. If $\ell / t$ is order one at equality, the photons reach thermal equilibrium with the axions and hence cool.

Gravitational interactions conserve particle number and therefore produce only kinetic (as opposed to chemical) equilibrium between the species involved. Also, after 500 eV photon temperature, the coupling between photons and baryons is in the kinetic, rather than chemical, equilibrium regime [74]. Upon cooling, the photons that cannot be accommodated in thermally excited states enter the ground state, a plasma oscillation with zero wavevector. Since the photon chemical potential remains zero, the final photon spectrum is Planckian, consistent with observation.

Eq. (5–20) does not apply to degenerate fermions because of Pauli blocking. The cosmic neutrinos are semi-degenerate since they have a thermal distribution with zero chemical potential. Their thermalization rate is less than that $\Gamma_r$ of relativistic bosons. Since $\Gamma_r / H \propto t n \ell \propto t^2 a^{-3}(t)$, that ratio does not grow after equality. Since the relativistic axions may only reach thermal contact with the cold axions at equality and the neutrinos are delayed relative to the relativistic axions, we believe it most likely that neutrinos remain decoupled from the axions, photons and baryons at all times.

It is straightforward to determine how much the photons cool if they reach thermal equilibrium with the axions. Energy conservation implies $\rho_{\text{i,}\gamma} = \rho_{\text{r,}\gamma} + \rho_{\text{r,a}}$ because the contributions to the energy density of the initial axions and of the baryons are negligible. The ratio between the final and initial photon temperature is thus $(2/3)^{1/4}$. Since their
number density is proportional to \( T^3 \), we find:

\[
\eta_{10,\text{BBN}} = \left( \frac{2}{3} \right)^{3/4} \eta_{10,\text{WMAP}} = 4.57 \pm 0.11
\]  

(5–21)

using \( \eta_{10,\text{WMAP}} = 6.190 \pm 0.145 \). Because the \( ^7\text{Li} \) abundance is proportional to \( \eta_{10,\text{BBN}}^2 \) in the range of interest, it is reduced by approximately the factor \( \left( \frac{2}{3} \right)^3 \simeq 0.55 \).

5.3.2 Effect on the other light element primordial abundances

Whether photon cooling by axion BEC solves the Lithium Problem remains to be seen. The data have been time dependent in addition to the usual uncertainties. In Fig 5-1, we plot the value of \( \eta_{10,\text{BBN}} \) in the standard cosmological model, labeled ‘WIMP’, and in the scenario described here, labeled ‘axion’, along with the values inferred from the observed light element abundances according to the review by G. Steigman in 2005 [58], the review by F. Iocco et al. in 2008 [76] and a private communication from G. Steigman updating his 2005 estimates in the light of recent observations [77]. The error bars indicate the range of \( \eta_{10,\text{BBN}} \) consistent with the estimated 1-\( \sigma \) uncertainties in the observations. The axion prediction agrees very well with the \( ^7\text{Li} \) abundance at the time of Steigman’s 2005 review (\( \eta_{10,^7\text{Li}} = 4.50 \pm 0.30 \)). However more recent observations indicate a lower primordial \( ^7\text{Li} \) abundance, worsening the Lithium Problem.

Perhaps more problematic is that a smaller \( \eta_{10,\text{BBN}} \) predicts an overproduction of deuterium (D). Traditionally, D has been the prime choice as a baryometer among the light elements, due to its sensitivity to \( \eta_{10,\text{BBN}} \) and simple post-BBN evolution (abundance monotonically decreasing). The major drawback with D is that its abundance is inferred from a very small set of (seven) spectra of QSO absorption line systems [78]. Worse yet, these few measurements have a large dispersion, and do not seem to correlate with metallicity, obscuring the expected deuterium plateau. Due to the various inadequacies in the D measurements mentioned, we have reservations about the common practice of attaching most significance to D in the comparison between data and BBN predictions. In comparison, \( ^7\text{Li} \) is inferred from a large number
Figure 5-1. Values of $\eta_{10,\text{BBN}}$ inferred from the abundances of $^7\text{Li}$, D, $^3\text{He}$ and $^4\text{He}$, and the predicted values in the standard cosmological model (WIMP) and in our proposal (axion). The data inferred values are taken from refs. [58], [76] and [77]. The error bars indicate the $\eta_{10,\text{BBN}}$ values consistent with the estimated 1-$\sigma$ uncertainties in the observations.

of measurements, which are more-or-less consistent. Also, since D is more easily destructible than $^7\text{Li}$, it is conceivable that unknown stellar processes further deplete D.

Finally the $^3\text{He}$ and $^4\text{He}$ inferred $\eta_{10,\text{BBN}}$ values have large error bars and hence carry less statistical weight. The $^4\text{He}$ inferred value has increased recently ($5.5 < \eta_{10,^4\text{He}} < 11$ according to ref. [76] and $7.5 < \eta_{10,^4\text{He}} < 20$ according to ref. [77]) compared to its accepted value a few years ago, creating additional uncertainty.

5.3.3 Effective number of neutrino species

After the axions are heated up and reach the same temperature as the photons, most of them are still in the ground state. The axions in the ground state behave as cold
dark matter. The axions in the excited states contribute one bosonic degree of freedom to radiation. The radiation content of the universe is commonly given in terms of the effective number $N_{\text{eff}}$ of thermally excited neutrino degrees of freedom, defined by

$$
\rho_{\text{rad}} = \rho_\gamma \left[ 1 + N_{\text{eff}} \frac{7}{8} \left( \frac{4}{11} \right)^{\frac{4}{3}} \right]
$$

(5–22)

where $\rho_{\text{rad}}$ is the total energy density in radiation and $\rho_\gamma$ is the energy density in photons only. The standard cosmological model with ordinary cold dark matter predicts $N_{\text{eff}} = 3.046$, slightly larger than 3 because the three neutrinos heat up a little during $e^+e^-$ annihilation. Taking account of the fact that not only is there an extra species of radiation (thermally excited axions) but also the contribution of the three ordinary neutrinos is boosted because the photons have been cooled relative to them, the proposed scenario predicts

$$
\rho_{\text{rad}} = \rho_\gamma + \rho_a + \rho_\nu
= \rho_\gamma \left[ 1 + \frac{1}{2} + 3.046 \frac{7}{8} \left( \frac{4}{11} \right)^{\frac{4}{3}} \left( \frac{3}{2} \right) \right],
$$

(5–23)

which yields $N_{\text{eff}} = 6.77$.

At present, the measured values are smaller than this prediction. The WMAP collaboration found $N_{\text{eff}} = 4.34^{+0.86}_{-0.88}$ (68% CL) based on their 7 year data combined with independent data on large scale structure and the Hubble constant. An analysis [79] using the Sloan Digital Sky Survey (SDSS) data release 7 halo power spectrum found $N_{\text{eff}} = 4.8 \pm 2.0$ (95% CL). The Atacama Cosmology Telescope (ACT) collaboration finds [80] $N_{\text{eff}} = 5.3 \pm 1.3$ (68% CL) using only their CMB anisotropy data and $N_{\text{eff}} = 4.56 \pm 0.75$ (68% CL) when combining that data with large scale structure data. The tendency for the measured values to be larger than 3.046 has been taken sufficiently seriously to prompt proposals for new physics involving extra neutrino species or a neutrino asymmetry [81].
CHAPTER 6
CONCLUSIONS

Although the three major candidates for CDM, axions, WIMPs, and sterile neutrinos, were thought until recently to be indistinguishable by purely astronomical observations, axions are very different from the other two in terms of statistical mechanics properties. The axions are a highly degenerate Bose fluid while the other two are not degenerate. Through gravitational self-interactions, axions thermalize when the photon temperature drops below $500\text{eV}$. They then form a Bose-Einstein condensate, i.e. almost all axions go to the lowest energy state. We find that if that state is time independent, axions behave as the other cold dark matter candidates on all scales of observational interest. However observational differences occur when the axions rethermalize and the axion state tracks the lowest energy state. We find that cold axions rethermalize when they fall into a galactic halo. As a result the cold axions acquire a state of net overall rotation. In contrast, ordinary cold dark matter falls in with an irrotational velocity field. The inner caustics of the galactic halo are different in the two cases. The occurrence of caustic rings of dark matter in galactic halos is inconsistent with ordinary cold dark matter, but consistent with axion BEC. In addition, cold axions may reach thermal contact with photons and baryons at the time of equality between matter and radiation. Thermal contact between cold axions, photons and baryons changes cosmological parameters, specifically the baryon to photon ratio at the time of primordial nucleosynthesis and the effective number of neutrino species at decoupling. The change in the baryon to photon ratio alleviates the famous "lithium problem". Future cosmological observations, such as by the Planck mission, may provide precise data on the effective number of neutrinos.
APPENDIX A
DETECTION OF AXION-LIKE PARTICLES BY INTERFEROMETRY

A.1 Introduction

Although astrophysical observations and cosmological considerations provide useful constraints on the axion parameter space, whether axions really exist can only be settled if they are actually detected in the laboratory, and as of today the hypothetical particle remains elusive. Initially, the prospect of detecting such weakly interacting particles was deemed unlikely, since a very large $f_a$ implies that axions couple very weakly to ordinary matter. However, it was pointed out that we may catch glimpses of the elusive particle by exploiting its coupling to two photons, which is given in the Lagrangian by [30]

$$L_{a\gamma\gamma} = \frac{g_{a\gamma\gamma}}{4} a F \tilde{F}.$$  \hspace{1cm} (A–1)

Through this coupling, the axion and photon can mix with each other in a background magnetic field. It is this principle that underlies all existing axion detection experiments. The ADMX experiment, for example, is a realization of the axion haloscope, in which axions in the halo are induced to convert in a cavity to microwave photons that are then picked up by an antenna. The CERN Axion Solar Telescope and the Tokyo Helioscope, on the other hand, are a realization of the helioscope and aim to detect axions originating from the Sun, by converting them into X-rays in a strong magnetic field. The photon-axion mixing can also manifest itself in the birefringence and dichroism in the vacuum, resulting in rotation and elliptization of the polarization of light in the presence of a magnetic field. Such signal is actively being sought, as in the PVLAS experiment.

Another type of experiment that makes use of this mixing is photon-regeneration (or “light shining through a wall”) [99], in which a small fraction of the photons in a laser beam traversing a region permeated by a magnetic field is converted to axions. Because of their weak coupling to ordinary matter, the axions travel unimpeded through a wall,
on the other side of which is an identical arrangement of magnets. There some of the
axions are induced to convert back to photons, which can be detected. The primary
advantage of photon-regeneration experiments is their greater control over experimental
conditions. Since the laser beam is prepared in the laboratory, one does not have to
rely on extraterrestrial axion sources. The major drawback is that the signal is very
weak ($\propto g_{a\gamma\gamma}^4$), since two stages of conversion are required. At the moment, photon
regeneration experiments do not have sufficient sensitivity to detect the QCD axion,
although they are in principle capable of detecting other particles that couple more
strongly to the photon in an analogous manner. Hence, their primary objective is to
detect “axion-like particles” (ALPs), rather than axions.

ALPs are predicted to exist generically in string theory [88]. While pseudoscalar
ALPs couple to photons as axions do, scalar ALPs couple to photons via a $aF_{\mu\nu}F^{\mu\nu}$
term in the Lagrangian, so they can be produced by photons whose polarization
is perpendicular to the background magnetic field. In general, there is no a priori
relationship between the mass and couplings of ALPs; hence their parameter space is a
lot less constrained compared to axions.

We propose a new experimental method based on interferometry to detect ALPs.
A laser beam is split into two beams of equal intensity. One of them acts as a reference
beam, while the other traverses a region permeated by a magnetic field which induces
conversion into ALPs, just as in the first half of photon-regeneration experiments.
However, instead of having a second stage behind a wall where ALPs are converted
back to photons, the beam is recombined with the reference beam. If photon-ALPs
conversion has occurred, the beam emerging from the conversion region would have
a slightly reduced amplitude and a phase shift relative to the reference beam. This
leads to a change in the combined intensity, which can then be measured by a detector.
Because only one stage of conversion is needed, the signal intensity is proportional to
only $g_{a\gamma\gamma}^2$, instead of $g_{a\gamma\gamma}^4$ for the photon-regeneration experiment. This, however, does
not straightforwardly improve sensitivity to $g_{\gamma\gamma}$ due to the presence of shot noise in any light source; we will expound on this later.

In order to avoid having the signal being overwhelmed by the background, the two beams are arranged to traverse paths of different lengths, such that they would be out of phase by $\pi$ at the detector when the magnetic field is switched off. Thus, without any conversion the two beams would interfere destructively at the detector. The detection of light would signal the occurrence of ALPs production. Unfortunately, at the dark fringe the signal is reduced to a second-order effect ($O(g_{\gamma\gamma}^4)$), so it is necessary to modulate the amplitude (or frequency) of the laser by using a Pockels cell. The presence of the two sidebands in addition to the carrier gives rise to a component in the power output that is of $O(g_{\gamma\gamma}^2)$, which can then be isolated and detected by the use of a mixer.

Any coherent light source is affected by shot noise. For an incoming laser beam of $N$ photons, the Heisenberg uncertainty principle implies an fluctuation of $\sqrt{N}$ in the photon number. This reduces our ability to place a limit on the ALPs-photon coupling: $g_{\gamma\gamma} \sim B^{-1}L^{-1}N^{-1/4}$, which is the same as that in ordinary photon regeneration (where $B$, $L$, and $N$ are the magnetic field, length of conversion region, and number of photons respectively). Fortunately, our design admits a straightforward implementation of light squeezing, which can reduce shot noise by an order of magnitude with current techniques.

Furthermore, by employing optical delay lines, we can enhance the signal by a factor of $n$, where $n$ is the number of times a laser beam is folded. So we can improve our constraint on $g_{\gamma\gamma}$ by $n^{1/2} \sim 10^{2.5}$. By comparison, the use of optical delay line in photon-regeneration results in a much weaker improvement of order $n^{1/4}$.

We also point out that in recent years there has been a proliferation of hypothesized particles, many of which couple to two photons as ALPs do, so they could also potentially be discovered in our proposed experiment. Some examples include chameleons, massive hidden photons, and light minicharged particles [91–95]. In
Figure A-1. Schematic diagram of our proposed experiment. A laser beam, whose amplitude is modulated by a Pockels cell, is split into two beams of equal intensity ($B_1$ and $B_2$). The beam $B_2$ (vertical) traverses a region permeated by a magnetic field, where photons convert to axions (and other particles with a two-photon vertex). It is then recombined at the detector with the beam $B_1$ (horizontal), which acts as a reference. The two arms are different in length, so that the two beams are out of phase by $\pi$ in the absence of a magnetic field. A change in intensity registered by the detector would signal the occurrence of a conversion. To extract the component of the overall signal that is proportional to $g_{a\gamma\gamma}^2$, we mix the output with the oscillator voltage that drives the Pockels cell.

In particular, using results in [91], it is straightforward to generalize our analysis to the detection of minicharged particles.

**A.2 Design of experiment**

Photon-axion mixing in a magnetic field is based on the $aF\tilde{F}$ coupling, where one of the photon legs is a virtual photon in the magnetic field. If the polarization of the photon is parallel to the magnetic field, the probability of conversion $\eta$ can be obtained from the
cross section of this process:

$$\eta_{\gamma \to a} = \frac{1}{4v_a} (g_{a\gamma\gamma}BL)^2 \left( \frac{2}{qL} \sin \left( \frac{qL}{2} \right) \right)^2, \quad (A-2)$$

where $v_a$ is the velocity of the axion, $B$ the magnetic field, $L$ the length of the conversion region, and $q$ the momentum transfer. Since $m_a \ll \omega_\gamma$, the frequency of the laser beam photons, $v_a \sim 1$. For $L \sim 10\,\text{m}$, this also implies that $qL \sim 10^{-5} \ll 1$, so (A-2) can be approximated by

$$\eta_{\gamma \to a} \approx \frac{1}{4} (g_{a\gamma\gamma}BL)^2. \quad (A-3)$$

If we use $B \sim 10\,\text{T}$, $L \sim 10\,\text{m}$, and $f_a^{-1} \sim 10^{-12}\,\text{GeV}^{-1}$, the probability of photon-axion conversion is of $O(10^{-26})$ to $O(10^{-25})$. After the conversion, the amplitude $A$ of the photon is reduced to $A - \delta A$, where

$$\delta A_{\gamma \to a} = \frac{A\eta_{\gamma \to a}}{2} \approx \frac{g_{a\gamma\gamma}^2 B^2 L^2 A}{8}. \quad (A-4)$$

We note that the discussion here is applicable to pseudoscalar ALPs, since they couple to the photon in exactly the same way. If the photon polarization is instead perpendicular to the magnetic field, the analysis is also valid for scalar ALPs, as they couple to photons via $a\mathcal{F}F \sim \vec{B} \cdot \vec{B}$ instead.

When a photon enters a region permeated by a magnetic field, the dispersion relation for the component orthogonal with respect to the magnetic field remains $\omega^2 = k^2$. However, if axion production occurs, that of the parallel component is modified:

$$\omega^2 = k^2 + \frac{1}{2} \left( m_a^2 + g_{a\gamma\gamma}^2 B^2 \right) \pm \sqrt{\left( m_a^2 + g_{a\gamma\gamma}^2 B^2 \right)^2 + 4 g_{a\gamma\gamma}^2 k^2 B^2}. \quad (A-5)$$

For $B \sim 10\,\text{T}$ and $g_{a\gamma\gamma} \sim 10^{-12}\,\text{GeV}^{-1}$, the value of $g_{a\gamma\gamma}^2 B^2$ is much less than $m_a^2$. Under these assumptions, the additional phase acquired $\delta \theta$ is then approximately

$$\delta \theta_{\gamma \to a} \approx \frac{g_{a\gamma\gamma}^2 B^2 m_a^2 L^3}{48k}. \quad (A-6)$$
The effect of the phase shift is negligible in comparison with $\delta A/A$.

In our proposed experiment, a laser beam first enters a Pockels cell (with a polarizer behind) to modulate its amplitude (the purpose of the modulation will be explained below). Subsequently, it is divided by a beamsplitter into two beams (which we label $B_1$ and $B_2$ in Figure A-1) with equal intensity. $B_2$ is essentially the laser beam used in the first half of the “shining-light-through-the-wall” experiment: it passes through a region permeated by a constant magnetic field, where a small fraction of the photons are converted into axions which carry energy away from the beam. For simplicity, we will consider here that the carrier of the modulated beam (both $B_1$ and $B_2$) is linearly polarized in the direction of the magnetic field, so our analysis in the previous section applies (For the detection of scalar ALPs, the polarization should be perpendicular to the magnetic field instead). The two beams are then recombined at the detector, and in the presence of a conversion, the slight amplitude reduction and phase shift would lead to interference, which can be detected.

The length of the path traversed by beam $B_1$ is by design slightly different from that by $B_2$, so that at the detector the two beams would be out of phase by $\pi$ if the magnetic field has been absent. Operationally, this can be achieved by adjusting one of the path lengths until destructive interference is observed at the detector when the magnetic field is turned off. Hence, in the absence of the sidebands, the two beams would interfere destructively at the detector. The purpose for this arrangement is to reduce the background, thereby enhancing the signal-to-noise ratio and minimizing shot noise.

Let the path lengths of the two arms be $L_x$ and $L_y$ (corresponding to beams $B_1$ and $B_2$), and the state of the laser after passing through the Pockels cell be described by

$$\vec{E}_{\text{in}} = \vec{E}_0 (1 + \beta \sin \omega_m t) e^{i\omega t},$$  \hspace{1cm} (A–7)
where $\beta$ is a constant, $\vec{E}_0$ the initial electric field at $t = 0$, and $\omega$ is the frequency of the laser. The amplitude is modulated at a frequency $\omega_m$. This can be recast as

$$
\vec{E}_{in} = \vec{E}_0 \left( e^{i\omega t} + \frac{\beta}{2i} e^{i(\omega+\omega_m)t} - \frac{\beta}{2i} e^{i(\omega-\omega_m)t} \right),
$$

(A–8)

where the first term is referred to as the “carrier”, and the latter two as “sidebands”.

For simplicity, we ignore the contribution of the additional phase in the present analysis, since it is negligible in comparison to that of $\delta A$. In this case, the state of the carrier after recombination at the detector is given by

$$
\vec{E}_{carrier} = -\frac{\vec{E}_0}{2} e^{i(\omega t + 2kL)} \times [2i \sin k\Delta L - \frac{\delta A}{A} e^{-iK\Delta L}],
$$

(A–9)

where $k = \omega/c$ is the wavenumber of the laser photons, $A = |\vec{E}_0|, \Delta L = L_x - L_y$ is the length difference between the two arms, and $L = (L_x + L_y)/2$ is the average. As mentioned, we will choose $k\Delta L = \pi$, so that the detector operates at a dark fringe, in order to eliminate the background signal. This leads to

$$
\vec{E}_{carrier} = e^{i(\omega t + 2kL)} \frac{\delta A}{2A} \vec{E}_0.
$$

(A–10)

Note that without the aid of the sidebands, this would be the entire signal. While the background is eliminated, the intensity ($\sim \vec{E}^2$) is of $\mathcal{O}(g_{3\gamma\gamma}^A)$ (for axions). This loss in sensitivity, as we will see, can be recovered by using the sidebands.

Meanwhile, the sidebands (second and third terms of (A–8)) are described by

$$
\vec{E}_\pm = \vec{E}_0 \beta e^{i(\omega t + 2kL)} e^{\pm i(\omega_m t + 2\omega_m L/c)} \times \left[ \sin \frac{\omega_m \Delta L}{c} \mp i \frac{\delta A}{2A} e^{\mp i\omega_m \Delta L/c} \right],
$$

(A–11)

where the subscripts + and − denote respectively the sideband components of frequency $\omega + \omega_m$ and $\omega - \omega_m$. 

99
If we set \( \omega_m \approx \pi c / 2\Delta L \), the total electric field at the detector is obtained by adding that of the carrier and sidebands:

\[
\vec{E} = \vec{E}_0 e^{i(\omega t + 2\pi \Delta L)} \left( \frac{\delta A}{2A} \right) + \beta \left( 2 - \frac{\delta A}{A} \right) \cos \left( \omega_m t + \frac{2\omega_m L}{c} \right). \tag{A-12}
\]

Note that this particular value of \( \omega_m \) is chosen to maximize the signal. Since \( \omega_m \to n\omega_m \) and \( k\Delta L \to n\pi \) (for \( n \) an odd integer) are equally valid choices, the experimenter has much freedom in choosing a suitable value for \( \omega_m \) that is experimentally feasible.

Hence, the power \( P \) that falls on the detector is

\[
P = P_n \left\{ \frac{(\frac{\delta A}{A})^2}{4} + \frac{\beta^2 (4 - 4(\frac{\delta A}{A}) + (\frac{\delta A}{A})^2)}{2} \right. \\
\left. + \frac{\delta A}{A} \beta (2 - \frac{\delta A}{A}) \cos \left[ \omega_m \left( t + \frac{2L}{c} \right) \right] \right. \\
\left. + \frac{\beta^2 (4 - 4\frac{\delta A}{A} + \frac{\delta A^2}{A^2})}{2} \cos \left[ 2\omega_m \left( t + \frac{2L}{c} \right) \right] \right\}. \tag{A-13}
\]

Thus the power has a dc component (first line), and two ac components with frequencies \( \omega_m \) and \( 2\omega_m \). If we multiply this with the oscillator voltage that drives the Pockels cell (plus an appropriate phase shift) via a mixer, we can extract the component of frequency \( \omega_m \). Neglecting the second-order contributions, the time-averaged output power of the mixer is given by

\[
P_{\text{out}} = \frac{1}{T} \int_T^{t+T} 2P_n \beta G \left( \frac{\delta A}{A} \right) \cos^2 (\omega_m t) \tag{A-14}
\]

\[
= \frac{P_n \beta G \delta A}{A} \tag{A-15}
\]

where \( G \) is the gain of the detector and \( T \) is taken to be sufficiently long to ensure that the time-averaging is accurate. Hence, the output signal is proportional to \( g_{\alpha\gamma\gamma}^2 \) for axions and \( G \) for gravitons.

In this analysis we choose to modulate the amplitude, rather than the phase, of the photons because the reduction in amplitude has comparatively a much larger effect. In
principle, we could instead modulate the phase, in which case the change in intensity registered by the detector would be primarily a consequence of the phase shift instead of the amplitude reduction. The corresponding analysis is highly analogous and will not be repeated here. The major difference is that the coefficients for the sidebands in (A–8), $\beta/2i$, are replaced approximately by $J_1(\beta)$, the first-order Bessel function of the first kind (higher harmonics now are also present, but are negligible). Since $J_1(\beta)$ is real, our earlier analysis would work if $\delta A/A$ is replaced by $i\delta \theta$, which is purely imaginary. This can be implemented by manipulating polarizers adjacent to the Pockels cell. Thus by switching between phase and amplitude modulation, we can infer information on both the amplitude reduction and phase shift. This is one conceivable way of identifying the particles that the photons have converted into.

**A.3 Discussion and conclusions**

In this appendix, we proposed a new method of ALP detection based on interferometry. A laser source is split into two beams, where one is exposed to a magnetic field permeating a confined region, within which photon-axion conversion occurs. This results in a phase shift and reduction in amplitude, which can be made manifest if the beam is then recombined and made to interfere with the other, which acts as a reference. Because only one stage of conversion is needed, the signal goes as $g_{\gamma\gamma}^2$, which is an improvement over that of existing photon-regeneration experiments. The key to the improvement is the realization that it is not necessary to convert the axions back to photons for detection; interference with a reference beam can reveal just as much.

However, what matters in practice is the signal to background ratio. In order to avoid the signal being overwhelmed by the background, it is necessary to have the detector operate at a dark fringe. Unfortunately this also reduces the signal to a second-order effect ($O(g_{\gamma\gamma}^4)$). This reduction can be nullified by modulating the photon amplitude, and mixing the output signal with the oscillator voltage that drives the Pockels cell.
Despite the improvement in signal size, the use of interferometers is inevitably accompanied by the presence of shot noise, which is a manifestation of the granular nature of the coherent state of photons in the laser beam. This limits the resolution of the interferometer therefore reducing the sensitivity to $g_{a\gamma\gamma}$ in our set up.

For a laser beam consisting of $N$ incoming photons, we expect the shot noise in our setup to have a magnitude of $\sqrt{N}$ due to Poisson statistics. The signal-to-noise ratio is thus reduced to $(g_{a\gamma\gamma}BL)^2 N/\sqrt{N}$. In the case of a non-detection, this allows us to constrain the axion-photon coupling to $g_{a\gamma\gamma,max} < (BL)^{-1}N^{-1/4}$, which is what can be achieved by conventional photon-regeneration experiments. (In their case, the signal is much smaller, of $O(g_{a\gamma\gamma}^4N)$, so dark count rate can be a problem.)

Our setup admits a straightforward implementation of squeezed light using standard optical techniques, which can help reduce shot noise. The idea is that we could reduce the uncertainty in an operator by enhancing that of its conjugate operator, so that the Heisenberg uncertainty principle is still satisfied. A $10\text{dB}$ suppression of shot noise can result in a $10^{1/2}$ improvement of the constraint to $g_{a\gamma\gamma}$. To further boost the sensitivity, we can incorporate in our setup optical delay lines to enhance the signal by a factor of $n$, where $n$ is the number of times the laser beam is folded. The resultant improvement in our ability to constrain $g_{a\gamma\gamma}$ is of order $n^{1/2} \sim 10^{2.5}$ v.s. $n^{1/4} \sim 10^{1.25}$ in photon regeneration experiment. Combined, the use of squeezed light and optical delay lines results in a gain in the sensitivity to $g_{a\gamma\gamma}$ of $10^2$ over simple photon regeneration experiment.

If we use $n \sim 10^3$, $B \sim 10\text{T}$, $L \sim 10\text{m}$ with a $10\text{W}$ ($\lambda = 1\mu\text{m}$) laser, after 240 hours running, the experiment can exclude ALPs with $g_{a\gamma\gamma} > 2.8 \times 10^{-11}\text{GeV}^{-1}$ to $5\sigma$ significance. If one also employs squeezed-light laser which improves signal-to-noise ratio by $10\text{dB}$ with similar setup, the exclusion limit can reach $g_{a\gamma\gamma} \sim 10^{-12}\text{GeV}^{-1}$.

Finally we point out that while we have as our principal aim the detection of ALPs, our design is theoretically applicable to any particle with a two photon vertex, so that
mixing in the presence of an external magnetic field is permitted. Given the possibility that more than one such particle exists, it is important to identify what the photons have converted into. We suggest two methods that can help shed light on this issue. First, we could repeat the experiment by modulating the phase instead of the amplitude of the laser, as this would reveal information about the phase shift as well. Secondly, scalar and pseudoscalar ALPs can be distinguished by modifying the polarization of the laser. Conversion can only occur if the polarization is parallel (perpendicular) to the external magnetic field for pseudoscalar (scalar) ALPs.

Even with the incorporation of squeezed light and an optical delay line, the sensitivity of our experiment still falls short for the detection of the QCD axion and the graviton. This is expected in light of the feebleness of their couplings to photons. It is hoped that future improvements in the technology of light squeezing and the advent of more powerful lasers might some day help bridge the gap in sensitivity required. If achievable, our experiment might serve as an excellent complement to existing experiments, such as ADMX and LIGO.
B.1 Introduction

We propose a new observational probe of axions or ALPs, based on the fact that charged particles propagating in a time-dependent axion field emit photons. The axion field can be viewed as a source of energy (due to its time dependence), but not three-momentum (due to its homogeneity). Consequently, some processes that are kinematically forbidden (energy and momentum conservation cannot be satisfied simultaneously) before now become possible. Admittedly, because the axion’s coupling to ordinary matter is very small, the rate of photon emission is extremely tiny. In fact, with current technologies, performing a laboratory experiment to observe the emission is definitely not feasible. A rough estimate shows that an electron accelerator with a reasonable length and electron flux will have to be in operation for more than the age of the universe before the emission of a single photon. Fortunately, this phenomenon occurs naturally in cosmology: by cosmic rays (primarily protons) propagating in a time-dependent axion field. As we will demonstrate, the abundance of axions (or ALPs) and cosmic rays in our galaxy might compensate for the smallness of the coupling. With the aid of a detector with a collecting area of $10^{10}$ cm$^2$ (for example, the Square Kilometre Array currently under construction), this can give rise to a weak but detectable signal for ALPS (but not for the QCD axion).

Because galactic cosmic rays also generate diffuse galactic radiation, it might appear difficult to disentangle our signal from the background, which is dominant. Fortunately, it turns out that the energy spectrum of the photons has a well-defined peak, which is located approximately at the mass of the axion or ALP. This is expected to be of order $\mu$eV, while diffuse galactic radiation tends to be much more energetic (GeV). The existence of such a peak can be understood since kinematics dictates that cosmic ray protons in the low energy end of their spectrum can only produce photons
whose frequency is approximately the mass of the axion or ALP. Since both the cross section and cosmic-ray energy spectrum decrease with increasing energy, photons whose frequency lies in the vicinity of the axion (or ALP) mass are most abundantly produced.

The propagation of charged particles in a spatially homogeneous, but time-dependent, pseudoscalar background was investigated in [105], in which the authors make the assumption that the time-varying background be treated as a constant in the Lagrangian (this is essentially the Lorentz-violating Chern-Simons term considered by Carroll et. al. in [106]). This additional term has the effect of modifying the dispersion relation of the photons. As a consequence, the process of photon emission is described at leading order by a single Feynman diagram with one vertex. According to [105], photon emission by charged particles is then possible only if the photon is massive, and the emission angle is small. In contrast, our calculation incorporates the time dependence of the axion field, which leads to three Feynman diagrams with two vertices each. Using this method, we find instead that emission is possible at all angles and for massless photons.

B.2 Theoretical analysis of photon emission by protons in a pseudoscalar field

The Lagrangian that describes the dynamics of protons ($\psi$) propagating in an axion (or ALP) field ($\phi$) is given by

$$
\mathcal{L} = -ig_{ap}\bar{\psi}\gamma^5\psi + \frac{\phi}{4} g_{a\gamma\gamma} F_{\mu\nu} \tilde{F}^{\mu\nu} + \mathcal{L}_{QED},
$$

(B–1)

where $\mathcal{L}_{QED}$ is the usual QED Lagrangian that describes protons and photons. The couplings $g_{ap} = c_{ap} m_p / f_a$ and $g_{a\gamma\gamma} = c_{a\gamma\gamma} \alpha / (2\pi f_a)$ are respectively the axion-proton and axion-photon coupling, where $c_{ap}$ and $c_{a\gamma\gamma}$ are dimensional-less model-dependent parameters, typically of order unity [107]. In this paper, we assume that $c_{ap} = c_{a\gamma\gamma} = 1$. The parameter $f_a$ essentially measures the strength of the ALP’s coupling to the photon and proton. For the QCD axion, $f_a$ is known as the axion decay constant, and it is
constrained to $10^9 < f_a < 10^{12}$ GeV by particle and nuclear physics experiments, stellar evolution, and cosmology. For the QCD axion, the mass is given by

$$m_a \approx 6 \times 10^{-6} \text{eV} \frac{10^{12} \text{GeV}}{f_a}.$$  (B–2)

For ALPs, no such relation holds, and so both $f_a$ and $m_a$ are free parameters of the theory.

The axion field is expected to oscillate, given by $\phi(t) = \phi_0 \cos(m_a t)$ [108]. Because of its time variation and spatial homogeneity, it can be viewed as a source of energy, but not three-momentum. This is what makes the ordinarily kinematically forbidden process of photon emission by a charged particle possible.

### B.2.1 Matrix elements

There are three processes that contribute to the emission of photons by cosmic ray protons. The four-momenta of the incoming proton and the axion are denoted by $p^\mu$ and $a^\mu$, while those of the outgoing proton and emitted photon are $q^\mu$ and $l^\mu$ (i.e. $p(p^\mu) + a(a^\mu) \to p(q^\mu) + \gamma(l^\mu)$). In the following, we adopt the (+−−−) metric convention.

The matrix element for process $I$ in which a cosmic ray proton emits a virtual photon, which then interacts with the axion field to emit a real photon, is

$$i\mathcal{M}_1 = \frac{ieg_{\gamma\gamma}}{(p-q)^2} \bar{u}(q)\gamma^\mu u(p)\epsilon_\mu^{\alpha\beta\rho} (p-q)_\alpha l_\beta \epsilon_\rho^*(l).$$  (B–3)
The matrix element for process II in which a proton, subsequent to emitting a real photon, interacts with the axion field, is

\[ iM_2 = \frac{i e g_{ak}}{2(l \cdot p)} \epsilon^*_\mu(l) \left[ \bar{u}(q) \gamma^5(2p^\mu - \gamma^\nu l_\nu \gamma^\mu) u(p) \right]. \tag{B–4} \]

The matrix element for process III in which a proton interacts first with the axion background, then emits a real photon, is given by

\[ iM_3 = -\frac{i e g_{ak}}{2(l \cdot q)} \epsilon^*_\mu(l) \left[ \bar{u}(q)(2q^\mu + \gamma^\nu l_\nu \gamma^5) u(p) \right]. \tag{B–5} \]

To calculate the differential cross section, we need to first compute \(|M_1|^2, |M_2|^2, |M_3|^2, M_1M_2^*, M_1M_3^*, M_2M_3^*\). Squaring \(M_1\), averaging over initial proton spins, and summing over final photon polarizations, we have

\[ \frac{1}{2} \sum_{\text{spins}} |M_1|^2 = -\frac{4 g_{a\gamma\gamma} e^2}{(p-q)^4} \left[ 4m_p^2 - 4(p \cdot q)(p \cdot l) + 4m_p^2(l \cdot a)^2 - 2(l \cdot a)^2(p \cdot q) \right]. \tag{B–6} \]

Meanwhile, squaring \(M_2\) and \(M_3\) yields

\[ \frac{1}{2} \sum_{\text{spins}} |M_2|^2 = \frac{4 e^2 g_{ak}}{(l \cdot p)^2} \left[ (p \cdot l)(q \cdot l) + m_p^2(a \cdot l) - m_p^2(p \cdot q) + m_0^4 \right], \tag{B–7} \]

and

\[ \frac{1}{2} \sum_{\text{spins}} |M_3|^2 = \frac{4 e^2 g_{ak}}{(l \cdot q)^2} \left[ (p \cdot l)(q \cdot l) + m_p^2(a \cdot l) - m_p^2(p \cdot q) + m_0^4 \right]. \tag{B–8} \]

The cross terms can likewise be straightforwardly computed:

\[ \frac{1}{2} \sum_{\text{spins}} M_1M_2^* = \frac{2im_p e^2 g_{a\gamma\gamma} g_{ak}(a \cdot l)^2}{(l \cdot p)(p \cdot q)^2} \tag{B–9} \]

\[ \frac{1}{2} \sum_{\text{spins}} M_1M_3^* = \frac{2im_p e^2 g_{a\gamma\gamma} g_{ak}(a \cdot l)^2}{(l \cdot q)(p \cdot q)^2} \tag{B–10} \]

\[ \frac{1}{2} \sum_{\text{spins}} M_2M_3^* = -\frac{2e^2 g_{ak}^2}{(l \cdot p)(l \cdot q)} \left[ (l \cdot p)(l \cdot q) + (p \cdot q)(l \cdot a) - (p \cdot q)^2 + m_p^2(p \cdot q) \right]. \tag{B–11} \]
B.2.2 Differential cross section

For simplicity, we evaluate the differential cross section in the rest frame of the axion field. In this case the four-momenta are given by $p^\mu = (E_p, 0, 0, p)$, $a^\mu = (m_a, \vec{0})$, $q^\mu = (E_q, q \sin \theta, 0, q \cos \theta)$, $l^\mu = (\omega, \vec{\omega})$, where $E_p = \sqrt{p^2 + m_a^2}$, $E_q = \sqrt{q^2 + m_a^2}$. Without loss of generality we align the z-axis with the direction of propagation of the initial proton, and restrict the scattering to the x-z plane. $\theta$ thus denotes the angle between the direction of the emitted photon and the z-axis. The photon’s frequency $\omega$ is

$$\omega = \frac{m_a^2 + 2m_a E_p}{2E_p + 2m_a - 2|p| \cos \theta}. \quad (B-12)$$

In this frame, the phase space for the final state particles is given by

$$\int \frac{d^3 \vec{q}}{(2\pi)^3 2E_q} \frac{d^3 \vec{\omega}}{(2\pi)^3 2\omega} (2\pi)^4 \delta^4(p + a - q - l), \quad (B-13)$$

which yields the differential cross section

$$\frac{d\sigma}{d \cos(\theta)} = \frac{m_a + 2E_p}{16\pi(2E_p + 2m_a - 2|p| \cos \theta)^2 \sqrt{E_p^2 - m_a^2}} \left[ \frac{1}{2} \sum \left| M_1 + M_2 + M_3 \right|^2 \right]. \quad (B-14)$$

B.2.3 Emission rate of the photons

Our galaxy is teeming with cosmic rays, whose primary constituent is protons, to which we will restrict our attention in this paper (hence $E = E_p$ below). Our calculation of the photon emission rate is thus a conservative estimate, as other charged constituents (e.g. electrons) would also contribute to the process. As cosmic ray protons propagate in this background time-dependent axion field, they undergo photon emission via processes described in the previous section.

To estimate the photon flux on Earth, we make the assumption that cosmic ray protons are homogeneous and isotropic within our galaxy. This is predicated on the observation that cosmic ray protons scatter off interstellar medium and traverse random trajectories within the Galaxy for an average of $10^7$ years in the galaxy [105]. For our
calculation, we adopt the following energy spectrum for cosmic ray protons \cite{110, 111}:

\[
\frac{d\mathcal{F}}{dE d\Omega} = \left( \frac{3.06 \text{ cm}^2 \text{ s sr GeV}}{\text{cm}^2 \text{ s sr GeV}} \right) \left( \frac{E}{\text{GeV}} \right)^{-2.70},
\]

which we assume to hold for \( E > 50 \) GeV. Since the flux is known with less certainty for low-energy protons, we impose a cutoff in our calculation and disregard all protons with an energy below 50 GeV as a conservative measure. We will also neglect the contribution of extragalactic cosmic rays, which should be subdominant as compared to the galactic ones.

Consider now a photon detector on Earth, with a field of view \( \delta \Omega_d \). With our assumptions on the distribution of cosmic rays, the detector can pick up photons originating from cosmic rays filling a region from the Earth to the edge of the Galaxy, whose angular boundary is determined by \( \delta \Omega_d \). Let \( \mathcal{V} \) denote this region, and \( d\mathcal{V}_r \) be a differential volume element in \( \mathcal{V} \) at a distance \( r \) from the detector on Earth. At each point in \( \mathcal{V} \), there are cosmic rays propagating in all possible directions. Locally, we define a spherical coordinate system (with the usual coordinates \( \theta \) and \( \phi \)), where \( \theta \), as we defined earlier, denotes the angle between the velocity vector of the cosmic ray and the line connecting \( d\mathcal{V}_r \) to the detector on Earth. The other variable \( \phi \) is the usual azimuthal angle confined to the plane perpendicular to the direction of propagation of the cosmic ray.

The number of photons emitted by the cosmic rays that fill up \( d\mathcal{V}_r \) over an interval \( dt \) is given by

\[
dN_\gamma = n_s n_p d\mathcal{V}_r d\sigma d\Omega d\nu dt.
\]

where \( n_s \) and \( n_p \) are the number density of axions and protons, \( \delta \nu \) the velocity of the cosmic rays, and \( d\sigma \) the differential cross section. From this, we can compute the flux of photons per unit time at the detector:

\[
\frac{dN_\gamma}{dt dA} = n_s \left( \frac{dn_p}{dE} dE \right) \frac{d\sigma}{dA d\Omega d\nu} d\mathcal{V}_r.
\]
where \( dA = r^2 \, d\Omega = r^2 \, d\cos \theta \, d\phi \) is the differential area at the detector. Using the fact that

\[
\frac{dn_p}{dEd\Omega} = \frac{1}{\delta \nu} \frac{d\mathcal{F}}{dEd\Omega},
\]  

(B–18)

the photon flux simplifies to

\[
\frac{dN_{\gamma}}{dtdA} = \left( \frac{n_a \, d\mathcal{F}_p}{r^2 \, dEd\Omega} \frac{d\sigma}{d\cos \theta} \right) d\mathcal{V}_r \, dEd \cos \theta.
\]  

(B–19)

Integrating over the volume that the detector can see, the proton energy, and \( \cos \theta \), and using (B–15), we obtain the photon flux at the detector,

\[
\frac{dN_{\gamma}}{dtdA} = n_a \left( \int_\gamma^1 \frac{1}{r^2} d\mathcal{V}_r \right) \int dE \left[ \frac{3.06}{\text{cm}^2 \, \text{s} \, \text{sr} \, \text{GeV}} \left( \frac{E}{\text{GeV}} \right)^{-2.70} \left( \int_{-1}^{1} \frac{d\sigma(E, \cos \theta)}{d\cos \theta} d\cos \theta \right) \right],
\]  

(B–20)

which can be computed numerically.

B.2.4 Energy spectrum of photons

Due to the uncertainties inherent in the energy spectrum of low-energy cosmic rays within the Galaxy, it is not possible to determine the precise spectral shape of the emitted photons. Nonetheless, the spectrum possesses a robust feature: it has a peak located at approximately the mass of the axion or ALP (more accurately, at \( \omega_c \equiv (m_a^2 + 2m_a m_p)/(2m_p + 2m_a) \), obtained by setting \( E_p = m_p \) in (B–12)).

The existence of such a peak can be understood as follows. From (B–12), we observe that for a given \( E \), the emitted photon’s frequency is confined to the range \( \omega_- \leq \omega \leq \omega_+ \), where

\[
\omega_{\pm} = \frac{m_a^2 + 2m_a E}{2E + 2m_a \mp 2\sqrt{E^2 - m_p^2}}.
\]  

(B–21)

The frequencies \( \omega_+ \) and \( \omega_- \) correspond to scattering at \( \theta = 0 \) and \( \pi \) respectively (see figure B-4). Hence, photons of frequency \( \omega \) can only be produced by cosmic ray protons with an energy larger than

\[
E = \frac{1}{2m_a^2 - 4m_a \omega} \left( -m_a^3 + 3m_a^2 \omega - 2m_a \omega^2 + \sqrt{-4m_p^2 m_a^2 \omega^2 + m_a^4 \omega^2 + 8m_p^2 m_a^2 \omega^3 - 4m_a^2 \omega^3 + 4m_p^2 \omega^4} \right).
\]  

(B–22)
Figure B-4. Photon frequency $\omega$ versus the proton energy $E$, for an axion mass of $10^{-6}$ eV, according to (B–21). To the right of the curve is the region in which photon generation by cosmic ray protons is allowed. The turnaround point corresponds to $\omega = \omega_c$, which is very near the axion or ALP mass. The top (bottom) half of the curve which is increasing (decreasing) with respect to $E$ corresponds to forward (backward) scattering.

The frequency at which $dE/d\omega$ vanishes is $\omega = \omega_c$ (equivalently, $E = m_p$). This implies that photons whose frequency is in the vicinity of $\omega_c$ are produced by cosmic rays over the widest range of energy. More specifically, photons with a frequency near $\omega_c$ can be produced as long as the proton energy is above its rest mass. On the other hand, photons with a very low frequency near $m_a/2$ can only be produced by enormously energetic protons which back-scatter at $\theta = \pi$, while those with a frequency $\omega \gg m_a$ can only be produced by protons whose energy exceeds approximately $\sqrt{\omega m_a^2/(2m_a)}$.

Since that the cosmic ray flux and cross section decrease monotonically with $E$, we thus conclude that a peak is present in the energy spectrum of the photons. This has been verified numerically (by linearly extrapolating the proton flux from $E = m_p$ to 50 GeV); see figure B-6 for a plot of the spectrum, which features a salient peak, as expected.
B.3 Observational Consequences

B.3.1 Insufficient sensitivity to detect the QCD axion

We numerically integrate \((B–20)\) and find that, for \(m_a = 10^{-6} \text{ eV}\) (corresponding to roughly the expected mass of the QCD axion),

\[
\frac{dN_\gamma}{dt dA} \sim 10^{-21} \text{ cm}^{-2} \text{s}^{-1}.
\]

This estimate is quite conservative, as we only included energy \(E > 50 \text{ GeV}\) in the integration, due to uncertainties in the spectrum of the protons. If we extend it to as low as, say, \(10 \text{ GeV}\), we gain a boost in the flux by approximately a factor of ten (From \([111]\), we know that the flux actually increases as \(E\) is decreased to approximately \(1.4 \text{ GeV}\)). For \(n_a\), we use \(\rho_a/m_a\), where \(\rho_a = 10^{-24} \text{ g/cm}^3\), the expected local halo density of the Galaxy \([109]\). For the radius \(r\) of the region \(\mathcal{V}\), we adopt the value \(10^{22} \text{ cm}\), which is approximately one tenth of the size of the stellar disk. The collecting area and field of view of the photon detector are taken to be that of the Square Kilometre Array: \(10^{10} \text{ cm}^2\) and \(200 \text{ deg}^2\). Even with such a huge surface area, we expect only one photon every \(10^{11} \text{ s}\). Clearly, it is not yet possible to detect the QCD axion.

B.3.2 Constraining the parameter space of ALPs

Although the detection of the QCD axion seems out of reach given existing technologies, an interesting constraint can still be placed on the \((m_b, f_a^{-1})\) parameter space of ALPs, under the assumption that they constitute dark matter. Their number density is thus taken to be \(\rho_a/m_a\), where \(\rho_a \sim 10^{-24} \text{ g cm}^{-3}\). The sensitivity of the detector is assumed to be ten nanoJanskys. Over a frequency range of \(10^{11} \text{ Hz}\), which is roughly what we need to see the peak in the spectrum, this translates to a minimum rate of \(10^{-3} \text{ photons/cm}^2\text{s}\) (for the Square Kilometre Array, about \(10^7 \text{ photons per second}\)). Note that the constraint is only valid for an ALP whose mass lies between \(2.9 \times 10^{-7} \text{ eV}\) and \(4.1 \times 10^{-5} \text{ eV}\), since the Square Kilometre Array can only detect photons in this range with a field of view \(\delta \Omega_d\) of \(200 \text{ deg}^2\). For higher photon energies (in the range
Figure B-5. Photon energy spectrum $dF/dE \times d\sigma/d\omega$ versus photon energy $\omega$ for an axion mass of $10^{-6}$ eV, up to a normalization factor. For $E > 50$ GeV, the proton spectrum is given by $3.06(E/\text{GeV})^{-2.70}$. For $m_p < E < 50$ GeV, it is chosen to be a linear function that extrapolates to zero at $E = m_p$. This underestimates the abundance of low-energy cosmic ray protons, but does however capture the fact that the flux vanishes at very low proton velocities. Note that in our actual calculation we make the even-more-conservative approximation of neglecting all cosmic rays with $E < 50$ GeV.
Figure B-6. Hypothetical exclusion limit on the mass $m_a$ and the decay constant $f_a$ for axion-like particles over the mass range $2.9 \times 10^{-7} \text{ eV} < m_a < 4.1 \times 10^{-5} \text{ eV}$, in the case of a non-detection by the Square Kilometre Array. The other limits come from laser experiments, solar searches, and stellar evolution. The bottom line corresponds to the QCD axion.

$4.1 \times 10^{-5} \text{ eV} < m_a < 1.2 \times 10^{-4} \text{ eV}$ that it is capable of detecting), the field of view diminishes substantially, resulting in a weaker exclusion limit.

Our numerical result is shown in figure B-6. The exclusion limit is slightly worse than that from laser experiments, but substantially weaker than that obtained from stellar evolution. It is foreseeable that improvements in radio telescopy in the future could help boost the sensitivity and thus improve the constraint. If so, our proposal would have the potential to complement existing axion experiments in the future, since it relies on very different physics from that in photon regeneration experiments (and, in fact, all other methods of axion search).

Finally, we stress that the constraint shown in figure B-6 is only approximate, given the uncertainties in the halo density, and the flux spectrum and exact distribution of
cosmic ray protons in the Galaxy. Throughout our calculation, we were conservative in our estimates. For instance, we 1) disregard the contribution of all cosmic ray protons below 50 GeV, 2) consider only proton cosmic rays, and 3) only take into account cosmic rays within a distance of $10^{22}$ cm (one-tenth of the radius of the stellar disk) from Earth.

### B.4 Conclusions

We proposed a new observational probe of axions and ALPs, based on the detection of photons emitted by cosmic ray protons which propagate in a time-dependent axion field within the Galaxy. Since the axions couple very weakly to ordinary matter, we find, unsurprisingly, that the emitted photon flux is exceedingly low, of order $10^{-21}$ cm$^{-2}$ s$^{-1}$. Even with a detector whose collecting area is as large as $10^{10}$ cm$^2$ (e.g. the Square Kilometre Array), we expect only approximately one photon every $10^{11}$ s. Thus detection of the QCD axion by this mean is out of the question at the moment.

Nonetheless, the same mechanism can be exploited to impose exclusion limits on the parameter space of ALPs. Since their decay constant $f_a$ can be smaller, ALPs can couple more strongly to ordinary matter, thereby increasing the rate of photon emission. Under the assumptions that dark matter is primarily ALPs, and that a detection rate of $10^7$ photons per day is sufficient, we numerically find that a non-detection of photons by the Square Kilometre Array translates to exclusion limits on $f_a$ (for the mass range $2.9 \times 10^{-7}$ eV $< m_a < 4.1 \times 10^{-5}$ eV) which are slightly worse than those obtained by current photon regeneration experiments. However, since the physics underlying our proposed observation is quite different from that used in any other axion/ALP searches, it holds the promise for serving as an excellent complement to the various axion/ALP detection experiments in the future.
REFERENCES


[77] G. Steigman, private correspondence.


BIOGRAPHICAL SKETCH

I was born in Kunming, China, where I attended primary school, middle school and high school. All three schools were within one mile from my family home. After graduating from high school, I went to the city of Nanjing, China, where I obtained my Bachelor degree of science from Nanjing university. I came to the U.S. in 2005 and enrolled in the University of Kentucky where I obtained my Master degree of science in 2007, and then transferred to the University of Florida to study for a Doctor of Philosophy in physics.

I am currently working in the high energy group where my Ph.D. advisor is Professor Pierre Sikivie. My current research interest is in dark matter, axion physics and cosmology.