Chapter 5

Out-of-equilibrium and freeze-out:
WIMPs, Big-bang nucleosynthesis and recombination

In the last chapter, we have studied the equilibrium thermodynamics, and calculated thermal and expansion history in the early Universe. There, as the majority of cosmic energy budget in the early time comes from relativistic particles, the equilibrium thermal history is just enough to describe the expansion history during the radiation dominated epoch.

We, at the same time, have also encountered the case for neutrino decoupling where electron-neutrino scattering rate falls below the Hubble expansion rate. After the decoupling epoch, neutrino hardly interacts with cosmic thermal plasma and evolves independently. This phenomena of freeze-out and thermal decoupling is pretty generic in the expanding universe, and there are number of interactions in the Universe that undergo out-of-equilibrium: baryon-antibaryon asymmetry, (presumably) thermal decoupling of dark matter, neutrino decoupling, neutron-to-proton freezeout, big-bang nucleosynthesis, and CMB recombination, to mention a few important examples. These are the subject of this chapter.

It is this out-of-equilibrium phenomena that make our Universe very rich and interesting in contrast to the majority of energy budget (at early times) which are in equilibrium. Had it stayed in a perfect equilibrium state up until now, the Universe would be a boring place characterized by one number $T_{CMB} = 2.726$ K. We have seen that there are residual non-relativistic particles, surviving well after the temperature fell below the rest mass of the particle. In addition, if dark matters coupled to the cosmic plasma at earlier times, their abundance is also freeze-out to a finite relic density. Although minorities during the radiation dominated epoch, these non-relativistic particles eventually dominate the Universe because energy density of the equilibrium part—thermal radiation—drops more quickly ($\rho \propto T^4$) compared with non-relativistic relics ($\rho \propto T^3$) survive through the freeze-out; thus, matter era is opened up. The other events like neutrino decoupling, CMB recombination, and big-bang nucleosynthesis all provide a window for us, cosmologists, to directly look at the early history of the Universe without much of the contamination— because they are decoupled, otherwise, all memory about the initial states of the Universe would be completely erased in the name of equilibrium!

General reasoning for the freeze-out goes as following with the argument that we have used before: $t_i < t_H$ to maintain the equilibrium and $t_i > t_H$ to decouple. Let's think about the interaction whose cross section scales as $\langle \sigma v \rangle \propto T^n$. Then, the interaction rate ($\Gamma = t_i^{-1}$) is $\Gamma \propto n \langle \sigma v \rangle \propto T^{n+3}$. During the radiation dominated epoch, the Hubble expansion rate is proportional to $H \propto T^2$, which means that

---

1This net number has to be generated by a process called baryogenesis —though, we do not have the theory that everybody agrees on.
when \( n > -1 \), the interaction rate drops quickly than the expansion rate, and eventually hit \( \Gamma \approx H \) after which the interaction freezes out. Is \( n > -1 \) reasonable? For the interaction mediated by massive gauge boons, e.g. weak interaction, the cross section goes as \( \sigma \approx (g_I/m_I^2)^2 T^2 \), where \( g_I \) is the interaction coupling constant, and \( 1/m_I^2 \) is the mass of the gauge bosons. Then, \( T^2 \) is added to get the right dimension for the cross section. For the relativistic particles that moves with \( v \approx c \), we have \( \langle \sigma v \rangle \propto T^2 \); thus, we have \( n > -1 \) for sure. For the scattering via massless particles such as photons, at high energy \( (T > m_e) \), Klein-Nishina cross section scales as \( \sigma_{KN} \propto T^{-2} \), but at low energy, Thomson scattering cross section is constant, \( \sigma_T \propto \text{const.} \)

Note that at the high energy \( (T \gtrsim M_W \approx 80 \text{ GeV}) \), the interaction rate for both weak interaction and electromagnetic interaction scales as \( \Gamma \approx \langle \sigma v \rangle n \approx \alpha^2 T \) while Hubble rate is \( H \approx T^2/M_{pl} \), where \( M_{pl} \) is the reduced Planck mass and \( \alpha = \mathcal{O}(0.01) \) is the electroweak coupling constant. This reads an interesting conclusion that the interactions are frozen out for \( T \gtrsim \alpha^2 M_{pl} \approx 10^{16} \text{ GeV} \), and they cannot maintain the thermal equilibrium during these super early times!

In this chapter, we will refine the time scale argument, which works pretty well, in a more rigorous way by using kinetic theory. On top of that, the kinetic theory allows us to calculate the decoupling and freezeout procedure more accurately. We will then move on to discuss individual event that we have mentioned earlier.

### 5.1 Kinetic theory in the FRW universe

The evolution of the phase space distribution function is described by Boltzmann equation:

\[
\frac{df}{d\lambda} = \mathcal{C}[f],
\]

which looks pretty simple, and is indeed easy to understand! The left-hand side of the Boltzmann equation is the total derivative of the phase space distribution function following the trajectory. That must be balanced by the change of particles (source or sink) along the pathway due to interaction, which is encoded by the collision term in the right hand side of Eq. (5.1). If there is no source or sink, the right-hand side vanishes and the Boltzmann equation simply states the Liouville theorem that we discussed in the last chapter.

One difference you might notice from the usual (classical mechanics) form of the Boltzmann equation is that we now use the affine parameter \( \lambda \) instead of the time. It is to make the Boltzmann equation manifestly covariant; otherwise, the Boltzmann equation would depend on the choice of coordinate system. Of course, you can relate the two derivatives by using (see, [?], [?] for more rigorous ways of writing down Boltzmann equation covariant way):

\[
\varepsilon = \frac{dt}{d\lambda}.
\]

#### 5.1.1 Boltzmann equation in the FRW Universe

In the homogeneous isotropic FRW Universe, the particle distribution function only depends on the momentum and time \( f = f(p, t) \). The usual procedure of proceeding with Boltzmann equation is rewriting the total derivative in the left-hand-side acting on \( f(p, t) \) as

\[
\frac{df}{d\lambda} = \frac{dt}{d\lambda} \frac{\partial f}{\partial t} + \frac{dp}{d\lambda} \frac{\partial f}{\partial p},
\]

\[
(5.3)
\]
then use geodesic equation (eq. 2.26)

\[
\frac{dp}{d\lambda} = -H\varepsilon p
\]  

(5.4)

to rewrite

\[
\frac{df}{d\lambda} = \varepsilon \frac{df}{dt} = \varepsilon \left[ \frac{\partial f}{\partial t} - Hf \frac{\partial f}{\partial p} \right].
\]  

(5.5)

The right hand side of the Boltzmann equation depends on the nature of the interaction. In this chapter, we will see the two body interaction (binary collision) of \(1 + 2 \leftrightarrow 3 + 4\) that particle 1 and particle 2 annihilate producing particle 3 and 4, or vice versa. Let us focusing on particle 1. The change of \(f(r, p, t)\) from affine parameter \(t \sim t + \delta t\) can be written as

\[
f(t + \delta t) - f(t) = \frac{df}{dt}\delta t \equiv (\bar{R} - R)\delta t.
\]  

(5.6)

where \(\bar{R}\) and \(R\) are the shorthand notation of interaction rate that create (\(\bar{R}\)) or destroy (\(R\)) particle 1 at position \(r\) and momentum \(p\). How do we write \(\varepsilon[f] = \bar{R} - R\) for these binary collisions?

Let us consider a binary collision happening within a spatial volume \(d^3r\) at location \(r\). The number of collision destroying particle 1 and 2 and creating 3 and 4 within an interval \(\delta t\) can be written as

\[
R\delta\lambda = dN_{12}dP_{12\rightarrow 34}\delta t,
\]  

(5.7)

where \(dN_{12}\) is the initial number of colliding pairs \((p_1, p_2)\):

\[
dN_{12} = f_1(p_1)f_2(p_2)\frac{d^4p_1}{(2\pi)^4}(2\pi)^3\delta^{(1)}(p_1^2 - m_1^2)\frac{d^4p_2}{(2\pi)^4}(2\pi)^3\delta^{(1)}(p_2^2 - m_2^2)d^3r.
\]  

(5.8)

Here, we use the molecular chaos assumption that phase space density of particle species 1, 2 are completely uncorrelated so that expected density of pairings is simply given by multiplying the two densities. \(dP_{12\rightarrow 34}\) is the transition rate to the final state within \(d^3p_3d^3p_4\) which is related to the transition matrix \(|\mathcal{M}_{fi}|^2 = |(3,4)|\mathcal{M}|1,2\|^2\) as

\[
dP_{12\rightarrow 34} = Id\sigma
\]  

\[
= \frac{d^4p_3}{(2\pi)^4}(2\pi)^3\delta^{(1)}(p_3^2 - m_3^2)\frac{d^4p_4}{(2\pi)^4}(2\pi)^3\delta^{(1)}(p_4^2 - m_4^2)|\mathcal{M}_{fi}|^2(2\pi)^4\delta^{(4)}(p_4 + p_3 - p_1 - p_2)
\]

\[
\times (1 \pm f_3(p_3))(1 \pm f_4(p_4)),
\]  

(5.9)

where \(I = n_2(p_2)|v_{12}|\) is the incident flux of particle 2 in the particle 1’s rest frame—in the state cat \(|1, 2\rangle\)—, and the ± in the last line encodes influence of particle 3 and 4 to the rate due to bose enhancement (+) and Pauli exclusion principle (−). Note that the expression above is manifestly covariant, as \(f\) is a scalar. When the interaction is invariant under reflections and time reversal,

\[
\mathcal{M}_{fi} \equiv \langle p_3, p_4|\mathcal{M}|p_1, p_2\rangle = \langle -p_3, -p_4|\mathcal{M}| -p_1, -p_2\rangle = \langle p_1, p_2|\mathcal{M}|p_3, p_4\rangle = \mathcal{M}_{if}.
\]  

(5.10)

Combining Eqs. (5.8)–(5.9), we have the expression for \(R\) as

\[
R = g_2g_3g_4 \int\frac{d^4p_2}{(2\pi)^4}(2\pi)^3\delta^{(1)}(p_2^2 - m_2^2)\int\frac{d^4p_3}{(2\pi)^4}(2\pi)^3\delta^{(1)}(p_3^2 - m_3^2)\int\frac{d^4p_4}{(2\pi)^4}(2\pi)^3\delta^{(1)}(p_4^2 - m_4^2)
\]

\[
\times |\mathcal{M}_{fi}|^2\delta^{(4)}(p_4 + p_3 - p_1 - p_2)f_1(p_1)f_2(p_2)(1 \pm f_3(p_3))(1 \pm f_4(p_4)).
\]  

(5.11)
Now we can do the mass shell integration\(^2\):

\[
\int_0^\infty dp^0 \delta^{(1)}((p^0)^2 - \mathbf{p}^2 - m^2) = \frac{1}{2} = \frac{1}{2\sqrt{|\mathbf{p}|^2 + m^2}} = \int_0^\infty d\epsilon \frac{\delta^{(1)}(\epsilon - \sqrt{|\mathbf{p}|^2 + m^2})}{2\epsilon} \tag{5.13}
\]

to rewrite \(R\) as

\[
R = g_2 g_3 g_4 \int \frac{d^3p_2}{(2\pi)^3 2\epsilon_2} \int \frac{d^3p_3}{(2\pi)^3 2\epsilon_3} \int \frac{d^3p_4}{(2\pi)^3 2\epsilon_4} \times (\mathcal{M}_{fi})^2 \delta^{(4)}(|\mathbf{p}_1 - \mathbf{p}_2| - p_1 - p_2)f_1(\mathbf{p}_1)f_2(\mathbf{p}_2)(1 \pm f_3(\mathbf{p}_3))(1 \pm f_4(\mathbf{p}_4)). \tag{5.14}
\]

Because \(\mathcal{M}_{fi}^2 = \mathcal{M}_{fi}^2\), we calculate \(\tilde{R}\) in a similar manner, and find the collision term as

\[
\mathcal{C}[f] = g_2 g_3 g_4 \int \frac{d^3p_2}{2\epsilon_1} \int \frac{d^3p_3}{(2\pi)^3 2\epsilon_2} \int \frac{d^3p_4}{(2\pi)^3 2\epsilon_3} \times (2\pi)^4 \delta^{(4)}(p_4 + p_3 - p_1 - p_2)|\mathcal{M}_{fi}|^2 \times [f_3(\mathbf{p}_3)f_4(\mathbf{p}_4)(1 \pm f_1(\mathbf{p}_1))(1 \pm f_2(\mathbf{p}_2)) - f_1(\mathbf{p}_1)f_2(\mathbf{p}_2)(1 \pm f_3(\mathbf{p}_3))(1 \pm f_4(\mathbf{p}_4))]. \tag{5.15}
\]

Combining the two, we arrive at the final form of the Boltzmann equation:

\[
\frac{\partial f_1}{\partial t} - H p_1 \frac{\partial f_1}{\partial p_1} = g_2 g_3 g_4 \int \frac{d^3p_2}{(2\pi)^3 2\epsilon_1} \int \frac{d^3p_3}{(2\pi)^3 2\epsilon_2} \int \frac{d^3p_4}{(2\pi)^3 2\epsilon_3} \times (2\pi)^4 \delta^{(4)}(p_4 + p_3 - p_1 - p_2)|\mathcal{M}_{fi}|^2 \times [f_3(\mathbf{p}_3)f_4(\mathbf{p}_4)(1 \pm f_1(\mathbf{p}_1))(1 \pm f_2(\mathbf{p}_2)) - f_1(\mathbf{p}_1)f_2(\mathbf{p}_2)(1 \pm f_3(\mathbf{p}_3))(1 \pm f_4(\mathbf{p}_4))]. \tag{5.16}
\]

### 5.1.2 Equation for particle number density

To get equation for the number density, we integrate over all momenta:

\[
n_1 = g_1 \int \frac{d^3p_1}{(2\pi)^3} f_1(\mathbf{p}_1, t) \tag{5.17}
\]

Then, the left hand side becomes

\[
\frac{dn_1}{dt} - H \frac{1}{2\pi^2} \int_0^\infty d^3p_1 \frac{\partial f_1(\mathbf{p}_1, t)}{\partial p_1} = \frac{dn_1}{dt} - 3Hn_1 = \frac{1}{a^3} \frac{d(a^3n_1)}{dt}. \tag{5.18}
\]

The right hand side becomes more symmetric now:

\[
g_1 \int \frac{d^3p_1}{(2\pi)^3} \mathcal{C}[f] = g_1 \int \frac{d^3p_1}{(2\pi)^3 2\epsilon_1} \int \frac{d^3p_2}{(2\pi)^3 2\epsilon_2} g_2 \int \frac{d^3p_3}{(2\pi)^3 2\epsilon_3} g_3 \int \frac{d^3p_4}{(2\pi)^3 2\epsilon_4} g_4 \times (2\pi)^4 \delta^{(4)}(|\mathbf{p}_1 - \mathbf{p}_2| - p_1 - p_2)|\mathcal{M}_{fi}|^2 \times [f_3(\mathbf{p}_3)f_4(\mathbf{p}_4)(1 \pm f_1(\mathbf{p}_1))(1 \pm f_2(\mathbf{p}_2)) - f_1(\mathbf{p}_1)f_2(\mathbf{p}_2)(1 \pm f_3(\mathbf{p}_3))(1 \pm f_4(\mathbf{p}_4))]. \tag{5.19}
\]

\(^2\)Here we use

\[
\int dx \delta^{(1)}(f(x)) = \int dy \left[ \frac{dx}{dy} \right] \delta^{(1)}(y) = \sum_{\text{zeros of } f(x)} \frac{1}{|f'(x)|}. \tag{5.12}
\]
Therefore, the equation for the number density becomes
\[
\frac{1}{a^3} \frac{d(a^3 n_i)}{dt} = g_1 \int \frac{d^3 p_1}{(2\pi)^3 2\epsilon_1} g_2 \int \frac{d^3 p_2}{(2\pi)^3 2\epsilon_2} g_3 \int \frac{d^3 p_3}{(2\pi)^3 2\epsilon_3} g_4 \int \frac{d^3 p_4}{(2\pi)^3 2\epsilon_4} \times (2\pi)^4 \delta^{(4)}(p_4 + p_3 - p_1 - p_2) |M_{fi}|^2 \times [f_3(p_3)f_4(p_4)(1 \pm f_1(p_1))(1 \pm f_2(p_2)) - f_1(p_1)f_2(p_2)(1 \pm f_3(p_3))(1 \pm f_4(p_4))].
\]
(5.20)

After some hard work, we just confront a lengthy equation that seems to be quite hard to integrate! Even worse, for the most general cases, we have to write something like Eq. (5.20) for all four particles and solve them at the same time. Yes, a great fun is always there to solve the integro-differential equation.

## 5.2 After decoupling: collisionless Boltzmann equation

Without diving into the equation above, we can extract an important conclusion from Eq. (5.20) about the decoupled particle species. Let’s consider a particle species, that was in thermal equilibrium state before, but decouples from thermal plasma when temperature of the Universe is \( T = T_D \), and scale factor \( a_D \). The chemical potential of the particle was \( \mu_D \) at the time of decoupling. For the particles decoupled from the rest of the Universe, the right hand side of the Boltzmann equation vanishes, \( \mathcal{E}[f] = 0 \), and collisionless Boltzmann equation
\[
\frac{\partial f}{\partial t} - H p \frac{\partial f}{\partial p} = 0,
\]
(5.21)
governs the evolution of the phase space distribution function. This equation, indeed tells that the change of distribution function in this case is solely due to the redshift of the momentum. Let’s see this more directly by solving the equation.

In principle, the phase space distribution function of the decoupled species can take any form. Let us describe this generic function by time varying temperature and chemical potential. Then, the Boltzmann equation reads
\[
f \left( \frac{T}{\mu} \right)^2 \left[ \mu \left( \frac{\mu}{\mu} - \frac{T}{T} \right) + \epsilon \left( \frac{T}{T} + \frac{p^2}{\epsilon^2} \frac{\dot{a}}{a} \right) \right] e^{(\epsilon - \mu)/T} = 0,
\]
(5.22)
where dot represent the time derivative. For the particles that were relativistic at decoupling, \( \epsilon \approx p \), and the equation reduces to
\[
\mu \left( \frac{\mu}{\mu} - \frac{T}{T} \right) + p \left( \frac{T}{T} + \frac{\dot{a}}{a} \right) = 0,
\]
(5.23)
and the only way to satisfy this relation is to having \( Ta = \text{const.} \), and \( T/\mu = \text{const.} \). that reads
\[
T(a) = a_D T_D, \quad \mu(a) = \frac{a}{a_D} \mu_D,
\]
(5.24)
for \( a > a_D \). For particles that are non-relativistic \( \epsilon \approx m + p^2/(2m) \) at decoupling, the equation reduces to
\[
\dot{\mu} - \mu \frac{\dot{T}}{T} + \left( m + \frac{p^2}{2m} \right) \frac{\dot{T}}{T} + \frac{p^2}{\epsilon} \frac{\dot{a}}{a} \approx (m - \mu) \left( \frac{\mu}{m - \mu} + \frac{\dot{\mu}}{\dot{T}} \right) + \frac{p^2}{2m} \left( \frac{\dot{T}}{T} + 2 \frac{\dot{a}}{a} \right) = 0,
\]
(5.25)
to the second order in \( p^2 \). Again, the only way to satisfy this relation for all \( p \) is to have \( (m - \mu)/T = \text{const.} \), and \( Ta^2 = \text{const.} \), or
\[
T(a) = T_D \left( \frac{a_D}{a} \right)^2, \quad \mu(a) = m + (\mu_D - m) \frac{T(a)}{T_D}.
\]
(5.26)
One can easily show that, for both cases, the comoving particle number density is conserved after the decoupling, which is consistent with Eq. (5.20) which says that the comoving number density \( (a^3 n) \) is conserved when the right hand side vanishes. Then, from the equation we have derived in the previous chapter, the comoving entropy of the particle is separately conserved because
\[
d(a^3) = -\frac{\mu}{T} d(na^3) = 0.
\]

In fact, for non-relativistic particles, this two constrains restrict the time evolution of the temperature and the chemical potential as you have figured out in the previous homework.

### 5.2.1 Prelude for dark matters

What are particles completely decoupled from the thermal bath, and do not interact with photons at all after the decoupling? We call this particles dark matters. Strictly speaking, dark matters do not have to be in thermal contact with plasma at all at early times. But, there is a class of dark matter candidates which do that: WIMPs, Weakly Interacting Massive Particles. These particles had interacted with standard model particles, or among themselves via weak interaction. Almost all—except for those searching for axionic dark matters—the dark matter search programs (either direct or indirect) are looking for the signatures from WIMPs.

We normally categorize dark matters into three classes, hot, warm and cold, depending on its role in the large-scale structure formation. The usual criteria is the smallest structure that the dark matter can form: cold dark matter can form structures all the way down to the sub-galactic scales, warm dark matters form structures from galactic scales, while hot dark matters can only form structures above the galactic scales.

When applying this criteria to the WIMPs that decouples at early time, WIMPs become

- **cold dark matter** if it is already non-relativistic at the time of decoupling,
- **warm dark matter** if it is relativistic at the time of decoupling, but non-relativistic now,
- **hot dark matter** if it has been relativistic for entire history of the Universe until now (or very recently).

Although the criteria above does not exactly coincide with the earlier definition, the division captures important physical characteristics of each category: hot dark matter first form a huge overdensity then fragmented into smaller pieces, warm dark matter can form galactic size halos but no/small substructures, cold dark matter is responsible for all the subgalactic scale minihalos.

One obvious candidate for warm/hot dark matters is neutrinos: the massless neutrinos are the perfect candidate for the hot dark matter, and the massive neutrinos can be warm/hot dark matter candidates depending on the mass. There is no cold dark matter candidates in the standard model of particle physics, but going beyond the SM, there are massive right-handed neutrinos and LSP (lightest supersymmetric particles) or neutralino (electric neutral super-partners of gauge bosons) as candidates for cold dark matter. Neutralinos are mixture of superpartners of the photon, \( Z^0 \) boson, and Higgs bosons (called photino, zino, higgsino, respectively), spin-1/2, and Majorana—i.e., it is its own antiparticle.

### 5.2.2 Massive neutrinos

In the standard model of particle physics, neutrinos are assumed to be massless particles. But, now we know that neutrinos must have mass from the neutrino oscillation observed through solar (\( \nu_e \) from \( p-p \)
chains — $4p \rightarrow ^4\text{He} + 2e^+ + 2\nu_e$ —; Kamiokande, Sudbury Neutrino Observatory, etc), atmospheric ($\pi^\pm \rightarrow \mu^\pm + \nu_\mu (\bar{\nu}_\mu)$, or $\pi^\pm \rightarrow \mu^\pm \rightarrow e^\pm + \nu_e (\bar{\nu}_e) + \bar{\nu}_\mu (\nu_\mu)$; Kamiokande, Super-Kamiokande, etc), accelerator ($\mu^+ \rightarrow e^+ \nu_e \bar{\nu}_\mu$; K2K, T2K, MINOS), and reactor ($\bar{\nu}_e$, KamLAND, Daya Bay, RENO, Double Chooz) neutrinos. This clearly is one of the important clues to find out the physics beyond the standard model of particle physics.

The neutrino oscillation is a result of the neutrino flavor mixing. That is, the flavor ($e$, $\mu$, $\tau$) state is not the eigenstate of the Hamiltonian (time evolution operator), and can be in general written as the linear combination of the neutrinos with mass $m_j$:

$$|\nu_{ll}\rangle = \sum_j U_{jl} |\nu_{jL}\rangle.$$  \hspace{1cm} (5.28)

Here, $|\nu_{jL}\rangle$ stands for the mass eigenstates (massive neutrinos) and $U_{jl}$ is the unitary matrix called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) mixing matrix. From the neutrino oscillation data (if we have three massive neutrino states), we find the mass gap between three mass states as $(1 - \sigma, 68\% \text{C.L.})$ range of:

$$\Delta m_{21}^2 = m_2^2 - m_1^2 \approx 7.53 \pm 0.18 \times 10^{-5} \text{eV}^2,$$  \hspace{1cm} (5.29)

$$|\Delta m_{32}^2| = |m_3^2 - m_2^2| \approx 2.52 \pm 0.07 \times 10^{-3} \text{eV}^2 \text{ (normal hierarchy : } m_3 > m_2 > m_1)$$

$$\approx 2.44 \pm 0.06 \times 10^{-3} \text{eV}^2 \text{ (inverted hierarchy : } m_2 > m_1 > m_3).$$  \hspace{1cm} (5.30)

Note that we cannot observe the absolute mass of the neutrinos from neutrino oscillation as the flavor changing probability only depends on the square of mass gap. $\Delta m_{21}^2$ is measured from the solar neutrinos, and $\Delta m_{32}^2$ is measured from the atmospheric neutrinos.

The mass gap, however, can be translated as the lower bound of the total mass of the massive neutrinos by setting the mass of the lightest neutrino species to zero. They are

$$0.067(0.106) \text{eV} < \sum_i m_{vi},$$  \hspace{1cm} (5.31)

for normal (inverted) hierarchy spectrum of neutrino mass. The upper bound comes from the tritium beta decay\(^4\), by very accurate measurement of kinematics, that

$$m_{vi} < 2 \text{eV} \rightarrow \sum_i m_{vi} < 6 \text{eV}. \hspace{1cm} (5.33)$$

As we will discuss in the later chapters, there might be a chance to detect the absolute mass of the neutrinos from cosmological observations and put a tighter constraint. Also, the next generation of $\beta$-spectroscopy experiments (e.g. KATRIN in Germany) are aiming at reaching a sensitivity limit $m(\nu_e) < 0.2 \text{eV}$, a cosmologically relevant range.

\(^3\)The numbers come from PDG http://pdg.lbl.gov

\(^4\)Tritium is a radioactive material with half life 12.32 yrs. Its beta decay

$$T \rightarrow ^3\text{He}^{++} + e^- + \bar{\nu}_e$$  \hspace{1cm} (5.32)

releases only 18.6 keV of energy shared by electron and anti-neutrino, and therefore can be used as a probe of the neutrino mass. Depending on the mass of neutrinos, the high-energy tail of the emitted electron energy distribution have a cut-off at 18.6 keV $- m_{vi}$. This kind of experiment is hard as we have to read off the high-energy tail of the electron distribution, which is very unusual.
But, for our purpose in this section, it is important to note that the lower limit of the mass is greater than the temperature of the neutrino

\[ T_\nu = 1.946 \, K \simeq 0.168 \, \text{meV} \ll 34(53) \, \text{meV} < m_{\nu i}, \tag{5.34} \]

at its absolute minimum when one of the three mass state is massless (number in the parenthesis is for the inverted hierarchy). It means neutrinos are non-relativistic particles at present epoch! As we have seen in the previous chapter, cosmic background neutrinos have decoupled from the thermal plasma at \( T_{\nu-\text{dec.}} \simeq 1.5 \, \text{MeV} \), and free-streaming since then. Because the neutrinos were relativistic at the time of decoupling (mass upper bound (2 eV) \( \ll T_{\nu-\text{dec.}} \)), the distribution function freezes as the form,

\[ f(p_D) = \frac{1}{e^{(p_D - \mu_D)/T_D} + 1} \tag{5.35} \]

and all that the expansion of the Universe changes is redshifting the neutrino momentum as \(|p| = a_D/a|p_D|\), to yield the distribution function as the same Fermi-Dirac function

\[ f(p) = \frac{1}{e^{(p - \mu)/T} + 1}, \tag{5.36} \]

with \( T = (a_D/a)T_D \) and \( \mu = (a_D/a)\mu_D \).

What is the chemical potential of the neutrinos? As we have seen in the previous chapter, chemical potential is related to the particle-antiparticle asymmetry. With neutrino-antineutrino asymmetry, the total energy density of neutrinos is different from the standard value, and this effect allow us to bound—from BBN, for example—the neutrino chemical potential to be (at 2\( \sigma \))

\[ -0.05 < \frac{\mu_\nu}{T_\nu} < 0.07. \tag{5.37} \]

Note that the mixing of neutrinos leads to equilibration of asymmetries in different flavors: so, we have only one number here. Therefore, we can safely neglect the chemical potential of the neutrinos. Then, the number of neutrinos is the same as the number of anti-neutrinos, and we have

\[ n_{\nu i} = \frac{3}{4} \frac{g_{*n}}{4} \left( \frac{T_{\nu i}}{T} \right)^3 n_\gamma = \frac{3}{4} \times \frac{4}{11} \times 411 \, \text{cm}^{-3} \simeq 112.09 \, \text{cm}^{-3} \tag{5.38} \]

neutrinos per species. Then, because the number density of all three species are the same, the neutrino mass density at present is

\[ \rho_\nu = n_{\nu i} \sum_i m_{\nu i} = 112.09 \sum_i m_{\nu i} \, \text{cm}^{-3}, \tag{5.39} \]

from which we calculate the density parameter for the neutrino as

\[ \Omega_\nu h^2 = \frac{\rho_\nu h^2}{\rho_{\text{crit}}} = \frac{\sum_i m_{\nu i}}{94.02 \, \text{eV}}. \tag{5.40} \]

By using the known bound from the terrestrial experiments, we can find the bound of the neutrino density parameter as

\[ 7.13 \times 10^{-4}(1.13 \times 10^{-3}) < \Omega_\nu h^2 < 0.0638. \tag{5.41} \]

Again, parenthesis is for the inverted hierarchy. As the best-fitting mass density from Planck is \( \Omega_{\text{dm}} h^2 \simeq 0.12 \), neutrinos are just not enough to explain the observed dark matter energy density. In terms of the energy density, it can be at most a half of the total dark matter energy budget. Looking at the opposite
angle, one can find an upper limit on the neutrino mass from \( \Omega_{\nu} h^2 < \Omega_{dm} h^2 \approx 0.12 \) that is translated as

\[
m_{\nu i} < 3.76 \text{ eV}
\]

for the mass of a neutrino species when neutrino oscillation equalize mass of three neutrino species. This cosmological bound on the neutrino mass is called \textit{Gershtein-Zeldovich bound}\(^5\). This bound is just the very first cosmological bound. The observables in the large-scale structure of the Universe such as galaxy power spectrum and angular power spectrum of CMB put more stringent upper limit on the neutrino mass. For example, from Planck alone, the \( 1 - \sigma \) (68% C.L.) upper bound is

\[
\sum_i m_{\nu i} < 0.403 \text{ eV.}
\]

A recent review\(^6\) presents a thorough overview about the effect of massive neutrinos on cosmological observations. We will come back to this issue when we discuss the large-scale structure, but basically the constraint comes about because neutrinos are warm dark matter. For warm dark matter, perturbations cannot grow until they become non-relativistic \( T_\nu < m_\nu \), and warm dark matters generate quite different large-scale structure than cold dark matter, which is the main guest of honor in the next section.

### 5.3 Boltzmann equation simplified

Now, let’s go back to the full Boltzmann equation in Eq. (5.20). To be exact, this integro-differential equations must be solved for all particle species 1, 2, 3 and 4 that are involved in the interaction. Fortunately, however, for the cases that we are interested in, Eq. (5.20) can be reduced to quite a simple form, as we make following simplification\(^6\):

- We don’t need to solve them for all four particles involved in the collosion process at all, because we will only consider a particle/process being non-equilibrium at a time. That is, other three particle species are safely in the thermal equilibrium state. Also, for many cases the particle/process that we are interested in is maintaining kinetic equilibrium, but goes out-of-equilibrium by departure from chemical equilibrium (that is, inverse reaction of chemical equation stops happening, but still there can be frequent scattering to maintain them in the same temperature).

- Because we are typically interested in the case where \( (\epsilon - \mu) \gtrsim T \) where the spin-statistics (Fermi-Dirac and Bose-Einstein) is not important, and the distribution function follows Maxwell-Boltzmann statistics. Therefore, we can neglect the Bose enhancement and Pauli blocking term altogether, and the phase-space distribution function becomes

\[
f(p) \simeq e^{-(\epsilon(p) - \mu)/T} \equiv e^{\mu/T} f^{(0)}(p).
\]

- Using this approximation, the number density of particle species \( i \) can be written as

\[
n_i = g_i \int \frac{d^3 p}{(2\pi)^3} f_i(p) = e^{\mu_i/2} \left[ g_i \int \frac{d^3 p}{(2\pi)^3} f_i^{(0)}(p) \right] \equiv e^{\mu_i/T} n_i^{(0)},
\]

where \( n_i^{(0)} \) is the equilibrium number density of the species \( i \) without chemical potential.

\(^5\)Gershtein S S, Zel’dovich Ya B Pis’ma Zh. Eksp. Teor. Fiz. 4 174 (1966) [English translation JETP Lett. 4 120 (1966)]

\(^6\)Here, we follow the logic in the textbooks of Kolb & Turner and Dodelson.
With these three simplifications, the only unknown that we want to figure out is the chemical potential \( \mu_i \), or equivalently, the number density \( n_i \) (they are related by the known number density \( n_i^{(0)} \) at a given temperature). Let’s see how it works. First, using Maxwell-Boltzmann distribution function,

\[
\left[ f_3(p_3)f_4(p_4)(1 \pm f_1(p_1))(1 \pm f_2(p_2)) - f_1(p_1)f_2(p_2)(1 \pm f_3(p_3))(1 \pm f_4(p_4)) \right]
\]

\[
\simeq e^{-(\epsilon_3 + \epsilon_4)/T} e^{(\mu_3 + \mu_4)/T} - e^{-(\epsilon_1 + \epsilon_2)/T} e^{(\mu_1 + \mu_2)/T} = f_1^{(0)}(p_1)f_2^{(0)}(p_2) \left[ \frac{n_3n_4}{n_3^{(0)}n_4^{(0)}} - \frac{n_1n_2}{n_1^{(0)}n_2^{(0)}} \right], \tag{5.46}
\]

and the definition of average \( \langle \sigma v \rangle \) as

\[
\langle \sigma v \rangle \equiv \frac{1}{n_1^{(0)}n_2^{(0)}} \prod_{i=1}^{4} \left[ g_i \int \frac{d^3p_i}{(2\pi)^3 2\epsilon_i} \right] (2\pi)^4 \delta^{(4)}(p_4 + p_3 - p_1 - p_2) \cdot \mathcal{M}_{fi}^2 f_1^{(0)}(p_1)f_2^{(0)}(p_2). \tag{5.47}
\]

We simplify the Boltzmann equation as

\[
\frac{1}{a^3} \frac{d(n_1a^3)}{dt} = \langle \sigma v \rangle n_1^{(0)}n_2^{(0)} \left[ \frac{n_3n_4}{n_3^{(0)}n_4^{(0)}} - \frac{n_1n_2}{n_1^{(0)}n_2^{(0)}} \right], \tag{5.48}
\]

which is much simpler than before, and is indeed a simple ordinary differential equation for the number densities for a given weighted average of the interaction cross section \( \sigma \) and relative velocity \( v \).

Let’s digest the simplified equation. This equation contains the time scale argument that we had before. The left-hand side is of order \( n_1/t \approx n_1H \), as Hubble time is the typical time scale that the number density of particle has changed. The right-hand side is of order \( n_1n_2 \langle \sigma v \rangle \approx n_1\Gamma_i \), which is the interaction rate per volume. First, when interaction rate is larger than the Hubble expansion rate \( \Gamma_i \gg H \), the individual terms in the right hand side is much larger than the left hand side. To make the equality, the large terms must cancel each other to yield

\[
\frac{n_3n_4}{n_3^{(0)}n_4^{(0)}} = \frac{n_1n_2}{n_1^{(0)}n_2^{(0)}}, \tag{5.49}
\]

which is nothing but a condition of chemical equilibrium! You can clearly see this from Eq. (5.45). One could also derive the same equation from the detailed balance as the time derivative in the left-hand side must vanish in the equilibrium state.

At the other extreme, when the interaction rate is smaller than the expansion rate \( \Gamma_i \ll H \), we can neglect the right hand side all together, and the situation is the same as no-interaction case where the comoving number density \( n_1a^3 \) is conserved.

Now that we have our master equation, we are ready to tackle some interesting out-of-equilibrium events. In the following sections of this chapter, we will apply this equation to three such cases: thermal decoupling of WIMPs, Big-bang nucleosynthesis and CMB recombination.

### 5.4 Weakly-Interacting-Massive-Particles

Because of the weak but non-zero interactions,

\[
\psi + \bar{\psi} \rightarrow X + \bar{X}, \tag{5.50}
\]
the WIMPs were in thermal equilibrium state with cosmic plasma (consists of particle $X = \gamma, e^-/e^+, \cdots$) at earlier times by means of pair creation/annihilation. But, at some point the interaction rate drops below the Hubble expansion rate to freeze out the interaction, and WIMP abundance. Let us assume that the annihilation products $X$ are in local thermodynamical equilibrium state with other particles because of interactions other than the one we consider. This should be the case for the standard model particles which participate the electromagnetic interactions, e.g. photons, electrons, etc. Also, for simplicity, let us further assume that the chemical potential for dark matters is negligible $\mu_\psi \ll 1$ so that $n_\psi = n_\bar{\psi}$. This is certainly true for the Majorana particles, when they pair-annihilate to some particle and antiparticle pairs. For Dirac particles, this means that the number density of particles and antiparticles are the same: $n_\psi = n_\bar{\psi}$. In this case, the Eq. (5.48) becomes

$$\frac{1}{a^3} \frac{d(n_\psi a^3)}{dt} = -\langle \sigma v \rangle \left[ n^2_\psi - \left( n^\text{eq}_\psi \right)^2 \right]. \tag{5.51}$$

That is, we have an equation only with the equilibrium number density of the dark matter which determines the source term. For the Majorana particles, because we create/remove two particles per interaction, one might think that we need a symmetric factor of 2. But, that factor of 2 is cancel by that for total $N$ identical particles, there are $N(N - 1)/2$ different ways of pairing (to annihilate) them.

Because the comoving number density of dark matters is conserved after decoupling, it is convenient to define a new variable

$$Y \equiv \frac{n_\psi}{s}, \tag{5.52}$$

which **takes a constant value after the decoupling**. In terms of $Y$, we rewrite the equation as

$$\frac{dY}{dt} = -\langle \sigma v \rangle s \left[ Y^2 - Y^2_\text{eq} \right]. \tag{5.53}$$

Finally, we use

$$x = m_\psi / T$$

as a time variable, which is related to the cosmic time from the Friedmann equation (here, we assume that $g_* \text{ varies slowly}):

$$t = \frac{1}{2H} = \left( \frac{45}{16\pi^3 G_N g_*(T)} \right)^{1/2} \frac{1}{T^2} \approx 0.3012 \left( \frac{m_{\text{pl}}}{T^2} \right)^{1/2} = 0.3012 \left( \frac{m_{\text{pl}}}{T^2} \right)^{1/2} \left( \frac{m_{\text{pl}}}{m^2} \right)^{1/2} \left( \frac{m}{m_{\text{pl}}} \right)^2 \equiv \frac{x^2}{2H(m,x)}, \tag{5.54}$$

where $m_{\text{pl}} = 1.2211 \times 10^{-19} \text{ GeV} = 1/\sqrt{G_N}$ is the Planck mass, and we define the function $H(m,x) = 1.6602 g_*^{1/2}(m^2/m_{\text{pl}})$ so that $H = H(m,x)x^{-2}$. Note that the $x$-dependence of $H(m,x)$ comes only through the time evolution of $g_*(T)$, but we can practically ignore (tolerating $\sim %$ level error) that by using the approximation that $g_*$ varies only slowly. With this in mind, but not forget that we need to put $x$-dependency back to achieve the $\sim %$ accuracy, we will write $H(m,x) = H(m)$ hereafter. Then, we trade the time derivative to

$$\frac{d}{dt} = \frac{dx}{dt} \frac{d}{dx} = \frac{H(m)}{x} \frac{d}{dx}, \tag{5.55}$$

then transform the Eq. (5.51) as

$$\frac{dY}{dx} = -\frac{1}{H(m)} \left[ \frac{x}{\langle \sigma v \rangle s} \left[ Y^2 - Y^2_\text{eq} \right] \right]. \tag{5.56}$$

We can recast Eq. (5.56) once more to the following suggestive form

$$\frac{x}{Y_\text{eq}} \frac{dY}{dx} = -\frac{\Gamma_\psi}{H} \left[ \left( \frac{Y}{Y_\text{eq}} \right)^2 - 1 \right], \tag{5.57}$$

where $\Gamma_\psi$ is the annihilation width, and $Y_\text{eq}$ is the equilibrium number density of the dark matter when they pair-annihilate to some particle and antiparticle pairs.
with $\Gamma_\psi = n_\psi^{eq} (\sigma v)$ being the annihilation rate. This equation, again, tells that the ratio $\Gamma_\psi / H$ is the key parameter to quantify the effectiveness of the annihilation, in a sense that when $\Gamma_\psi \lesssim H$, the annihilation stop being efficient and the dark matter fraction $Y = n_\psi / s$ freezes out.

### 5.4.1 Hot/warm relic dark matter

For the hot/warm dark matter WIMPs, which decouples when they are still relativistic $m_\psi \ll T$, the decoupling procedure is exactly same as what happens to the neutrinos at $T \approx 1.5 \text{ MeV}$: once the decoupling temperature has reached, the particle species thermally decouples with the relativistic particle number density at the time of decoupling. Because the number density of relativistic particles is solely determined by the temperature, as long as the decoupling happens during the time $g_{*s} = \text{constant}$, the equilibrium number density are the same as the actual number density and the right hand side of Eq. (5.51) vanishes:

$$Y = Y_{eq} = \frac{45 \zeta(3)}{2\pi^4} \frac{g_{\psi n}}{g_{*s}} \approx 0.2777 \frac{g_{\psi n}}{g_{*s}},$$  

(5.58)

where $g_{\psi n} = g_\psi$ for bosons and $3/4 g_\psi$ for fermion dark matters. Then, the relic abundance $Y_\infty \equiv Y(x \rightarrow \infty)$ is also given by the $Y_{eq}$ at freeze-out time $x_f$.

$$Y_\infty = Y_{eq}(x_f) = 0.2777 \frac{g_{\psi n}}{g_{*s}(x_f)}.$$  

(5.59)

Using this, we also calculate the number density of hot/warm relic at present time as

$$n_{\psi 0} = s_0 Y_\infty = 803.4 \left( \frac{g_{\psi n}}{g_{*s}(x_f)} \right) \text{cm}^{-3},$$  

(5.60)

using

$$s_0 = \frac{2\pi^2}{45} g_{*s} T_{cmb}^3 = 2893.1 \text{ cm}^{-3},$$  

(5.61)

with $T = 2.726 \text{ K}$. Note that the relic density fo hot/warm dark matter is very insensitive to the freezeout temperature $T_f = m / x_f$ as the dependence comes only through $g_{*s}$.

The energy density of hot/warm dark matter can then be written with the mass $m$ as

$$\rho_{\psi 0} = mn_{\psi 0} = 803.4 \left( \frac{g_{\psi n}}{g_{*s}(x_f)} \right) \left( \frac{m}{\text{eV}} \right) \text{eV/cm}^3,$$

(5.62)

which corresponds to the density parameter of

$$\Omega_{\psi} h^2 = 0.07623 \left( \frac{g_{\psi n}}{g_{*s}(x_f)} \right) \left( \frac{m}{\text{eV}} \right).$$

(5.63)

Because the hot/warm dark matter can be at most the total matter density, we can find the upper bound for the mass from $\Omega_{\psi} h^2 \lesssim 0.12$:

$$m \lesssim 1.5742 \left( \frac{g_{*s}(x_f)}{g_{\psi n}} \right) \text{eV}.$$  

(5.64)

Also, to be counted as dark matter, the particle must be non-relativistic at present time:

$$m \gtrsim T_{\psi 0} = T_{\psi f} \frac{a_f}{a_0} = \left( \frac{3.909}{g_{*s}(x_f)} \right)^{1/3} T_{cmb} = 0.2350 \left( \frac{3.909}{g_{*s}(x_f)} \right)^{1/3} \text{meV}.$$  

(5.65)
This gives us a mass bound for the hot/warm relics as

\[
2.350 \times 10^{-4} \left(\frac{3.909}{g_{s\psi}(x_f)}\right)^{1/3} \lesssim \frac{m}{\text{eV}} \lesssim 1.5742 \left(\frac{g_{s\psi}(x_f)}{g_{\psi/n}}\right). \tag{5.66}
\]

As we have discussed earlier, neutrinos are the most well-known example of the warm dark matter. Neutrinos decoupled at around 1.5 MeV where \(g_{\psi/n} = 10.75\), with \(g_{\psi/n} = 1.5\), we reproduce the upper bound that we found earlier:

\[
\sum_i m_{\nu i} < 11.28 \text{ eV}. \tag{5.67}
\]

If hot/warm dark matter interacts with heavier gauge bosons, e.g. one in the supersymmetric models, the weak interaction cross section can be modified as \(G_F \rightarrow G_F \left(\frac{M_W}{M}\right)^2\), and the decoupling temperature can be increased by a factor of \(T_f \propto G^{-2/3} \propto (M/M_W)^{4/3}\). For example, the heavy gauge boson of mass \(M \gtrsim 560\) TeV can lead to the decoupling temperature of \(T_f \gtrsim 200\) GeV where \(g_{s\psi} = 106.75\), and the mass bound becomes (for spin-1/2 fermions)

\[
m < 112.03 \text{ eV}. \tag{5.68}
\]

These hypothetical dark matter particles turned to non-relativistic in the radiation dominated epoch.

### 5.4.2 Cold relic dark matter

For the cold dark matter that decouples when \(T \lesssim m_{\psi}\), the situation becomes a bit more complicated as the equilibrium number density \(Y_{eq}\) of non-relativistic particles depends exponentially on the temperature for \(x = m/T \gg 1\):

\[
Y_{eq}(x) = \frac{45\zeta(3)}{2\pi^4} \frac{g_{\psi/n}}{g_{s\psi}} = \frac{45}{4\pi^4} \frac{g_{\psi/n}}{g_{s\psi}} \int_x^\infty \frac{u^2 - x^2}{e^u + 1} du \rightarrow \frac{90}{(2\pi)^{7/2} g_{s\psi}(x)} x^{3/2} e^{-x} \simeq 0.1447 \frac{g_{\psi/n}}{g_{s\psi}(x)} x^{3/2} e^{-x}, \tag{5.69}
\]

using the number density of non-relativistic particles and entropy density:

\[
n_{\psi} = g_{\psi} \left(\frac{m^2}{2\pi x}\right)^{3/2} e^{-x}, \quad s = \frac{2\pi^2}{45} g_{s\psi} m^3 x^{-3}. \tag{5.70}
\]

That is, when the temperature drops below the mass of the dark matter, the number density of dark matter particles drops drastically to reduce the interaction (annihilation) rate, which is proportional to the number density. Therefore, if interaction rate is not too strong, the annihilation interaction rate freezes out around \(T \sim O(1) m_{\psi}\) and it leaves some relic that can survive from exponentially suppressed equilibrium number density\(^7\).

Let us start from the annihilation cross section, which must take the form of

\[
\langle \sigma v \rangle \equiv \sigma_0 \left(\frac{T}{m}\right)^n = \sigma_0 x^{-n}, \tag{5.71}
\]

\(^7\)As we have seen in the previous chapter, non-relativistic baryons and leptons are thermally coupled to plasma (therefore also have to suffer from the exponential suppression) but still have non-zero number density for those particles because they have the chemical potential of order \(\mu/T \approx n_{b/s} \approx 10^{-10}\) from the beginning. We don’t yet have a concrete theory about how to get such a large chemical potential from the fundamental theory, and this is an active research area called baryogenesis. We will come back to this later in this section.
Figure 5.1: Freeze-out of cold relics abundance $Y = n_\psi/s$ as a function of $x = m/T$ for $\lambda = 10^5$, $10^7$ and $10^9$ (red, green, blue, respectively) along with the decoupling temperature $x_f = m/T_f$ shown as a vertical dashed lines with the same color. Black dashed line show the case for thermal equilibrium which suffers from strong—exponential—suppression for $x \gg 1$.

for non-relativistic dark matter particles ($T \lesssim m$) where $\langle v \rangle \propto T^{1/2}$. Here, $n$ depends on the decay channel: $n = 0$ for $s$-wave, $n = 1$ for $p$-wave, etc. Then, we rewrite the Boltzmann equation Eq. (5.56) as

$$\frac{dY}{dx} = -\lambda x^{n-2} \left( Y^2 - Y_{eq}^2 \right),$$  

with

$$\lambda \equiv \left[ \frac{x \langle \sigma v \rangle_s}{H(m)} \right]_{s=1} = \frac{\sigma_0 s(m)}{H(m)} \simeq 0.2642 \left( \frac{g_{*s}}{\sqrt{g_{*s}}} \right) \frac{m_{Pl} \sigma_0}{m m}$$  

a constant depending on the mass of dark matter, and $g_{*s}, \sqrt{g_{*s}}$, and $\langle \sigma v \rangle$ measured at $T = m$.

We can then integrate the equation for a given value of $\lambda$ and $m$. As an example, let us consider the dark matter with $g_{\psi} = 2, m_\psi \gtrsim 100$ GeV (this makes $g_{*s} \simeq g_* \simeq 106.75$), and $\lambda = 10^{5-9}$. We show the result of numerically integrating Eq. (5.72) in Fig. 5.1. As we expected earlier, for $x < x_f$, the abundance $Y$ closely follows the equilibrium values, while it freezes out after the decoupling and asymptotes to the constant value $Y_\infty$. As increasing $\lambda$, decoupling temperature decreases to reduce the final relic abundance.

Let’s estimate the freezeout temperature and the abundance $Y_\infty$ first. Freeze-out happens at $\Gamma_\psi(x_f) = H(x_f)$. Transforming the equation to the form of Eq. (5.57),

$$\frac{x}{Y_{eq}} \frac{dY}{dx} = -\lambda Y_{eq} x^{-n-1} \left[ \left( \frac{Y}{Y_{eq}} \right)^2 - 1 \right]$$  

(5.74)
we find the criteria of $x_f$ as 
\[
\frac{\Gamma_\psi}{H}(x_f) = 1 = \lambda Y_{\text{eq}}(x_f) x_f^{-n-1}
\]
\[
= 0.1447 \frac{g_\psi}{g_*} x_f^{-n+1/2} e^{-x_f} = 0.03823 \frac{g_\psi}{\sqrt{g_*}} m_{\text{pl}} \sigma_0 x_f^{-n+1/2} e^{-x_f}. \tag{5.75}
\]
Note that here we drop all the time dependence of $g_\psi$’s as the time scale that is relevant here is usually much smaller than the time scale that $g_{\psi i}$ are varying (which is of order tens of Hubble time). We take the logarithm of both side,
\[
\ln \left[ 0.03823 \frac{g_\psi}{\sqrt{g_*}} m_{\text{pl}} \sigma_0 \right] = \left( n - \frac{1}{2} \right) \ln x_f + x_f, \tag{5.76}
\]
which can be iteratively solved to yield
\[
x_f = \ln \left( 0.03823 \frac{g_\psi}{\sqrt{g_*}} m_{\text{pl}} \sigma_0 \right) - \left( n - \frac{1}{2} \right) \ln \left[ \ln \left( 0.03823 \frac{g_\psi}{\sqrt{g_*}} m_{\text{pl}} \sigma_0 \right) \right] + \cdots. \tag{5.77}
\]
After the freeze-out, the actual number density is much bigger than the equilibrium number density $Y \gg Y_{\text{eq}}$ and the differential equation becomes
\[
\frac{dY}{dx} = -\lambda x^{-n-2} Y^2, \tag{5.78}
\]
which can be integrated as
\[
\frac{1}{Y_f} - \frac{1}{Y_\infty} = -\frac{\lambda}{n+1} x_f^{n-1}. \tag{5.79}
\]
Typically $Y_f$ is significantly larger than $Y_\infty$, as the number density further decays even after the freeze-out time, then we find
\[
Y_\infty = \frac{n+1}{\lambda} x_f^{n+1} \approx \frac{3.785(n+1)x_f^{n+1}}{g_{\psi}/g_{\ast}^{1/2} m_{\text{pl}} \sigma_0} = \frac{3.785(n+1)\left( g_{\ast}^{1/2} / g_\psi \right) x_f}{m_{\text{pl}} \langle \sigma v \rangle_f}, \tag{5.80}
\]
where $\langle \sigma v \rangle_f$ refers to the value at freezeout. Note that the relic density is larger for smaller cross-section. It is because the smaller cross section means earlier decoupling, which saves the dark matters out of the equilibrium density (which is exponentially suppressed) earlier. Dark matters with larger annihilation cross-section, on the other hand, decouple later and yield smaller relic abundance.

The estimate of $Y_\infty$ is directly related to the dark matter abundance and number density:
\[
n_{\psi 0} = n_0 = 1.095 \times 10^4 \frac{(n+1)\left( g_{\ast}^{1/2} / g_\psi \right) x_f^{n+1}}{m_{\text{pl}} \sigma_0} \text{cm}^{-3}, \tag{5.81}
\]
\[
\rho_{\psi 0} = m n_{\psi 0} = 1.095 \times 10^4 \frac{(n+1)\left( g_{\ast}^{1/2} / g_\psi \right) x_f^{n+1}}{m_{\text{pl}} \sigma_0} \left( \frac{1}{\text{eV}} \right) \text{eV/cm}^3, \tag{5.82}
\]
that corresponds to the density parameter of
\[
\Omega_{\psi 0} h^2 = 1.039 \times 10^9 \frac{(n+1)\left( g_{\ast}^{1/2} / g_\psi \right) x_f}{m_{\text{pl}} \langle \sigma v \rangle} \text{GeV}^{-1}. \tag{5.83}
\]
Fourth generation massive neutrinos

As an example of the cold relic, let us consider the fourth generation massive neutrinos with $m \gg \text{MeV}$ which is non-relativistic at decoupling ($x_f \gtrsim 1$). The annihilation cross-section depends on the nature of particle: for the Dirac type neutrinos, annihilations process is dominated by $s$-wave ($n = 0$) with

$$\sigma_0 \simeq G_F^2 m^2 = 1.36044 \times 10^{-10} \left( \frac{m}{\text{GeV}} \right)^2 \text{GeV}^{-2},$$  

(5.84)

then $\lambda = 3.400 \times 10^9 m_{\text{GeV}}^3$, and the freezeout happens at (using $g_{\psi} = 2$, $g_{\ast} \simeq 60$)

$$x_f \simeq 16.613 + 3 \ln m_{\text{GeV}} + \frac{1}{2} \ln (16.613 + 3 \ln m_{\text{GeV}}) \simeq 18.02 + 3 \ln m_{\text{GeV}},$$  

(5.85)

which gives the WIMP number density

$$n_{\psi 0} \simeq 8.5096 \times 10^{-7} \frac{18.02 + 3 \ln m_{\text{GeV}}}{m_{\text{GeV}}^3} \text{cm}^{-3} \simeq 3.744 n_{p 0} \frac{18.02 + 3 \ln m_{\text{GeV}}}{m_{\text{GeV}}^3},$$  

(5.86)

Here, we use the proton density of

$$n_{p 0} = \left(1 - \frac{Y_p}{2}\right) \rho_{\text{crit}} \Omega_b = 2.273 \times 10^{-7} \left( \frac{\Omega_b h^2}{0.023} \right) \text{cm}^{-3},$$  

(5.87)

with $Y_p = 0.24$. The density parameter for the cold relic is

$$\Omega_{\psi 0} h^2 \simeq 0.08074 \frac{18.02 + 3 \ln m_{\text{GeV}}}{m_{\text{GeV}}^2} \simeq \frac{1.455}{m_{\text{GeV}}^2} \left[1 + \frac{1}{6} \ln m_{\text{GeV}}\right],$$  

(5.88)

because we have the same amount of anti-particles, we have

$$\Omega_{\bar{\psi} \psi 0} h^2 \simeq \frac{2.91}{m_{\text{GeV}}^2}.$$

(5.89)

The relic density in this case gets smaller for more massive particles, just because cross-section becomes bigger for the larger mass.

From the measured dark matter density parameter $\Omega_{\text{dm}} h^2 < 0.12$, we have a firm lower bound of the right-handed neutrinos as

$$m \gtrsim 4.9 \text{ GeV}.$$  

(5.90)

This bound is called Lee-Weinberg bound [?], although it was discovered many others (all published in the same year, 1997) including P Hut, K. Sato and, of course, Y. Zel’dovich.

WIMP miracle

To get the observed cold dark matter fraction of $\Omega_{\text{cdm}} h^2 \simeq 0.12$, we need

$$\Omega_{\psi 0} h^2 = 1.039 \times 10^9 \frac{(n + 1) \left( g_{\ast}^{1/2} / g_{s} \right) x_f}{m_{\text{Pl}} \langle \sigma v \rangle} \text{GeV}^{-1} \simeq 0.12,$$

(5.91)

or, equivalently,

$$\langle \sigma v \rangle \simeq 7.0906 \times 10^{-10} (n + 1) g_{\ast}^{1/2} / g_{s} x_f \text{GeV}^{-2} \simeq 8.2824 \times 10^{-27} (n + 1) g_{\ast}^{1/2} / g_{s} x_f \text{cm}^3 / \text{s}.$$  

(5.92)
5.5. **BIG-BANG NUCLEOSYNTHESIS**

For a weak scale mass $m \simeq 100\text{ GeV}$, $\sigma_0 \simeq \alpha_W^2/m^2 \simeq 10^{-8}\text{ GeV}^{-2}$, then $x_f \simeq 25$. Having $g_{\ast s} \simeq g_\ast \simeq 106.75$, we need (for $s$-wave annihilate),

$$\langle \sigma v \rangle \simeq 2.004 \times 10^{-26}\text{ cm}^3/\text{s} \simeq 1.713 \times 10^{-9}\text{ GeV}^{-2},$$  \hspace{1cm} (5.93)

which is remarkably close to the cross section at the weak-scale $\sigma_0 \simeq 10^{-8}\text{ GeV}^{-2}$.

That is, *if there is new physics at the electroweak scale that involves the introduction of a new stable neutral particle, then the particle must have a relic density in the ballpark of that required to explain the cold dark matter density.* It’s truly a remarkable coincidence as there is no *a priori* reason that a constant in the elementary particle physics (the weak scale) has anything to do with the dark matter density in the unit of critical density, which is set by the expansion rate of the Universe. Completely different scale, and completely different physics! This is called a *WIMP miracle*, and this coincidence led a number of theorists and experimentalists to take the idea of WIMP dark matter very seriously!

### 5.4.3 Need for baryogenesis

Another application of this calculation is the freeze-out of baryon-antibaryon. Consider the Universe begins without any baryon asymmetry. With the exactly same logic that we study in this section, when annihilation rate drops below the Hubble expansion rate (happens after baryons turns non-relativistic), baryon-antibaryon pair annihilation freezes out and we have a finite relic density of baryons and antibaryons. The annihilation cross section is $\langle \sigma_{\text{ann}} v \rangle \simeq 1\text{ fm}^2 = 10^{-26}\text{ cm}^2 \simeq 25.683\text{ GeV}^{-2}$, with $m \simeq 1\text{ GeV}$, it makes $g_\star \simeq 60$ and we find $x_f \simeq \ln(0.03823(2/\sqrt{60})(10^{19})(25)) \simeq 42$, then $T_f \simeq 23\text{ MeV}$.

That is, 10 decades increase in the cross section increases $x_f$ by a factor of 2. But, that means that the relic abundance decrease by twice longer time in the exponent to yield $Y_\infty = n_b/s = n_{\bar{b}}/s \simeq 6.5 \times 10^{-20}$! This, of course is 9 orders of magnitude too small compared to the observed baryon-to-photon ratio. Therefore, there must have been an initial baryon asymmetry, and the abundance of antibaryons should be entirely negligible today.

### 5.5 Big-bang Nucleosynthesis

The next case of the freeze-out that we consider is Big-Bang Nucleosynthesis. The observed Helium abundance from the metal-poor part of the Universe suggests that the primordial Helium abundance is about 24% of total baryonic mass: $Y_p = 4n_{\text{He}}/n_b \simeq 0.24$.

What does this number implies about the formation of Helium nucleus? Let us suppose that all the Helium nuclei in the Universe is formed in the core of the stars. Becase the binind energy of a Helium nucleus is 28.3 MeV, 7.1 MeV ($\simeq 1.14 \times 10^{-5}\text{ erg}$) of energy is released per baryon ($m_b \simeq m_p \simeq 1.67 \times 10^{-24}\text{ g}$) when Helium nucleus is formed. If we assumes that this process has been converted a quarter of baryons to Heliums during the past ten billion years (the Hubble time, $t_H \simeq 4.41 \times 10^{17}\text{ s}$), then the average baryonic mass-to-light ratio must be

$$\frac{M}{L} = \frac{M}{(E/t_H)/4} \approx 4 \frac{1.67 \times 10^{-24}\text{ g}}{1.14 \times 10^{-5}\text{ erg}/(4.41 \times 10^{17}\text{ s})} \approx 0.258 \frac{\text{g}}{\text{erg/s}} \approx 0.5 \frac{M_\odot}{L_\odot}.\hspace{1cm} (5.94)$$

This is far much smaller than the observed value of $M/L \gtrsim 20M_\odot/L_\odot$. Therefore, only a small fraction of Helium can form inside the stars, and majority of Helium nuclei must form at early time as the cosmological event called **Big-bang nucleosynthesis (BBN)**!
By the way, where are the fusion energy of BBN now? Because the number density of neutron is about $10^{-9}$ of the photon number density, and BBN happens at around 0.1 MeV scale, the fusion energy released from BBN is only a small fraction of the total radiation energy. At this early time, free-free emission and double-compton scattering is so efficient that any energy release can be very efficiently thermalized to absorb and digest the fusion energy as a blackbody radiation field that we observe today as Cosmic Microwave Background.

BBN happens around $T \approx 0.1$ MeV which happens a few minutes after the Big-bang. As end products, BBN forms the light nuclei such as D, T, $^3$He, $^4$He, $^7$Li, etc, whose primordial abundance can be measured from the observations such as QSO absorption lines (D), 3.46 cm spin-flip transition ($^3$He) as well as the abundance in the most metal poor environments ($^4$He and $^7$Li). The dominant processes of forming those elements is the two-body interaction of protons and neutrons and their compound nuclei like Deutrium and Tritium:

\[
\begin{align*}
    n + p & \rightarrow D + \gamma \\
    D + D & \rightarrow ^3\text{He} + n \\
    D + p & \rightarrow ^3\text{He} \\
    D + D & \rightarrow ^3\text{He} + p \\
    ^3\text{He} + n & \rightarrow ^4\text{He} + p \\
    ^3\text{He} + ^4\text{He} & \rightarrow ^7\text{Be} \\
    ^4\text{He} + T & \rightarrow ^7\text{Li} \\
    ^7\text{Li} + p & \rightarrow ^4\text{He} + ^4\text{He} \\
    ^7\text{Be} + n & \rightarrow ^7\text{Li} + p.
\end{align*}
\]

(5.95)

Using these interactions \[ ? \] find out analytic approximation to the final light element abundance from BBN (see also \[ ? \] for the fixed point approach).

General BBN calculation follows three steps:

1. $n/p$-freeze out,
2. Deutrium bottle-neck,
3. onset of light-element formation.

It is easy to see from the Chemical reactions in Eq. (5.95) that Deutrium is the key element of this process: Deutriums are needed to form Tritium and Helium-3 that are required to form Helium-4. Therefore, if there is not enough Deutriums in the Universe, the reaction rate will not be enough to form other nuclei. It is only after the Deutrium abundance becomes large enough that the Deutrium bottle-neck is opened to form elements like T, $^3$He and $^4$He. The abundance of Deutrium is sensitive to the number density of protons and neutrons from which the Deutrium is formed. Therefore, we shall start our study on BBN from neutrons and protons that set up the initial condition.

In this section, we discuss each item in a greater detail.
5.5. BIG-BANG NUCLEOSYNTHESIS

5.5.1 Setting up the initial condition: neutron freeze-out

Even without any two-body interactions, neutrons can $\beta$-decay into protons:

$$n \rightarrow p + e^- + \bar{\nu}_e,$$

with lifetime of $\tau_n = 880.3$ s. The mass of neutron is $m_n \approx 939.5654$ MeV and of proton is $m_p \approx 938.2720$ MeV, therefore the $\beta$-decay releases $Q \equiv m_n - m_p \approx 1.2933$ MeV. But, in the energy scale that we are in here, $T \approx 1 \sim 0.5$ MeV, age of the Universe is $t \lesssim 10$ second after the Big-Bang and the $\beta$-decay of neutrons is not yet important.

At these high temperatures, the following weak interactions are dominant channel to convert neutrons to protons and vice versa:

$$n + \nu_e \leftrightarrow p + e^-$$

$$n + e^+ \leftrightarrow p + \bar{\nu}_e,$$

whose combined interaction rate is given\(^8\) by

$$\lambda_{n \rightarrow p} = \frac{255}{\tau_n} \left( \frac{T}{Q} \right)^5 \left[ \left( \frac{Q}{T} \right)^2 + 6 \left( \frac{Q}{T} \right) + 12 \right] \text{s}^{-1},$$

which is about $5.5 \text{s}^{-1}$ at $T = Q = 1.2933$ MeV, faster than the expansion rate ($H \approx 1.133 \text{s}^{-1}$) at $T = Q$. Such a fast interaction rate leads the neutron to proton ratio to be the equilibrium value:

$$X_n \bigg|_{T \gtrsim Q} = \frac{n_n}{n_p} \bigg|_{T \gtrsim Q} = \left( \frac{m_n}{m_p} \right)^{3/2} e^{-\left( m_n - m_p \right)/T} \left( \frac{\mu_n - \mu_p}{\mu_p - \mu_{\nu_e}} \right) \approx e^{-Q/T} e^{\left( \mu_{\nu_e} - \mu_p \right)/T} \approx e^{-Q/T}. \quad (5.99)$$

Here, I use the chemical equilibrium condition $\mu_n - \mu_e = \mu_p - \mu_{\nu_e}$. This means that number density of neutrons are the same as the number density of protons at high temperature $T \gtrsim Q$.

Like the previous section of dark matter, let us define the dimensionless quantity $X_n$ as

$$X_n = \frac{n_n}{n_n + n_p}, \quad (5.100)$$

and use

$$x = \frac{Q}{T} \quad (5.101)$$

as the time variable. The equilibrium condition becomes

$$X_n(x \lesssim 1) = X_n^{\text{eq}} \simeq \frac{1}{e^x + 1}. \quad (5.102)$$

We the apply Eq. (5.48) to the case at hand to find out the evolution equation for neutron number density:

$$\frac{1}{a^3} \frac{d}{dt} (a^3 n_n) = -\lambda_{n \rightarrow p} \left( n_n - \frac{n_n^{\text{eq}}}{n_p^{\text{eq}}} n_p \right) = -\lambda_{n \rightarrow p} \left( n_n - e^{-Q/T} n_p \right), \quad (5.103)$$

\(^8\)Because the temperature is greater than the mass of electron $T > m_e \approx 0.511$ MeV, one can safely assume that the number density of non-relativistic particles (neutrons and protons) are much smaller than the relativistic particles (electrons and neutrinos): $n_n, n_p \ll n_e, n_{\nu_e}$.\n
where $\lambda_{n \rightarrow p} = \sum n \langle \sigma v \rangle$ is the total interaction rate that sums over the two reactions, and we use the detailed balance to fix the ratio between the source term and the sink term of the neutrons. Using $n_n = (n_n + n_p) X_n$, we rewrite the equation in terms of $X_n$,

$$\frac{dX_n}{dx} = -\frac{\lambda_{n \rightarrow p} x}{H(Q)} \left[ X_n - e^{-x} (1 - X_n) \right] = -\frac{\lambda_{n \rightarrow p} x}{H(Q)} \left( 1 + e^{-x} \right) \left( X_n - X_n^{eq} \right). \quad (5.104)$$

Here, we convert the time derivative to the $x$ derivative by using Eq. (5.55) with

$$H(Q) = 1.6602 g_\ast^{1/2} (Q^2/m_p) \approx 7.4561 \times 10^{-16} \text{MeV} \approx 1.133 \text{s}^{-1} \quad (5.105)$$

is the Hubble parameter at $T = Q$: $H(x) = H(Q)x^{-2}$. So, we end up with an ordinary differential equation Eq. (5.104) describing the evolution of the neutron fraction $X_n$, that we can solve with an initial condition of Eq. (5.102). We show the result of numerical integration in Fig. 5.2. The asymptotic value that we find numerically is $X_n^{\infty} \approx 0.147617$ when we ignore the $\beta$-decay that follows the freeze-out that further reduce the neutron fraction as:

$$X_n(t > t_f) \approx X_n^{\infty} e^{-t/\tau_n} \approx 0.147617 e^{-t/\tau_n}. \quad (5.106)$$

Let us estimate the freeze-out temperature of the neutrons. The story goes pretty much similar to what we have discussed earlier for the dark matter. As temperature drops, the interaction rate reduces sharply so that it eventually falls below the Hubble expansion rate when $\lambda_{n \rightarrow p} = H$, or

$$\frac{255}{\tau_n x_f^5} \left[ x_f^2 + 6x_f + 12 \right] = 1.133 x_f^{-2}, \quad (5.107)$$
which yields
\[ x_f = 1.9056 \rightarrow T_f = \frac{Q}{x_f} = 0.6787 \text{MeV}. \] (5.108)

Indeed, the freeze-out happens before electron-positron pair annihilation.

Now we estimate the freeze-out fraction \( X_n^\infty = X_n(x \to \infty) \) without the \( \beta \)-decay. To do so, let's find the formal solution of Eq. (5.104). Let \( f(x) = X_n - X_n^{\text{eq}} \), then Eq. (5.104) becomes

\[ \frac{df(x)}{dx} + \frac{\lambda_{n\to p}x}{H(Q)}(1 + e^{-x})f(x) = \frac{e^x}{(e^x + 1)^2}. \] (5.109)

Homogeneous solution is

\[ f^{(0)}(x) \propto \exp \left[ - \int x d\xi \frac{\lambda_{n\to p}(x')}{x'H(x')} (1 + e^{-x'}) \right], \] (5.110)

then general solution may be written as \( f(x) = g(x)f^{(0)}(x) \) where \( g(x) \) satisfies

\[ \frac{dg(x)}{dx} = \frac{1}{f^{(0)}(x)} \frac{e^x}{(e^x + 1)^2}. \] (5.111)

Integrating once more, we find

\[ f(x) = f^{(0)}(x)g(x) = \int^x d\tilde{x} \frac{e^{\tilde{x}}}{(e^{\tilde{x}} + 1)^2} \exp \left[ - \int^\xi d\xi' \frac{\lambda_{n\to p}(x')}{x'H(x')} (1 + e^{-x'}) \right], \] (5.112)

and

\[ X_n(x) = X_n^{\text{eq}} + \int_0^x d\tilde{x} \frac{e^{\tilde{x}}}{(e^{\tilde{x}} + 1)^2} \exp \left[ - \int^\xi d\xi' \frac{\lambda_{n\to p}(x')}{x'H(x')} (1 + e^{-x'}) \right]. \] (5.113)

Note that we apply the boundary condition \( X_n(x \to 0) \to X_n^{\text{eq}} \) at the last equation. At freeze-out, \( X_n \gg X_n^{\text{eq}} \) and we estimate it by setting \( x \to \infty \):

\[ X_n^\infty = \int_0^\infty d\tilde{x} \frac{e^{\tilde{x}}}{(e^{\tilde{x}} + 1)^2} \exp \left[ - \int^\xi d\xi' \frac{\lambda_{n\to p}(x')}{x'H(x')} (1 + e^{-x'}) \right]. \] (5.114)

The integral in the bracket can be done analytically to read:

\[ \int^\infty d\xi' \frac{\lambda_{n\to p}(x')}{x'H(x')} (1 + e^{-x'}) = 0.25567 \left[ \frac{1}{\tilde{x}^2} + \frac{3}{\tilde{x}^3} \right] e^{-\tilde{x}} + \frac{1}{\tilde{x}} + \frac{3}{\tilde{x}^2} + \frac{4}{\tilde{x}^3}. \] (5.115)

Then, the freeze-out neutron fraction is

\[ X_n^\infty = 0.147617, \] (5.116)

as we found in the numerical solution earlier! Wait a minute. Didn't I intend to 'estimate' the result? Well, if there is a full analytical solution, why not?
5.5.2 Deuterium bottleneck

After neutron fraction freezes out, Deutrium \((m_D \approx 1875.612 \text{ MeV})\) must form first via

\[ n + p \rightleftharpoons D + \gamma. \quad (5.117) \]

But, if temperature of the Universe is high enough so that there are enough photons to destroy the Deuteriums \( (\epsilon(\gamma) > B_D = m_n + m_p - m_D \approx 2.225 \text{ MeV}) \), the reverse process dominates. As we see earlier, no further interactions will be proceeded until we have enough Deuterium. This is called Deuterium bottleneck.

Let us consider the equilibrium abundance of Deuterium. In the last chapter, we derived the number density of non-relativistic particles

\[ n_i(T) = g_i \left( \frac{m_i T}{2\pi} \right)^{3/2} e^{-\left(\frac{m_i - \mu_i}{T}\right)}, \quad (5.118) \]

then

\[ e^{\mu_i/T} = \frac{n_i}{g_i \left( \frac{m_i T}{2\pi} \right)^{3/2} e^{m_i/T}} \quad (5.119) \]

In chemical equilibrium, \( \mu_n + \mu_p = \mu_D \), which yields

\[ \left( \frac{1}{2} \frac{n_n}{(m_n T/2\pi)^{3/2} e^{m_n/T}} \right) \left( \frac{1}{2} \frac{n_p}{(m_p T/2\pi)^{3/2} e^{m_p/T}} \right) = \left( \frac{1}{3} \frac{n_D}{(m_D T/2\pi)^{3/2} e^{m_D/T}} \right), \quad (5.120) \]

or the mass fraction of Deuterium \((X_D)\)

\[ X_D = \frac{2n_D}{n_n + n_p} = \frac{3}{2} \frac{n_n n_p}{2 n_n + n_p} \left( \frac{2\pi}{m_n m_p} \right)^{3/2} e^{\left(\frac{m_n + m_p - m_D}{T}\right)} \]

\[ = \frac{3}{2} X_n X_p n_b \left( \frac{2\pi}{m_D m_p} \right)^{3/2} e^{B_D/T}. \quad (5.121) \]

Using the baryon-to-photon ratio parameter\(^9\)

\[ \eta_{10} \equiv 10^{10} \frac{n_b}{n_\gamma} = 10^{10} \frac{n_b}{\frac{2\pi(3)}{\pi^2} T^3} \approx 4.105 \times 10^{10} n_b T^{-3}, \quad (5.123) \]

the Deuterium fraction becomes

\[ X_D = 3.654 \times 10^{-11} \eta_{10} X_n X_p (2\pi T)^{3/2} \left( \frac{m_D}{m_n m_p} \right)^{3/2} e^{B_D/T} \]

\[ \approx 5.647 \times 10^{-14} \eta_{10} X_n X_p (2.225/\text{MeV})^{3/2} e^{2225/T_{\text{MeV}}}. \quad (5.124) \]

This fraction is quite small earlier. For example, at \( T_{\text{MeV}} = B_D \approx 2.225 \) (> \( Q \)), \( X_n \approx 0.359 \) and \( X_p \approx 0.641 \) (their equilibrium fraction), then \( X_D \approx 7.369 \times 10^{-13} (\Omega_b h^2 / 0.023) \ll 1 \). We show the equilibrium

\[ \eta_{10} \equiv 10^{10} \frac{n_b}{n_\gamma} = 10^{10} \frac{\rho_{\text{crit}}}{m_p n_\gamma} \approx 10^{10} \frac{1.1232 \times 10^{-3} \Omega_b h^2}{411} = 6.286 \left( \frac{\Omega_b h^2}{0.023} \right) \quad (5.122) \]
fraction of neutrons, protons and Deuteriums in Fig. 5.3 as a thin cyan dashed line (exact calculation is shown as the thick cyan line). As shown in Fig. 5.3, the thermal equilibrium approximation does a fair job to explain the full result, although it shows some deviation at earlier time when \( X_n \) abundance freezes out (because we assume the full thermal equilibrium), and later time when nucleosynthesis starts.

Anyway, for both approximation and exact solution, it is clear that it takes some time for the deuterium abundance \( X_D \) to catch up with the neutron and proton abundance even after the temperature of the Universe drops below \( T \lesssim B_D \simeq 2.225 \text{ MeV} \). For example, let us estimate the time it takes to achieve \( X_D \simeq 0.01 \) from Eq. (5.124). For this estimate, we ignore the neutron decay and use \( X_n = X_n^\infty \) and \( X_p = 1 - X_n^\infty \) to have the equation for the temperature as

\[
X_D \simeq 1.4824 \times 10^{-13} \left( \frac{T_{1\%}}{B_D} \right)^{3/2} e^{B_D/T_{1\%}} = 0.01, \tag{5.125}
\]

or

\[
\frac{B_D}{T_{1\%}} - \frac{3}{2} \ln \left( \frac{B_D}{T_{1\%}} \right) \simeq 24.93, \tag{5.126}
\]

then, we find

\[
\frac{B_D}{T_{1\%}} \simeq 24.93 + \frac{3}{2} \ln(24.93) + \cdots \simeq 29.76, \tag{5.127}
\]

so that \( T_{1\%} \simeq B_D/29.76 \simeq 0.0748 \text{ MeV} \). Although Deuterium abundance in the exact solution does not quite reach that value (maximum abundance is about \( X_D \simeq 0.0036 \) at \( T \simeq 0.0729 \text{ MeV} \)), it is close enough to the temperature when the bottleneck is opened (that we will estimate shortly).

This delay of Deuterium formation comes about because the baryon-to-photon ratio \( \eta = n_b/n_r \) is so small. That is, because we have about a billion photons per every baryon, there still exist photons at Wein tail of the Planck curve—the photons energetic enough to destroy the Deuterium nucleus—even when the mean photon energy is below the binding energy of the Deuterium. Therefore, we have wait until the Universe is cool enough so that photons with \( \epsilon > B_D \) disappear. This happens at \( T \lesssim B_D/30! \)

From Eq. (5.124) (using \( X_n = X_n^\infty \simeq 0.1476 \)), we calculate the equilibrium temperature \( (T_{\text{MeV}}) \) abundance \( (X_D) \) relation as,

\[
\frac{B_D}{T_{\text{MeV}}} - \frac{3}{2} \ln \left( \frac{B_D}{T_{\text{MeV}}} \right) = \ln \left( \frac{X_D}{5.647 \times 10^{-14} \eta_{10} X_n X_p B_D^{3/2}} \right) \simeq \ln \left( \frac{X_D}{(\eta_{10}/6.286)} \right) + 29.54, \tag{5.128}
\]

then using the iteration,

\[
\frac{B_D}{T_{\text{MeV}}} \simeq 29.54 + \ln \left( \frac{X_D}{(\eta_{10}/6.286)} \right) + \frac{3}{2} \ln \left[ 29.54 + \ln \left( \frac{X_D}{(\eta_{10}/6.286)} \right) \right], \tag{5.129}
\]

we find the temperature (in MeV) corresponding to the Deuterium concentration \( X_D \):

\[
T_{\text{MeV}}^{-1} = 13.28 + 0.449 \ln \left( \frac{X_D}{(\eta_{10}/6.286)} \right) + 0.674 \ln \left[ 29.54 + \ln \left( \frac{X_D}{(\eta_{10}/6.286)} \right) \right]. \tag{5.130}
\]

This equation is the key to estimating the final Helium abundance.

Then, when does the Deuterium bottleneck open? Let us estimate \( T^{\text{(open)}} \). The Deuterium bottleneck opens up when following interactions destroying Deuterium are efficient to form heavier elements:

\[
D + D \rightarrow ^3\text{He} + n
\]
\[
D + D \rightarrow T + p \tag{5.131}
\]
The interaction rate of these reactions between $T = 0.06 \text{ MeV}$ and $T = 0.09 \text{ MeV}$ is measured to be
\[
\langle \sigma v \rangle_{\text{DD} \rightarrow \gamma \text{Hen}} = (1.3 - 2.2) \times 10^{-17} \text{cm}^3 \text{s}^{-1} \\
\langle \sigma v \rangle_{\text{DD} \rightarrow \text{Tp}} = (1.2 - 2) \times 10^{-17} \text{cm}^3 \text{s}^{-1}.
\] (5.132)

Then, the change of number density of Deuterium due to these two interactions are
\[
\frac{1}{a^3} \frac{d(a^3 n_D)}{dt} = -\left[ \langle \sigma v \rangle_{\text{DD} \rightarrow \gamma \text{Hen}} + \langle \sigma v \rangle_{\text{DD} \rightarrow \text{Tp}} \right] n_D^2 \equiv -\langle \sigma v \rangle n_D^2.
\] (5.133)

In terms of the abundance, $X_D = 2n_D/n_b$, the equation can be written as
\[
\frac{dX_D}{dt} = -\frac{1}{2} \langle \sigma v \rangle X_D^2 n_b,
\] (5.134)
and the change due to the interactions are visible when
\[
\frac{1}{2} \langle \sigma v \rangle X_D^2 n_b t \simeq X_D,
\] (5.135)
or
\[
X_D^{(\text{open})} \simeq \frac{2}{\langle \sigma v \rangle n_b t} = 2.2697 \times 10^{-6} \eta_{10}^{-1} \left( \frac{\eta_{10}}{6.286} \right)^{-1} \left( \frac{g_*}{3.363} \right)^{1/2}.
\] (5.136)

Here, we use the time-temperature relation $t = 2.420/\sqrt{g_* T_{\text{MeV}}^2}$, and $n_b = \eta_{10}/(4.105 \times 10^4)T^3 \simeq 3.171 \times 10^{21} \eta_{10} T_{\text{MeV}}^3 \text{cm}^{-3}$ (Eq. (5.123)). Plugging the temperature-concentration relation Eq. (5.130) into Eq. (5.136), we have
\[
X_D^{(\text{open})} \simeq \left[ 3.014 \times 10^{-5} + 1.019 \times 10^{-6} \ln \left( \frac{X_D^{(\text{open})}}{\eta_{10}/6.286} \right) \right] \left( \frac{\eta_{10}}{6.286} \right)^{-1} \left( \frac{g_*}{3.363} \right)^{1/2},
\] (5.137)
whose solution is, again using iteration,
\[
X_D^{(\text{open})} \simeq 3.014 \times 10^{-5} \left( \frac{\eta_{10}}{6.286} \right)^{-1} \left( \frac{g_*}{3.363} \right)^{1/2} \left[ 1 + 0.03381 \ln \left( \frac{3.012 \times 10^{-5} \left( \frac{g_*}{3.363} \right)^{1/2}}{\eta_{10}/6.286} \right)^2 \right]
\approx 1.953 \times 10^{-5} \left( \frac{\eta_{10}}{6.286} \right)^{-1} \left( \frac{g_*}{3.363} \right)^{1/2} \left[ 1 - 0.1043 \ln \left( \frac{\eta_{10}}{6.286} \right) + 0.02609 \ln \left( \frac{g_*}{3.363} \right) \right].
\] (5.138)

Therefore, for our reference cosmology, $\Omega_b h^2 = 0.023$, $g_* = 3.363$, the Deuterium bottleneck opens when the concentration becomes $X_D^{(\text{open})} \simeq 1.800 \times 10^{-5}$ which happens at $T \simeq 0.0967 \text{ MeV}$. Note that one can calculate the concentration $X_D$ at the bottle neck opening for given $g_*$ and $\eta_{10}$ from Eq. (5.138).

### 5.5.3 BBN: after bottleneck is opened

In the previous section, we show that the Deuterium bottleneck opens up at $T \simeq 0.0967 \text{ MeV}$ when $X_D^{(\text{open})} \simeq 1.8 \times 10^{-5}$. Then, the things happens so rapidly and the equilibrium abundance of the Deuterium reaches about 1% of the total baryon mass shortly after (when $T \simeq 0.075 \text{ MeV}$). Around that time, the Deuterium abundance peaks because the deuterium production rate $(n + p \rightarrow D + \gamma)$ and the consumption rate $(D + D \rightarrow ^3\text{He} + n$ and $D + D \rightarrow T + p)$ are equal around that time:
\[
\frac{\lambda_{pn} X_p X_D}{\lambda_{DD} X_D^2} \simeq 10^4 \left( \frac{10^{-4}}{X_D} \right)^2,
\] (5.139)
Figure 5.3: Evolution of the mass fraction of various nuclides relevant in the Big-bang nucleosynthesis process: neutron (n), proton (p), Deuterium (D), Triton (T), 3-Helium ($^3$He), 4-Helium ($^4$He) 7-Lithium ($^7$Li), 7-Belilium ($^7$Be) from about 1 sec after Big-bang to one hour. We use the publically available version (v4.1) of Kawano code with the standard cosmology: $N_\nu = 3$, $\Omega_b h^2 = 0.023$. 
where $\lambda \equiv \langle \sigma v \rangle n_b$ is the interaction rate normalized to total baryon, and $\lambda_{pn}/\lambda_{DD} \approx 10^{-3}$ at $T \approx 0.07 - 0.08$ MeV. Fig. 5.3 also shows that the nucleosynthesis ends around this time so that the abundance of Triton, 3-Helium, and 4-Helium saturate and freeze out. Because the binding energy per nuclide of 4-Helium nuclei is highest among the light elements, 4-Helium is the energetically favored among all the light element formed during BBN. Therefore, once the nucleosynthesis starts, essentially all neutrons contribute to form 4-Helium to yield

$$Y_p \approx 2X_n(T_{\text{BBN}}),$$

(5.140)

where $T_{\text{BBN}}$ is the temperature at the beginning of the nucleosynthesis. One indication of the beginning of the nucleosynthesis is when the neutron concentration drops below the level that the $\beta$-decay predicts, or the Deuterium concentration peaks ($X_D \approx 0.01$), which happens at

$$T_{\text{BBN}} \approx 0.075 \text{ MeV},$$

(5.141)

the cosmic time corresponding to this temperature is

$$t_{\text{BBN}} \approx 225.6 \text{ sec}$$

(5.142)

after the Big-band, and the helium abundance is

$$Y_p \approx 2 \times 0.147617 e^{-225.6/880.3} \approx 0.23,$$

(5.143)

which is close to the magic number! The numerical integration gives $Y_p \approx 0.2474$.

The final abundance of light elements, in particular Helium and Deuterium, depends sensitively on the assumed value of the baryon number density ($\eta_{10}$) as shown in Fig. 5.4 which is the result of the full numerical calculation.

When increasing the baryon density but fix photon number density, the deuterium can form earlier because there are less photons per Deuterium that can break the Deuterium bound state. Then, there are more Deuterium available to start the nucleosynthesis earlier. That means, more neutrons (in Deuterium) are available (simply because they avoid $\beta$-decay) to form Helium. This increases the final Helium abundance. However, $Y_p$ may not exceed the maximum value $Y_p = 2X_n^\infty \approx 0.2952$ unless BBN happens before the neutron-to-proton ratio freeze out. In the other extreme case, if the baryon density is too low, the Deuterium formation process ($n + p \rightarrow D + \gamma$) had freeze out before the nucleosynthesis began (then neutrons $\beta$-decay into protons). Then, $X_n$ does not drop below $X_D$ and only fraction of neutrons end up processed into light elements. Therefore, the final Helium fraction is much less than the standard value.

The relic Deuterium abundance traces the freeze-out concentration. For higher $\eta_{10}$, once $X_n$ drops below $X_D$, essentially all neutrons go into the Deuterium, then $X_D$ is controlled by its depletion via D+D channel and D+p channel. Schematically, the rate of change of Deuterium concentration is given by

$$\frac{dX_D}{dT} \propto \eta_{10} (X_D^2 + RX_p X_D),$$

(5.144)

where $R \propto \lambda_{pd}/\lambda_{DD} \approx 10^{-5}$ is relative ratio between the interaction rate of DD-channel and Dp-channel. For small $\eta_{10} < 10$, the freeze-out concentration $X_D^f$ is larger than $RX_p$, then the first term (DD-channel) dominates to have $X_D^f \propto \eta_{10}^{-1}$ dependence. On the other hand, when $\eta_{10} > 10$, the Dp-channel dominates and we have $\ln X_D^f \propto \eta_{10}$ dependence on the baryon density. That is the reason why the relic Deuterium abundance is a sensitive probe of the baryon density.
5.6. HYDROGEN RECOMBINATION AND CMB

Nucleosynthesis calculation can also be altered if we have different fundamental parameters/ constants. As an example, we plot the dependence of \( Y_p \) on the additional relativistic degrees of freedom (parameterized in terms of effective number of neutrino species \( N_\nu \)) from \( N_\nu = 2 \) to \( N_\nu = 5 \). The effect here is changing \( g_\ast \) from the standard value 3.363 to \( g_\ast = 2.908, 3.817, 4.271 \) for \( N_\nu = 2, 4, 5 \), respectively. Then, because the cosmic time depends \( t \propto g_\ast^{-1/2} \) for a given temperature, adding (subtracting) relativistic degrees of freedom decreases (increases) time for neutron-to-protonation freeze-out and the beginning of BBN. Therefore, for larger (smaller) \( N_\nu \), A) the freeze-out concentration of neutrons (\( X_n^f \)) increases (decreases) and B) more (less) neutrons are available to form Helium nuclei that increases (decreases) \( Y_p \). This qualitative argument is consistent with what we see in the top panel of Fig. 5.4. You will quantify the change of Helium concentration under the change of various fundamental parameters in the homework.

5.6 Hydrogen recombination and CMB

After the Big-bang nucleosynthesis, besides the neutrinos that have been free-streaming after \( T \approx 1.5 \text{ MeV} \), the thermal bath of the Universe is filled with Helium nuclei, Hydrogen nuclei (protons), electrons, and photons. Next major event is the recombination of Helium that the Helium nuclei captures one electron to form a \( \text{He}^+ \) ion, then a \( \text{He}^+ \) ion captures another electron to form a He atom. The ionization energy of \( \text{He}^+ \) and He are, respectively, \( I_{\text{He}^+} = 54.4 \text{ eV} \) and \( I_{\text{He}} = 24.62 \text{ eV} \), greater than the ionization of Hydrogen, \( I_H = 13.6 \text{ eV} \) and this process must happen earlier than the hydrogen recombination. You will have some fun with Helium recombination in the homework.

5.6.1 Recombination in equilibrium

Our main interest in this section is the recombination of Hydrogen, the process that the free electrons and protons combines to form hydrogen atoms in the ground \( 1s \) state:

\[
p + e \rightleftharpoons \text{H}_{1s} + \gamma (13.6 \text{ eV}). \tag{5.145}
\]

At early times when the photon temperature is high, the reaction should be in an equilibrium due to the frequent interaction between protons and electrons. In this case, we can use chemical equilibrium condition

\[
\mu_e + \mu_p = \mu_H, \tag{5.146}
\]

or

\[
\left( \frac{n_e}{2(m_e T / 2\pi)^{3/2} e^{m_e/T}} \right) \left( \frac{n_p}{2(m_p T / 2\pi)^{3/2} e^{m_p/T}} \right) = \left( \frac{n_H}{4(m_H T / 2\pi)^{3/2} e^{m_H/T}} \right), \tag{5.147}
\]

then

\[
\frac{n_en_p}{n_H} = \left( \frac{m_p m_e}{m_H} \right)^{3/2} e^{(m_H-m_p-m_e)/T} \approx \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{-I_H/T}. \tag{5.148}
\]

Let us define the ionization fraction as

\[
X_e \equiv \frac{n_e}{n_e + n_H} = \frac{n_p}{n_p + n_H}, \tag{5.149}
\]

then the equation becomes

\[
\frac{X_e^2}{1-X_e} \simeq \frac{1}{n_p + n_H} \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{-I_H/T}, \tag{5.150}
\]
Figure 5.4: The mass fraction of various nuclides relevant in the Big-bang nucleosynthesis
which is called Saha equation. Because we assume Heliums are completely neutral,

\[ n_{\text{H,tot}} = n_p + n_\text{H} = (1 - Y_p)n_\text{H} \simeq 0.76 \times 10^{-10}\eta_{10} n_{\gamma} \simeq 3.124 \times 10^{-8}\eta_{10} \left(\frac{T}{2.726\text{K}}\right)^3 \text{cm}^3, \quad (5.151) \]

the right hand side becomes

\[ \frac{1}{(1 - Y_p)10^{-10}\eta_{10} 2\zeta(3)/\pi^2 T^3} \left(\frac{m_e T}{2\pi}\right)^{3/2} e^{-I_H/T} \simeq 2.498 \times 10^{16}\eta_{10}^{-1} \left(\frac{I_H}{T}\right)^{3/2} e^{-I_H/T}, \quad (5.152) \]

and the Saha equation becomes

\[ \frac{X_e^2}{1 - X_e} = \exp\left[37.76 - \frac{I_H}{T} + \frac{3}{2} \ln\left(\frac{I_H}{T}\right) - \ln\eta_{10}\right]. \quad (5.153) \]

When \( T \simeq I_H \simeq 157,820\text{K} \), the right hand side of the Saha equation is \( e^{30.474} \simeq 1.72 \times 10^{13} \), extremely large, which dictates \( X_e \simeq 1 \). The ionization fraction \( X_e \) decreases and the neutral fraction is of order unity when the right hand side of the Saha equation is order unity; that happens when

\[ 37.76 - \frac{I_H}{T} + \frac{3}{2} \ln\left(\frac{I_H}{T}\right) - \ln\eta_{10} \simeq 0 \rightarrow \frac{I_H}{T} = - \frac{3}{2} \ln\left(\frac{I_H}{T}\right) \simeq 37.76 - \ln\eta_{10} = 35.922, \quad (5.154) \]

and the temperature can be calculated from (again, using iteration)

\[ \frac{I_H}{T} = 35.922 + \frac{3}{2} \ln(35.922) \simeq 41.294, \quad (5.155) \]

which gives \( T \simeq 3822\text{K} \) at which \( X_e^{(\text{eq})} \simeq 0.618 \). Again, the reason why we have to wait long after the temperature reaches the ionization temperature until the formation of neutral Hydrogen is because the baryon-to-photon ratio is small. Even though the mean photon temperature is small, we have large number of photons at the tail of the Planck distribution that can ionize all the hydrogen atom. Below this temperature, the equilibrium number density drops rapidly. The equilibrium ionization fraction becomes half \( (X_e^{(\text{eq})} = 0.5) \) at

\[ T_{\text{rec}} \simeq 3737\text{K}, \quad (5.156) \]

but the ionization fraction at, say, \( T \simeq 3000\text{K} \) is already (using that \( X_e \ll 1 \)),

\[ X_e^{(\text{eq})}(T = 3000\text{K}) \simeq e^{-15.19} \simeq 2.53 \times 10^{-7}. \quad (5.157) \]

The redshift at the recombination is \( z_{\text{rec}} = T_{\text{rec}}/T_{\text{cmb}} - 1 \simeq 1370 \), and the cosmic time is \( t_{\text{rec}} \simeq 252,000\text{yrs} \) from the Big-bang.

### 5.6.2 Non-equilibrium process

The equilibrium calculation above is not what happens in our Universe, in particular after the ionization fraction drops significantly. Instead, the sharp drop of the number density of electrons and protons freezes out the recombination process, and we have to use full kinetic equation. Yes, as you can easily suspect at this point, we will solve the freeze-out process by using the Boltzmann equation for the electron number density:

\[ \frac{1}{a^3} \frac{d(a^3 n_e)}{dt} = \langle \sigma v \rangle \left( n_e^{(\text{eq})} n_p^{(\text{eq})} - n_\text{H} n_e - n_e n_p \right). \quad (5.158) \]
You might tempt to use the Saha equilibrium ratio in Eq. (5.148) into the equation and solve for it, but that is NOT the way. We need to include more physics here.

First of all, the recombination to the ground state will not read a net change of the ionization fraction. It is because the recombination photon with $E > 13.6$ eV will be emitted in this process. This photon, has a photo-ionization cross section of (at $E = 13.6$ eV)

$$a_{\nu_0} = \frac{2^9 \pi \alpha}{3e^4} \pi a_0^2 \simeq 6.30 \times 10^{-18} \text{cm}^2,$$

(5.159)

where $\alpha = 1/137$ is the fine structure constant, and $a_0 = 0.5292 \text{Å}$ is the Bohr radius. The total number density of hydrogens and protons at recombination temperature $T \simeq 3500$ is ($\Omega_b h^2 = 0.023$)

$$n_{H,\text{tot}} \simeq 416 \text{cm}^{-3},$$

which gives the mean-free-time of

$$\lambda_\gamma = \frac{1}{n_H a_{\nu_0} c} = 12720(1 - X_e)^{-1} \text{s}. \quad (5.160)$$

Therefore, the mean-free-time for the ionizing photon is much smaller than the Hubble time ($\sim 10^5$ yrs) as long as the neutron fraction $1 - X_e > 10^{-9}$ (which is almost always true for our case of interest). This means that the recombination photon emitted from the ground-state recombination always ionize another neutral hydrogen to yield no net change. Therefore, the recombination must be done first to higher-$n$ state then cascade down to the ground state. This process is called case-B (on-the-spot) approximation. The case-B recombination rate coefficient is

$$\alpha_B \equiv \langle \sigma v \rangle_B = \frac{4.309 \times 10^{-13} T_4^{-0.6166}}{1 + 0.6703 T_4^{0.5300}} \text{cm}^3/\text{s}, \quad (5.161)$$

where $T_4 \equiv T/(10^4 \text{K}).$

But, even after excluding the recombination to the ground state, recombination to the higher-$n$ level then cascade down will produce Lyman-series photons that have a large cross sections for captured by a Hydrogen atom. These photons excite adjacent Hydrogen atoms to the high-energy states so that it can be easily photoionized again. Therefore, both of them are not the main channel through which cosmic recombination has happened.

The two main routes to the recombinations are A) two-photon decay and B) redshifted Lyman-alpha photons. The details of the recombination calculation is presented in [? ,?] using the three-level approximation (that we will follow in this section later). This calculation is refined later (including effectively 300 atomic levels and improved treatment of Helium recombination—same issue is there for Helium that we have to use case-B approximation) by [?] which reads to the public code RecFAST. More recently, further refinement has been done for the calculation to less than 1% accuracy by heroic personals like Jens Chluba, Chris Hirata, etc, and that effort reads public code HyRec and CosmoRec. The codes calculate a detailed physical processes including other two-photon decays, recombination to higher levels (including some 500 shells), feedback from other Lymann lines, etc. You can find them in the following URL 10. As for the discussion here, we will use Peebles’s prescription.

**Peebles’s recombination : three-level approximation**

Here, we model the cosmic recombination history by using the three-level approximation: ground state with fraction $x_1$, $n = 2$ state with fraction $x_2$, and ionization state with $x_p = x_e$. We assume that the

---

10[HyRec: http://www.sns.ias.edu/ yacine/hyrec/hyrec.html](http://www.sns.ias.edu/ yacine/hyrec/hyrec.html)

Cosmological Recombination Project:

[http://www.cita.utoronto.ca/~jchluba/Science_Jens/Recombination/Welcome.html](http://www.cita.utoronto.ca/~jchluba/Science_Jens/Recombination/Welcome.html)
higher $n \geq 2$ energy level is populated according to the equilibrium ratio:

$$\frac{x_n}{x_2} = \frac{g_n e^{-(E_n - E_2)/T}}{g_2}. \tag{5.162}$$

Because the net decay into $1s$ state is very slow, we can write the rate of change of $n = 2$ level from the competition between case-B recombination and photo-ionization:

$$\frac{1}{a^3} \frac{d(n_2 a^3)}{dt} = [\alpha_B n_e n_p - \beta n_2]. \tag{5.163}$$

where we can figure out $\beta$ from the detailed balance as

$$\beta = \frac{\alpha_B n_e^{eq} n_p^{eq}}{n_2^{eq}} = \alpha_B \frac{g_e g_p}{g_2} \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{-\nu_H/4T} = \frac{\alpha_B}{4} \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{-\nu_H/4T}. \tag{5.164}$$

Here, we use $g_2 = 8 [\text{electron} : 2(1s^2) + 6(2p^6)] \times 2(\text{proton spin}) = 16$, then the equation becomes

$$\frac{dx_2}{dt} \bigg|_{n>2} = \alpha_B \left[ n_{H_{\text{tot}}} x_2^2 - \frac{1}{4} \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{-\nu_H/4T} x_2 \right]. \tag{5.165}$$

On top of the recombination and photoionization to $n = 2$, we have to include the two-photon decay and the redshifted Lyman-alpha lines that give the net transition from $n = 2$ to $n = 1$ state.

Two photon decay is a forbidden ($2s \rightarrow 1s$) dipole transition with the transition rate

$$\Lambda_{2\gamma} = 8.2206 s^{-1}. \tag{5.166}$$

Neither of the two photons have enough energy to excite the ground-state hydrogen; thus, this process reads to the net production of the ground state hydrogen atom. Rate of change of the $x_2$ fraction due to the two-photon decay is thus

$$\frac{dx_2}{dt} \bigg|_{2\gamma} = -\Lambda_{2\gamma} \left[ \frac{x_2}{4} - \beta_{2\gamma} x_1 \right], \tag{5.167}$$

where $\beta_{2\gamma}$ is the detailed-balance correction that makes sure that the total rate does not change in the equilibrium state. That is, from the Saha equation for $H(n = 1) + \gamma(3/4I_{\text{H}}) \leftrightarrow H(n = 2)$, we find

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} e^{-3I_{\text{H}}/4I} = 4 e^{-3I_{\text{H}}/4I} \tag{5.168}$$

that reads

$$\beta_{2\gamma} = e^{-3I_{\text{H}}/4I}, \tag{5.169}$$

and

$$\frac{dx_2}{dt} \bigg|_{2\gamma} = -\Lambda_{2\gamma} \left[ \frac{x_2}{4} - x_1 e^{-3I_{\text{H}}/4I} \right]. \tag{5.170}$$

Finally, we consider the redshfit of the Lyman-alpha line. The decay rate for the Lyman-alpha transition ($2p \rightarrow 1s$) is $A_{\text{Ly}\alpha} \approx 6.25 \times 10^8 s^{-1}$. Let us first calculate the optical depth of the Lyman-alpha line. Because the line width is narrow we estimate the cross section as

$$\sigma(v) = \sigma_0 \delta(v - v_{\text{Ly}\alpha}), \tag{5.171}$$
where
\[ \sigma_0 = 3\pi^2 A_{\text{Ly}\alpha}^\lambda \lambda_{\text{Ly}\alpha}^2, \]  
(5.172)
is the integral constant\(^{11}\).

Then, the optical depth of Lyman-alpha photons is
\[ \tau = \int n_1 \sigma d t = n_{H,\text{tot}} x_1 \int \sigma(v) / \nu \, d \nu, \]  
(5.177)
and we use the cosmological redshift \( \nu = -H \nu \) to calculate the optical depth as
\[ \tau = \frac{n_{H,\text{tot}} x_1}{H \nu_{\text{Ly}\alpha}} \sigma_0 = \frac{3\pi^2 A_{\text{Ly}\alpha}^\lambda n_{H,\text{tot}} x_1}{H \nu_{\text{Ly}\alpha}^3}. \]  
(5.178)

This is called the Sobolev optical depth that plays a key role in the line formation in expanding media. During recombination, the optical depth is typically of order \( \sim 10^8 \), and Lyman-alpha photons are immediately re-absorbed. The escape probability\(^{12}\) of the Lyman-alpha photons is then
\[ P = \frac{1 - e^{-\tau}}{\tau} \approx \frac{1}{\tau} \approx 10^{-8} \]  
(5.182)
which makes the final rate \( A_{\text{Ly}\alpha}^\lambda P \approx 1 \).

Then, the change of the \( x_2 \) due to Lyman-alpha emission is
\[ \frac{dx_2}{dt}_{\text{Ly}\alpha} = -\frac{3 A_{\text{Ly}\alpha}}{4 \tau} \left[ x_2 - 4x_1 e^{-3\nu_{\alpha}/4T} \right] = -\frac{H \nu_{\text{Ly}\alpha}^2}{4\pi^2 n_{H,\text{tot}} x_1} \left[ x_2 - 4x_1 e^{-3\nu_{\alpha}/4T} \right] \]  
(5.183)

\(^{11}\) This constant can be found from the detailed balance as following. Put a Hydrogen atom in a black-body, then in equilibrium we have
\[ A_{\text{Ly}\alpha}^1 + f(n) n_{2p} = n_{1s} \int_0^\infty \frac{d\nu}{\nu} \sigma(\nu) d \nu. \]  
(5.173)

Then, using
\[ f(\nu) = \frac{1}{e^{\nu/T} - 1} \]  
(5.174)
we find
\[ A_{\text{Ly}\alpha}^e \frac{\nu}{e^{\nu/T} - 1} n_{2p} = n_{1s} \sigma_0 \frac{\nu_{\text{Ly}\alpha}^2}{\pi^2 (e^{\nu_{\text{Ly}\alpha}/T} - 1)}, \]  
(5.175)
or
\[ \sigma_0 = \frac{\pi^2}{\nu_{\text{Ly}\alpha}^2} A_{\text{Ly}\alpha}^e e^{\nu_{\text{Ly}\alpha}/T} n_{2p} / n_{1s} = 3 \frac{\pi^2}{\nu_{\text{Ly}\alpha}^2} A_{\text{Ly}\alpha}. \]  
(5.176)

\(^{12}\) The escaping probability of a line profile \( P(\nu) d \nu \) emitted at time \( t \) can be written as
\[ P(t) = \int_{-\infty}^{\infty} d \nu \mathcal{P}(\nu) e^{-\tau(\nu,t)}, \]  
(5.179)
which is the profile-weighted \( e^{-\tau(\nu)} \), the probability that a photon emitted at frequency \( \nu \) will escape. When the photon is emitted to an expanding media, the optical depth is
\[ \tau(\nu,t) = -\int_\infty^t d t' \nu_{1s} \sigma(\nu a(t)/a(t')) = -\int_\nu^0 d \nu' \frac{\nu_{1s}}{\nu'} \sigma(\nu') = \frac{n_{1s}}{H \nu} \tau_0 \int_\nu^0 d \nu' \delta(\nu' - \nu_{\text{Ly}\alpha}) = -\tau \int_\nu^0 d \nu' \delta(\nu' - \nu_{\text{Ly}\alpha}). \]  
(5.180)
with \( \nu' = \nu a(t)/a(t') \) being the redshifted frequency. For the final equality, we assume that the emitted photons will be re-absorbed in a time much less than the Hubble time. Then, \( \nu' \approx -H \nu \). Putting all pieces together, with \( \mathcal{P}(\nu) = \delta(\nu - \nu_{\text{Ly}\alpha}) \) we calculate the escaping probability as
\[ P(t) = \int_{-\infty}^{\infty} d \nu \mathcal{P}(\nu - \nu_{\text{Ly}\alpha}) e^{-\tau} \int_t^\infty d \nu' \delta(\nu' - \nu_{\text{Ly}\alpha}) = \frac{1}{\tau} \int_0^\tau dx e^{-x} = \frac{1 - e^{-x}}{\tau}. \]  
(5.181)
Combining all three contributions, we find
\[
\frac{dx_2}{dt} = \alpha_B n_{H,\text{tot}} x_2^2 - \beta x_2 - (\Lambda_{2\gamma} + \Lambda_a) \left( \frac{x_2}{4} - x_1 e^{-3I_H/4T} \right).
\] (5.184)

Here, we define
\[
\Lambda_a \equiv \frac{H \nu_{\text{Ly}a}^3}{\pi^2 n_{H,\text{tot}} x_1}.
\] (5.185)

We assume that the excited Hydrogen atoms are in equilibrium (it makes sense as they are short-lived), then set \( \dot{x}_2 = 0 \) to find
\[
x_2 = 4 \frac{\alpha_B n_{H,\text{tot}} x_2^2 + (\Lambda_{2\gamma} + \Lambda_a) x_1 e^{-3I_H/4T}}{\Lambda_a + \Lambda_{2\gamma} + 4\beta}.
\] (5.186)

Feeding this into the net rate of loss of electron (note that two-photon decay and Lyman-\( \alpha \) re-absorption do not involve electrons), and using that \( x_2 \ll 1 \):
\[
\frac{dx_e}{dt} = -\alpha_B n_{H,\text{tot}} x_e^2 + \beta x_2 = -\alpha_B n_{H,\text{tot}} x_e^2 + 4\beta \left[ \frac{\alpha_B n_{H,\text{tot}} x_e^2 + (\Lambda_{2\gamma} + \Lambda_a) x_1 e^{-3I_H/4T}}{\Lambda_a + \Lambda_{2\gamma} + 4\beta} \right]
= -\left( \frac{\Lambda_a + \Lambda_{2\gamma}}{\Lambda_a + \Lambda_{2\gamma} + 4\beta} \right) \alpha_B n_{H,\text{tot}} x_e^2 - 4\beta x_1 e^{-3I_H/4T}
= -\left( \frac{\Lambda_a + \Lambda_{2\gamma}}{\Lambda_a + \Lambda_{2\gamma} + 4\beta} \right) \alpha_B n_{H,\text{tot}} x_e^2 - \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{-I_H/4T} (1 - x_e).
\] (5.187)

with
\[
\Lambda_{2\gamma} = 8.225 \text{s}^{-1}, \quad \Lambda_a = \frac{H \nu_{\text{Ly}a}^3}{\pi^2 n_{H,\text{tot}} (1 - x_e)}, \quad 4\beta = \alpha_B \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{-I_H/4T}
\] (5.188)

This equation is called **Peebles equation**! Note that the terms in the bracket satisfy the detailed balance because of the Saha equation Eq. (5.148). Had we ignored the delay of the recombination due to the inefficiency of two-photon decay and redshifted Lyman-alpha channels, we could have ended up with the equation without the prefactor, simply from the detailed balance and the case-B approximation. The prefactor can be interpreted as the **branching ratio** of two-photon decay and Lyman-alpha escape to total of two-photon decay, Lyman-alpha escape and photoionization from \( n = 2 \).

**time-temperature relation during recombination**

To solve the Peebles equation numerically, we need to supplement the time-temperature relation. The recombination happens close to the matter-radiation equality, the two main player here is the radiation density and the matter density. As usual, we start from the Friedmann equation (let's ignore the mass of neutrinos for this calculation):
\[
3H^2 = 8\pi G (\rho_m + \rho_R).
\] (5.189)

It is convenient to define the new time variable \( y = a/a_{\text{eq}} \), where \( a_{\text{eq}} = \Omega_R/\Omega_M \) is the scale factor at equality, to rewrite the Friedmann equation as
\[
3H^2 = 3 \left( \frac{1}{y} \frac{dy}{dt} \right)^2 = 4\pi G \rho_{\text{eq}}^0 \left( 1 + \frac{1}{y} \right),
\] (5.190)
where \( \rho^{(eq)} = 2 \rho_M^{(eq)} = 2 \rho_R^{(eq)} \) is the total energy density at equality. Then, we can solve the differential equation to have

\[
t = \left( \frac{4\pi G \rho^{(eq)}}{3} \right)^{-1/2} \left\{ \frac{2}{3} \left[(1 + y)^{3/2} + 2 \right] - 2(1 + y)^{1/2} \right\}.
\] (5.191)

Then, now let’s work out some numbers. From \( g_\star = 3.363 \) and \( T_{\text{cmb}} = 2.726 \text{ K} \), the radiation energy density now is

\[
\rho_R = \frac{\pi^2}{30} g_\star T_{\text{cmb}}^4 = 7.8177 \times 10^{-34} \text{ g/cm}^3
\] (5.192)

and the matter density now is

\[
\rho_M = \Omega_M \rho_{\text{crit}} = 1.8788 \times 10^{-29} \text{ g/cm}^3.
\] (5.193)

This gives the scale factor at equality as

\[
a_{eq} = 4.161 \times 10^{-5} \left( \Omega_M h^2 \right)^{-1},
\] (5.194)

the redshift at equality as

\[
1 + z_{eq} = 1/a_{eq} \approx 24033 \Omega_M h^2,
\] (5.195)

Then, the photon temperature at equality is

\[
T_{\gamma}^{(eq)} = T_{\text{cmb}}/a_{eq} \approx 65513 \Omega_M h^2 \text{ K} \approx 5.6456 \Omega_M h^2 \text{ eV},
\] (5.196)

and the energy density at equality is

\[
\rho^{(eq)} = 2\rho_M/a_{eq}^3 = 5.2158 \times 10^{-16} \left( \Omega_M h^2 \right)^4 \text{ g/cm}^3,
\] (5.197)

which makes the Hubble parameter at equality as

\[
H^{(eq)} = \sqrt{\frac{8\pi G}{3}} \rho^{(eq)} = \sqrt{\frac{16\pi G}{3} \Omega_M \rho_{\text{crit}} a_{eq}^{-3}} = H_0 \sqrt{\frac{2\Omega_M a_{eq}^{-3}}{2}}
\] (5.198)

The time equation Eq. (5.191) is then becomes

\[
t = \frac{2.625 \text{ kyr}}{(\Omega_M h^2)^2} \left\{ \frac{2}{3} \left[(1 + y)^{3/2} + 2 \right] - 2(1 + y)^{1/2} \right\},
\] (5.199)

and temperature is

\[
T_{\gamma} (y) = \frac{T_{\gamma}^{(eq)}}{y} \approx 65513 \Omega_M h^2 \text{ K} \approx 5.6456 \Omega_M h^2 \text{ eV}.
\] (5.200)

Finally, the redshift is

\[
1 + z = \frac{1}{a} = \frac{1}{ya_{eq}} = \frac{24033 \Omega_M h^2}{y},
\] (5.201)

and the Hubble parameter is

\[
H(y) = \frac{H^{(eq)}}{y^2} \sqrt{\frac{1 + y}{2}} \approx 7.357 \times 10^{-14} \text{ s}^{-1} \left( \frac{1 + y}{2} \right)^{1/2}.
\] (5.202)

For example, at the recombination temperature, \( T_{\text{rec}} = 3737 \text{ K} \), the redshift is \( z_{\text{rec}} = 1370 \), which gives (for \( \Omega_M h^2 = 0.143 \)), \( y_{\text{rec}} = 2.507 \). This gives the age of the Universe at recombination as \( t_{\text{rec}} = 252.41 \text{ kyr} \), and the Hubble parameter at the time as \( H(y_{\text{rec}}) = 7.357 \times 10^{-14} \text{ s}^{-1} = (430.74 \text{ kyr})^{-1} \).
Result: recombination history

Using \( y = a / a_{\text{eq}} \) as a time variable, we can rewrite the left-hand side of the Peebles equation as

\[
\frac{dX_e}{dt} = \frac{dX_e}{dy} \frac{dy}{dt} = \frac{dX_e}{dy} y H(y).
\] (5.203)

Then, the Peebles equation becomes

\[
\frac{dX_e}{dy} = -\left( \frac{\Lambda_\alpha + \Lambda_{2\gamma}}{\Lambda_\alpha + \Lambda_{2\gamma} + 4\beta} \right) \frac{\alpha_B}{y H(y)} \left[ n_{\text{H,tot}} X_e^2 e^{-\left( \frac{m_e T}{2\pi} \right)^{3/2} e^{-\frac{I_H}{T} \left( 1 - X_e \right)}} \right].
\] (5.204)

We listed all parameters for convenience:

\[
n_{\text{H,tot}} = 1.964 \times 10^{-7} \left( \frac{\Omega_b h^2}{0.023} \right) \left( \frac{T}{2.726 \text{K}} \right)^3 \text{cm}^3, \quad \alpha_B = \frac{4.309 \times 10^{-13} T^4 - 0.6166}{1 + 0.6703 T^{0.5300}} \text{cm}^3/\text{s},
\]

\[
\Lambda_{2\gamma} = 8.2206 \text{s}^{-1}, \quad \Lambda_\alpha = \frac{H \gamma_{\text{Ly} \alpha}^3}{\pi^2 n_{\text{H,tot}} (1 - X_e)}, \quad 4\beta = \alpha_B \left( \frac{m_e T}{2\pi} \right)^{3/2} e^{-\frac{I_H}{T} \left( 1 - X_e \right)}. \] (5.205)

One cautionary note. Because we use \( \hbar = 1 \) in our natural unit, we need to multiply \( \nu_{\text{Ly} \alpha} \) by \((2\pi)^3\):

\[
\Lambda_\alpha = \frac{8\pi H \gamma_{\text{Ly} \alpha}^3}{n_{\text{H,tot}} (1 - X_e)} = \frac{8\pi H}{n_{\text{H,tot}} \lambda_{\text{Ly} \alpha}^3 (1 - X_e)},
\] (5.206)

where \( \lambda_{\text{Ly} \alpha} = 1216 \text{Å} \). In other words, for \( \nu_{\text{Ly} \alpha} \), what we really meant was \( \omega_{\text{Ly} \alpha} = 2\pi \nu_{\text{Ly} \alpha} \).

We show the result of the numerical calculation in Fig. 5.5. As can be seen in Fig. 5.5, the recombination proceeds quite slowly compared to the other non-equilibrium processes. For the previous two cases, the the non-equilibrium solutions follow the equilibrium solution quite well before the freezeout, then freezeout happens to yield the deviation. But, in the case of cosmic recombination, the non-equilibrium solution already diverges from the equilibrium solution quite early in the process and the ionization fraction gradually decreases toward the residual value. Of course, the reason is because the recombination process is delayed due to the slowness of the relevant channels—two-photon decay and redshifted Lyman-alpha photons (with a rate of a couple happening in a second)—compared to the normal radiative processes \((A_{2p-1s} \approx 0.6 \text{ billion in a second}) \) would yield the thermal equilibrium. But, why isn’t the recombination delayed by the same factor? It is because the time scale that set the equilibrium density to drop is the rate at which the number density of ionizing photon \((E_\gamma > 13.6 \text{eV}) \) decreases due to the cosmological redshift, which is much slower than the radiative decay time scale.

Now, let us estimate the freezeout redshift and the residual electron number density from the usual method. The interaction rate for the case-B recombination is

\[
\Gamma_{ep} = n_e \alpha_B = n_{\text{H,tot}} X_e \alpha_B.
\] (5.207)

As the decoupling happens when \( X_e \ll 1 \), we approximate it as

\[
X_e(T_4) \approx \exp \left[ 17.96 - \frac{7.891}{T_4} + \frac{3}{4} \ln \left( \frac{15.782}{T_4} \right) \right],
\] (5.208)

and

\[
n_{\text{H,tot}} \approx 9695.36 T_4^3 \text{cm}^{-3},
\] (5.209)
Figure 5.5: Recombination history calculated from Saha equation (Eq. (5.148)), Peebles equation (Eq. (5.204)), Peebles equation with a fudge factor 1.14 to $\alpha_B$, and the full calculation from HyRec code. We use $\Omega_M h^2 = 0.143$, $\Omega_b h^2 = 0.023$. 
5.6. HYDROGEN RECOMBINATION AND CMB

to estimate the temperature at which the recombination rate is equal to the Hubble rate as

\[
\Gamma_{ep} = \frac{3.1133T_4^{1.6334}}{1.49187 + T_4^{0.533}}e^{-7.891/T_4} = H = 2.8131 \times 10^{-13}\sqrt{(0.9368 + T_4)T_4^3}
\]  

(5.210)

to find out

\[
T_{\text{rec}}^f \approx 2714 \text{ K},
\]

(5.211)

with the ionization fraction

\[
X_e^f \approx 3.129 \times 10^{-4}.
\]

(5.212)

Surprise! surprise! The ionization fraction from the full calculation at \( z \simeq 200 \) is \( 3.2696 \times 10^{-4} \)! It’s not too different from our crude guesstimation!!

The residual free electron plays very important role in the primordial chemistry, as it is the main catalyst of \( \text{H}^- \) ion that is needed to form a molecular hydrogen, which is an important coolant for forming the first generation of stars and galaxies.

5.6.3 Decoupling of photon

Because the ionization fraction drops in the course of the Hydrogen recombination, the light can finally decoupled from the electrons\(^{13}\)! When does it happens? The cross-section for light scattering off the electron at this energy scale is given by the Thomson scattering cross section:

\[
\sigma_T = \frac{8\pi}{3}r_0^2 = \frac{8\pi}{3} \frac{e^4}{m_e^2} \simeq 6.6524 \times 10^{-25} \text{ cm}^2.
\]

(5.213)

Then, the mean free path of the photon is

\[
\lambda = \frac{1}{n_e \sigma_T}.
\]

(5.214)

The mean free path is the mean path length that the photon can travel without being scattered once. Because the volume swiped by a incident particle with vertical (to the moving direction) cross section, \( \sigma \), after moving a distance \( x \) is \( V = x\sigma \), the expected mean number of electrons inside of that volume is \( n_e \sigma_T x \). As the mean number of electrons is the same as expected number of scattering, the distance at which one scattering is expected on average is \( \lambda = (n_e \sigma_T)^{-1} \). With the same logic, the probability of photon being scattered in the path length between \( x \) and \( x + \delta x \) is

\[
P(x \sim x + \delta x) = \frac{\delta x}{\lambda(x)},
\]

(5.215)

which gives the scattering rate of

\[
\Gamma_{\text{scat.}} = \frac{P(x \sim x + \delta x)}{\delta x/c} = n_e \sigma_T c
\]

(5.216)

that we have been using so far. The decoupling happens when the scattering rate is the same as the expansion rate:

\[
n_e \sigma_T c = X_e n_{\text{H},\text{tot}} \sigma_T c = H(T),
\]

(5.217)

\(^{13}\)Note that we ignore the photon-proton scattering because the cross section is far much smaller, suppressed by the inverse-mass-square ratio (\( \sim 10^{-7} \)), for this process.
or

\[
X_e(T) = \frac{H(T)}{n_{H,\text{tot}} \sigma_T c} = \frac{2.8131 \times 10^{-13} \sqrt{(0.9368 + T_4)T_4^3}}{9695.36 T_4^3 \times (6.6524 \times 10^{-25}) \times (2.9979 \times 10^{10})} = 1.4549 \times 10^{-3}. \tag{5.218}
\]

Using the equilibrium ionization fraction, I find that the decoupling happens at

\[
T_{\text{dec}}^{(\text{est})} = 3086 \text{ K}, \quad z_{\text{dec}}^{(\text{est})} = 1131, \tag{5.219}
\]

with \(X_{e,\text{dec}}^{(\text{est})} = 0.0095\). The result from the full calculation is

\[
T_{\text{dec}} = 2461 \text{ K}, \quad z_{\text{dec}} = 901.82, \tag{5.220}
\]

and \(X_{e,\text{dec}} = 0.013\).

What we have calculated above was the time at which the scattering rate is the same as the Hubble rate. The more rigorous way of defining the decoupling is through the visibility function of photon. The visibility function of photon is the probability distribution of photon's last scattering redshift. That is, what we have to calculate is the probability that the photon had last-scattered at \(z_s\), but has been freely streaming to the observer at \(z = 0\) ever since (ignoring for the reionization for a moment).

Let's go back to the path-length argument. The probability that the photon has not scattered until \(x\), and does not scatter inbetween the path length \(x\) and \(x + \delta x\) is

\[
P_{\text{no}}(x + \delta x) = P_{\text{no}}(x) \left[ 1 - \frac{\delta x}{\lambda(x)} \right]. \tag{5.221}
\]

The equation above can be translated into the differential equation:

\[
\frac{d \ln P_{\text{no}}(x)}{dx} = -\frac{1}{\lambda(x)} = -n_e \sigma_T, \tag{5.222}
\]

from which we calculate the probability of no scattering from \(z = 0\) to all the way to the redshift \(z_s\) (path length \(x\)) as

\[
P_{\text{no}}(x) = e^{-\tau(x)} \tag{5.223}
\]

where

\[
\tau(x) = \int_0^x n_e(x') \sigma_T dx'. \tag{5.224}
\]

is the optical depth of the photon. Probability that the photon had scattered at least one times between \(x\) and now is

\[
P_{>1}(x) = 1 - P_{\text{no}}(x), \tag{5.225}
\]

which is called opacity. The probability that the last-scattering happens further away than \(x + \delta x\) is

\[
P_{ls}(> x + \delta x) = P_{\text{no}}(x + \delta x), \tag{5.226}
\]

and the probability that the last-scattering happens closer than \(x\) is

\[
P_{ls}(< x) = 1 - P_{\text{no}}(x). \tag{5.227}
\]
Combining the two, the probability that the last-scattering happens between \( x \) and \( x + \delta x \) is

\[
P_{ls}(x)dx = P_{ls}(< x + dx) - P_{ls}(< x) = -P_{no}(x + \delta x) + P_{no}(x),
\]

so that

\[
P_{ls}(x) = -\frac{dP_{no}(x)}{dx} = \frac{d\tau}{dx} e^{-\tau(x)} = n_e \sigma_T e^{-\tau(x)}. \tag{5.229}
\]

The probability distribution function \( P_{ls}(x) \) of the last-scattering is the desired visibility function. More often, the visibility function is defined as the last-scattering redshift instead of the optical path length \( x \) (physical distance that each photon travels). That is,

\[
g(z) = \frac{d\tau}{dz} e^{-\tau(z)} = \frac{\sigma_T n_{H,\text{tot}}(z) X_e(z)}{(1+z)H(z)} \exp \left[ -\sigma_T \int_0^z \frac{rn_{H,\text{tot}}(z')}{(1+z')H(z')} dz' \right]. \tag{5.230}
\]

We show the visibility function in Fig. 5.6. The visibility function peaks at around \( z \approx 1077 \), which means that the most of photons were last-scattered off electrons at that redshift. The isotropic sphere around us with radius \( \chi(z = 1077) \) therefore forms a last-scattering surface of photon after which the photons free-stream without being scattered.

The photons at higher redshift underwent frequent scattering so that the primordial fluctuations on scales smaller than diffusion scale got equalized and erased. At redshift \( z > 2 \times 10^6 \), the double Compton scattering and Bremsstrahlung process is efficient to generate new entropy (photon number) out of the erased fluctuations, but below that redshift the photon number density is conserved. Conserved number density, but changing energy (from the small scale fluctuation) will induce the spectral distortion of the observed CMB spectrum. Currently, spectral distortion is limited to less than \( 10^{-5} \) level, but the limit should be improved by future CMB mission with many frequency channels such as PIXIE.

### 5.6.4 Thermal decoupling of matter from radiation: Compton heating

We have discussed the photon’s last-scattering off the electron in the previous section, and show that most of photons have decoupled from electron at \( z \approx 1100 \). Did electrons set free about the same time? The answer is no! It is because there are so many photons per an electron.

The simplest way of looking at this problem is to calculate the mean free path. For photon, mean free path is given by \( \lambda_\gamma = (n_e \sigma_T)^{-1} \) and for electron it is \( \lambda_e = (n_\gamma \sigma_T)^{-1} \), therefore,

\[
\frac{\lambda_\gamma}{\lambda_e} = \frac{n_e}{n_\gamma} = (1 - Y_p) \frac{n_b}{n_\gamma} X_e = 7.6 \times 10^{-11} \eta_{10} X_e. \tag{5.231}
\]

That is, the electron mean-free path is MUCH smaller compared to the photon’s mean-free path. Well, photons do NOT care for one scattering per ten billion, but electrons certainly do.

This frequent scattering of electron off the photon result energy exchange between electrons and photons. In a adiabatic expansion, the temperature of electron must scale as \( T_e \propto 1/a^2 \). But, as a result of the tight thermal coupling due to energy exchange (via Compton scattering), the electron temperature scales as the same way as the photon temperature \( T_e \propto 1/a! \)

Let’s calculate when does that thermal coupling stop. The average energy transfer per Compton scattering (Compton heating) is

\[
\Delta \epsilon_b = \frac{4}{3} \beta^2 \langle \epsilon_\gamma \rangle = 4 \left( \frac{k_b T_e}{m_e c^2} \right) \frac{\rho_\gamma}{n_\gamma}.
\]

\[
\tag{5.232}
\]
Figure 5.6: The visibility function of the Cosmic Microwave Background photons. Most of photons were last-scattered off electrons at $z \simeq 1077$ for $\Omega_M h^2 = 0.143$, $\Omega_b h^2 = 0.023$. Note that the wing at the lower redshift side is broader due to the delayed recombination.
where we use the equipartition theorem \(3/2k_B T = 1/2mv^2\), and \(\rho_\gamma = n_\gamma \langle \epsilon_\gamma \rangle\). Note that the energy that electron gains from photon is shared by all baryons. Then, the rate of change of baryonic energy density is given by

\[
\frac{dE_b}{dt} = n_e n_\gamma \sigma_T c \Delta \epsilon_b = 4n_e \sigma_T c \rho_\gamma \left( \frac{k_B T_e}{m_e c^2} \right),
\]

from which we can read off the time scale for the thermal energy exchange:

\[
t_{\text{Compton}}^{-1} = \frac{1}{E_b} \frac{dE_b}{dt} = 4n_e \sigma_T c \rho_\gamma \left( \frac{k_B T_e}{m_e c^2} \right).
\]

Using \(E_b = 3/2(n_b + n_e)k_B T_e\), we find

\[
t_{\text{Compton}}^{-1} = \frac{8}{3} \frac{X_e}{(1 - Y_p)^{-1} + X_e} \left( \frac{\rho_\gamma \sigma_T c}{m_e c^2} \right).
\]

Let's work out for the numbers (in the natural units).

\[
\rho_\gamma = \frac{\pi^2}{15} T_\gamma^4 = 13216.5 (1 + z)^4 \text{ cm}^{-4}
\]

\[
m_e c^2 = 2.5897 \times 10^{10} \text{ cm}^{-1},
\]

from which we find

\[
t_{\text{Compton}}^{-1} = \frac{2.714 \times 10^{-20} X_e (1 + z)^4}{1.3158 + X_e}.
\]

We have to equate this with Hubble rate at the matter domination:

\[
H(z) = 1.22546 \times 10^{-18} (1 + z)^{3/2},
\]

with the freezeout value \(X_e \simeq 3 \times 10^{-4}\), we find that the baryons thermally decouple from photon at

\[
(1 + z) = 4.6 \left( \frac{1.316 + X_e}{X_e} \right)^{2/5} \simeq 131.72,
\]

or \(z_{\text{dec}} \simeq 130!\) The actual calculation using \(X_e(z)\) gives \(z_{\text{dec}} \simeq 134\), which agree pretty well with our estimation! The temperature of the baryons were the same as the temperature of the photons before that redshift. After the decoupling redshift, temperature of baryons drops well below the photon temperature (with different power-law index). That is, CMB has been influencing the evolution of the Universe until that redshift.