Week 2: Thermal History of the Universe

January 17, 2017

1 Thermodynamics in the Expanding Universe

As discovered by Penzias and Wilson in 1965, and determined much more precisely in the early 1990s by the *Far Infrared Absolute Spectrometer (FIRAS)* on NASA's Cosmic Background Explorer (COBE) satellite, the Universe is filled with a gas of microwave photons with a blackbody spectrum and a temperature $T_0 = 2.7$ K, or $T = 2.4 \times 10^{-13}$ GeV in units in which the Boltzmann constant $k_B = 1$. In other words, in every direction we look, we see a specific intensity (ergs/cm^2/sec/steradian/Hz),

$$I_\nu = \frac{2h\nu^3/c^2}{e^{h\nu/kT_0} - 1}, \tag{1}$$

as a function of frequency ν. It is important to keep in mind that although this is the energy distribution that a gas of massless particles has in thermal equilibrium, the photons in this gas are most certainly *not* in thermal equilibrium. Thermal equilibrium implies that the energy-momentum distribution of the particles in the thermal bath is maintained by frequent collisions. These photons have extremely long mean-free paths through the Universe, comparable to the size of the observable Universe, and thus are not in thermal equilibrium, even though they have the energy distribution characteristic of thermal equilibrium. We refer to this gas of photons as the *cosmic microwave background (CMB)*.

Interestingly enough, since the frequency of each photon scales as $1 + z$ with redshift z—$\nu(z) = (1 + z)\nu_0$—the frequency spectrum of the CMB always maintains a blackbody distribution, albeit one with a temperature $T(z) = T_0(1+z)$—i.e., the earlier Universe was hotter. Although they have not yet been observed, we will also see that the Universe also constains a gas of neutrinos (all three mass eigenstates) at a temperature $T_\nu = 1.96$ K. We also measure, through a variety of techniques that we will discuss in a bit, a baryon (i.e., neutrons and protons) density (in units of critial), $\Omega_b = \rho_b/\rho_c \simeq 0.019\, h^{-2}$. (And keep in mind that $\rho_c = 10^{-5}\, h^2\, \text{GeV cm}^{-3} = 1.9\, h^2 \times 10^{-29}\, \text{g cm}^{-3}$.)

Needless to say, although the Universe is quite cool and diffuse today, it must have been hotter and denser at earlier times. For example, at a redshift $z = 10^9$, the density of the Universe approached that of water, and the temperature somewhere around an MeV. It is thus reasonable to surmise that at some sufficiently early time, the contents of the Universe must have been described by a gas of elementary particles in thermal equilibrium, rather than a distribution of far-flung galaxies that interact only gravitationally. We therefore must recall some thermodynamics to describe the early Universe.

2 Review of relevant statistical mechanics and thermodynamics

In the following, we will use particle-physics units ($\hbar = c = 1 = k_B = 1$). Then a dilute gas of weakly interacting particles has a number density,

$$n = \frac{g}{(2\pi)^3} \int f(\vec{p})\, d^3p, \tag{2}$$

an energy density,

$$\rho = \frac{g}{(2\pi)^3} \int f(\vec{p})\, E(\vec{p})\, d^3p, \tag{3}$$

and pressure,

$$P = \frac{g}{(2\pi)^3} \int f(\vec{p})\, \frac{|\vec{p}|^2}{3E}\, d^3p, \tag{4}$$

where

$$f(\vec{p}) = \left[\exp\left(\frac{E-\mu}{T}\right) \pm 1\right]^{-1}, \tag{5}$$

is the distribution function for a gas of particles in thermal equilibrium, and the plus (minus) is for Fermi-Dirac (Bose-Einstein) statistics, and g is the degeneracy factor. The chemical potentials μ_i for particle species i that undergo the reactions $i + j \leftrightarrow k + l$ is $\mu_i + \mu_j = \mu_k + \mu_l$ in chemical equilibrium. The energy E of a particle of mass m and momentum p is $E(p) = (p^2 + m^2)^{1/2}$. At very high temperatures ($T \gg m$ and $T \gg \mu$),

$$\rho = \begin{cases} (\pi^2/30) g T^4 & \text{bosons} \\ (7/8)(\pi^2/30) g T^4 & \text{fermions} \end{cases}, \tag{6}$$

$$n = \begin{cases} [\zeta(3)/\pi^2] g T^3 & \text{bosons} \\ (3/4)[\zeta(3)/\pi^2] g T^3 & \text{fermions} \end{cases}, \tag{7}$$

where $\zeta(3) \simeq 1.2$ is the Riemann zeta function, and

$$P = \frac{1}{3}\rho. \tag{8}$$

For degenerate fermions ($\mu \gg T$) in the relativistic limit (this limit is important for white dwarfs (near the high-mass end) and neutrons stars, but not so much for cosmology),

$$\rho = \frac{g\mu^4}{8\pi^2}, \qquad n = \frac{g\mu^3}{6\pi^2}, \qquad P = \frac{g\mu^4}{24\pi^2}, \tag{9}$$

and in the nonrelativistic limit ($m \gg T$),

$$n = g\left(\frac{mT}{2\pi}\right)^{2/3} e^{-(m-\mu)/T}, \qquad \rho = mn, \qquad P = nT \ll \rho. \tag{10}$$

For $T \gg m$ and $T \gg \mu$, the mean particle energy is

$$\langle E \rangle \equiv \frac{\rho}{n} = \begin{cases} [\pi^4/30\zeta(3)]T \simeq 2.701\,T & \text{bosons} \\ \frac{7\pi^4}{180\zeta(3)}T \simeq 3.151\,T & \text{fermions} \end{cases} \tag{11}$$

Degenerate relativistic fermions have mean particle energy $\langle E \rangle \equiv \rho/n = (3/4)\mu$. For a nonrelativistic gas, $\langle E \rangle = m + (3/2)T \simeq m$.

3 Particle-antiparticle balance

The possibilty for, e.g., brehmsstrahlung reactions ($e^- + p \leftrightarrow e^- + p + \gamma$), implies that the photon has zero chemical potential, $\mu_\gamma = 0$ in chemical equilibrium. If so, then since a charged particle X^+ and its antiparticle X^- can annihilate to photons, $X^+X^- \leftrightarrow \gamma\gamma$, we must have $\mu_+ = -\mu_-$. If so, then the formula for n above yields for fermions a particle-antiparticle asymmetry (in terms of the chemical potential μ_+,

$$n_+ - n_- = \begin{cases} (gT^3/6\pi^2)\left[\pi^2(\mu/T) + (\mu/T)^3\right] & T \gg m, \\ 2g(mT/2\pi)^{3/2}\sinh(\mu_+/T)e^{-m/T} & T \ll m, \end{cases} \tag{12}$$

The total energy density and pressure are the sum of the contributions from each particle species in the thermal bath. If there is no particle-antiparticle asymmetry (or no significant particle-antiparticle asymmetry), then the density and pressure contributed by nonrelativistic particles is exponentially suppressed. In this case, the pressure and energy density are dominated by the relativistic particles, and we arrive at a radiation energy density,

$$\rho_R = (\pi^2/30)g_* T^4, \qquad P_R = \rho_R/3, \tag{13}$$

where

$$g_* \equiv \sum_{i=\text{bosons}} g_i(T_i/T)^4 + \frac{7}{8}\sum_{i=\text{fermions}} g_i(T_i/T)^4, \tag{14}$$

is the effective number of relativistic degrees of freedom (and the sum is taken only over relativistic species, $m_i \ll T$). For example, today, g_* receives contributions from the two degrees of freedom of the photon plus 3×2 neutrino degrees of freedom with $T_\nu = (4/11)^{1/3}T_\gamma$, so $g_*(T \ll \text{MeV}) = 2 + (7/8)6(4/11)^{4/3} \simeq 3.36$. For temperatures MeV$\ll T \ll 100$ MeV, $T_\nu = T_\gamma$ (as we will see), and the electron is relativistic ($m_e \ll T$), so $g_* = 2 + 6(7/8) + 4(7/8) = 10.75$. At higher temperatures, quarks contribute, as do muons, τ leptons, and at even higher temperatures, the electroweak gauge bosons $W^\pm$ and Z^0 and Higgs boson(s). For the standard electroweak model (with one Higgs doublet), $g_* = 106.75$ at $T \gg 100$ GeV.

During radiation domination ($z \gtrsim 22{,}000\,\Omega_m h^2$ ($t \lesssim 10^5$ yr), $\rho \simeq \rho_R$, $p \simeq \rho/3$, and the scale factor $a(t) \propto t^{1/2}$. At these times, $H = (8\pi G\rho_R/3)^{1/2} = 1.66\,g_*^{1/2}T^2/m_{\text{Pl}}$, where $m_{\text{Pl}} = G^{-1/2} = 1.2 \times 10^{19}$ GeV is the Planck mass. Under the approximation that $g_* \simeq$constant, the age of the Universe during radiation domination is $t \simeq (2H)^{-1} \simeq 0.301\,g_*^{-1/2}m_{\text{Pl}}/T^2 \simeq (T/\text{MeV})^{-2}$ sec. Therefore, the Universe was roughly one second old when the temperature was an MeV, and it scales roughly as T^{-2}.

4 Entropy

When the expansion timescale of the Universe is long compared with the timescales for reactions that maintain thermal equilibrium, then the cosmological gas can be considered to be in thermal equilibrium, but undergoing adiabatic changes with the slow expansion. Under these conditions, the entropy per comoving volume will be constant. Returning to the second law of thermodynamics, we have

$$TdS = d(\rho V) + PdV = d[(\rho + P)V] - VdP. \tag{15}$$

We then note that if $\mu = 0$ [so that the T dependence of f appears only in the combination E/T, then it can be shown (from the expression for P in terms of $f(E)$) that $dP = (\rho + P)(dT/T)$, so

$$dS = \frac{1}{T}d[(\rho + P)V] - (\rho + P)V\frac{dT}{T^2} = d\left[\frac{(\rho + P)V}{T} + \text{constant}\right]. \tag{16}$$

Therefore, the entropy per comoving volume is $S = a^3(\rho + P)/T$. But if the expansion is adiabatic, then $dS = 0$, and so the *entropy density* $s \equiv S/V = (\rho+p)/T \propto a^{-3}$ as long as local thermodynamic equilibrium is maintained. What this means is that if the energy density is dominated by radiation, then $s \propto a^{-3}$, where $s = (2\pi^2/45)g_{*s}T^3$, and

$$g_{*s} \equiv \sum_{i=\text{bosons}} g_i(T_i/T)^3 + (7/8) \sum_{i=\text{fermions}} g_i(T_i/T)^3, \tag{17}$$

and as before, the sum is taken only over relativistic species, $m_i \ll T$. Note that g_{*s} is almost always equal to g_*, and differs only if some species becomes thermally decoupled from the rest of the plasma. It is important to note that since $s \propto a^{-3}$, the temperature T is $T \propto a^{-1}$ only as long as g_{*s} remains constant. When the temperature T drops below the mass m_i of some particle in the thermal bath, then the temperature T drops a little more slowly with the expansion than $1/a$ as g_{*s} decreases. Thus, for example, when the temperature drops below m_e, electrons and positrons annihilate to produce additional photons. It is often misstated that electron-positron "heats" the photons. This is not true. What happens is that when electrons and positrons annihilate, they transfer their entropy to the photons, and this simply slows the temperature drop relative to $1/a$. Note that $s = 1.80\, g_{*s}n_\gamma$, and today, the photon number density is $n_{\gamma 0} \simeq 411\,\text{cm}^{-3}$, and $s_0 = 7.04\, n_\gamma \simeq 3000\,\text{cm}^{-3}$.

Since sa^3 is constant, we can use s to mark comoving volumes, and also the relation between scale factor a and temperature T: i.e., $g_{*s}T^3a^3$ =constant. Thus, for example, if baryon number is conserved, then $n_B/s \equiv (n_b - n_{\bar{b}})/s$ is the same at all times. It is common to hear people speak of the baryon-to-photon ratio $\eta \equiv (n_B/n_\gamma 0)_0 = 1.8\, g_{*s}(n_B/s)$, but this is not constant (if g_{*s} changes). In particular, the photon number, $N_\gamma = a^3 n_\gamma$ increases by 11/4 at e^+e^- annihilation near $T \sim 0.5$ MeV, since $g_{*s} = 2 + 4(7/8)$ before and $g_{*s} = 2$ afterwards. This is also why the neutrino temperature is $(4/11)^{1/3}$ relative to that of the photons. Neutrinos decouple at a temperature $T \sim MeV$, *before* e^+e^- recombination. Thus, while the (decoupled) neutrino temperature is falling as $T_\nu \propto a^{-1}$, the photon temperature is dropping as $T_\gamma \propto g_{*s}^{-1/3}a^{-1}$.

Note finally that since the momentum p of a massive particle decreases as $p \propto a^{-1}$, while its kinetic energy is $E = p^2/2m$ a massive decoupled species maintains a thermal energy distribution, but with a temperature $T \propto a^{-2}$. And, of course, its density decreases as a^{-3}. If a massive species decouples when it is relativistic ($m_D \ll T$), it does *not* maintain a thermal energy distribution when the temperature drops to $T \lesssim m$. Thus, for example, neutrinos decouple when their temperature is $T \sim$ MeV, and their temperature today is $T_{\nu 0} \sim 2 \times 10^{-4}$ eV. If they have a mass $m_\nu \gtrsim T_{\nu 0}$, then they are moving nonrelativistically today, and their energy distribution is *not* a Fermi-Dirac distribution.

5 Brief Overview

Before moving on, let's review some of the high points in the history of the Universe:

- $T \sim 10^{-4}$ eV, $t \sim 10^{10}$ yr: today; baryons and the CMB are entirely decoupled, stars and galaxies have been around a long time, and clusters of galaxies (gravitationally bound systems of $\sim$ 1000s of galaxies) are becoming common.
- $T \sim 10^{-3}$ eV, $t \sim 10^9$ yr; the first stars and (small) galaxies begin to form.
- $T \sim 10^{-1-2}$ eV, $t \sim$ millions of years: *baryon drag* ends; at earlier times, the baryons are still coupled to the CMB photons, and so perturbations in the baryon cannot grow; before this time, the gas distribution in the Universe remains smooth.
- $T \sim$eV, $t \sim 400,000$ yr: electrons and protons combine to form hydrogen atoms, and CMB photons decouple; at earlier times, photons are *tightly coupled* to the baryon fluid through rapid Thomson scattering from free electrons.
- $T \sim 3$ eV, $t \sim 10^{-(4-5)}$ yrs: matter-radiation equality; at earlier times, the energy density of the Universe is dominated by radiation. Nothing really spectacular happens at this time, although perturbations in the dark-matter density can begin to grow, providing seed gravitational-potential wells for what will later become the dark-matter halos that house galaxies and clusters of galaxies.
- $T \sim$keV, $t \sim 10^5$ sec; photons fall out of chemical equilibrium. At earlier times, interactions that can change the photon number occur rapidly compared with the expansion rate; afterwards, these reactions freeze out and CMB photons are neither created nor destroyed.
- $T \sim 10 - 0.1$ MeV, $t \sim$seconds-minutes. *Big-bang nucleosynthesis* (BBN). Neutrons and protons first combine to form D, ^{4}He, ^{3}He, and ^{7}Li nuclei. Quite remarkably, the theory for this is very well developed and agrees very impressively with a variety of observations.
- $T \sim 100 - 300$ MeV, $t \sim 10^{-5}$ sec. The quark-hadron phase transition. This is when quarks and gluons first become bound into neutrons and protons. We are pretty sure that this must have happened, but the details are not well understood, and there are still no signatures of this phase transition that have been observed. This is also when axions are produced, if they exist and are the dark matter.
- $T \sim$ 10s$-$100s GeV, $t \sim 10^{-8}$ sec. If dark matter is composed of supersymmetric particles, or some other *weakly interacting massive particle (WIMP)*, then this is when their interactions freeze out and their cosmological abundance fixed. This is speculative.

- $T \sim 10^{12}$ GeV, $t \sim 10^{-30}$ sec. The Peccei-Quinn phase transition, if the Peccei-Quinn mechanism is the correct explanation for the strong-CP problem. This is highly speculative.
- $T \sim 10^{16}$ GeV, $t \sim 10^{-38}$ sec. This is when the GUT (grand-unified theory) phase transition occurs. At earlier times, the strong and electroweak interactions are indistinguishable. This is highly speculative.
- $T \sim 10^{19}$ GeV, $t \sim 10^{-43}$ sec. Strings? quantum gravity? quantum birth of the Universe? This is all *highly* speculative. Basically, at these temperatures, the energy densities are so high that the classical treatment of general relativity is no longer reliable. Such early times can only be understood with a quantum theory of gravity, which we don't have.

6 Thermal History of the Universe

Thermal equilibrium of particle species in the early Universe is maintained by scattering of particles. Thus, equilibrium is maintained (and the expansion of the Universe can be considered to be adiabatic) as long as the interaction rate $\Gamma = n\sigma v$ is greater than the expansion rate, $H \sim T^2/m_{\rm Pl}$. Here, n is the number density of (other) particles that the particle in question can scatter from, and σ is the cross section for the scattering process in question, and v the relative velocity between the reacting particles. At this point, it is important to distinguish between *statistical equilibrium* and *chemical equilibrium*. In *chemical equilibrium*, the reactions that create and destroy the particles in question occur faster than the expansion, and so the chemical potential retains its equilibrium value (as determined by, e.g., $\mu_i + \mu_j = \mu_k + \mu_l$ for reactions $i + j \leftrightarrow k + l$). In *statistical equilibrium*, the *elastic* scattering reactions that maintain the thermal *energy* distribution of particles occur faster than the expansion. It is possible for a particular particle species to be in statistical equilibrium, but not in chemical equilibrium. I think you may be able to prove that if chemical equilibrium is maintained, then so is statistical equilibrium, but I'm not entirely sure. True *thermal equilibrium* occurs when the particle species in question is in statistical and chemical equilibrium.

Suppose now that the interaction rate Γ scales with temperature as $\Gamma \propto T^n$. Then, the number of interactions the particle will undergo after some time t is

$$N_{\rm int} = \int_t^\infty \Gamma(t')\, dt' = \frac{\Gamma(T)}{2H(T)} \int_0^T \left(\frac{T'}{T}\right)^n \frac{dT'}{T'^3} T^2 = \left(\frac{\Gamma}{H}\right)_t \frac{1}{n-2}. \tag{18}$$

Therefore, for $n > 2$ particles interact $\lesssim 1$ time after the time when $\Gamma \simeq H$. This is simply because during radiation domination, $H \propto T^2$, so if $n > 2$, the scattering rate Γ falls below H after $\Gamma = H$, and the interactions of the particles are said to *freeze out*.

Let's start with neutrino decoupling. At temperatures $T \gtrsim m_e \simeq 0.5$ MeV, electrons and positrons are abundant in the Universe, and chemical equilibrium of neutrinos is maintained by $\nu\bar{\nu} \leftrightarrow e^+e^-$ reactions through a virtual Z^0 boson and additionally (for the electron neutrino) through $W^\pm$ exchange, and statistical equilibrium is additionally maintained by neutrino-electron elastic scattering, which occurs through $W^\pm$ exchange (for electron neutrinos) and Z^0 exchange (for all three neutrinos).

The cross section for these processes can be calculated using standard weak-interaction theory, which is beyond the scope of this course. The result for the cross sections turns out to be $\sigma \sim (\alpha^2/m_W^4)E_\nu^2$, where $\alpha \sim 10^{-2}$ is the fine-structure constant, $m_{W,Z} \sim 100$ GeV is the gauge-boson mass, and E_ν the neutrino energy. The α^2 factor arises because there are two vertices in the matrix element (the Feynman diagram), and the cross section is proportional to the matrix element squared. The $m_{W,Z}^{-4}$ arises because of a factor $m_{W,Z}^2$ appears in the propagator for the virtual gauge boson that is exchanged. We then add the factor E_ν^2 to get the right dimensions [(energy)$^{-2}$] for the cross section. Noting that the abundance of electrons (when they are still relativistic is $n \sim T^3$, that neutrinos travel at velocities $v \simeq c$, and that the typical neutrino energy is $E_\nu \sim T$, the interaction rate for neutrino conversion to electrons and for neutrino-electron elastic scattering is $\Gamma_{\rm int} \sim \alpha^2 T^5/m_{W,Z}^4$, and the ratio to the expansion rate is $(\Gamma_{\rm int}/H) \sim \alpha^2 T^3 m_{\rm Pl}/m_{W,Z}^4 \sim (T/{\rm MeV})^3$. Therefore, for temperatures $T \gtrsim$MeV, neutrinos are in thermal equilibrium, and they decouple at temperatures $T \sim$MeV. Of course, all this has been pretty fast and loose, but it turns out that when you put all the factors of π, 2, etc. in the right place, you get this result. We then note that these neutrinos then free stream throughout the remaining history of the Universe, and their temperature today is $(4/11)^{1/3}$ the photon temperature because, as discussed last week, the photon number increases by (11/4) during electron-positron annihilation.

7 Recombination

In the early Universe, at temperatures $T \gg$eV, the atoms in the Universe are ionized, and photons are tightly coupled to the baryons through Thomson scattering from the electrons. At a temperature $T \sim$eV, electrons and protons combine to form hydrogen atoms, free electrons disappear, and without any electrons to scatter from, photons decouple and free stream through the Universe. These are the CMB photons we see today.

The temperature at which recombination occurs can be determined roughly from the *Saha equation*. Let's ignore the helium in the Universe. Then, there can be electrons, protons, and bound hydrogen atoms, with abundances n_e, n_p, and n_H, and the baryon density is $n_B = n_p + n_H$. In terms of the temperature and chemical potentials, the abundances are

$$n_i = g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_i - m_i}{T}\right). \tag{19}$$

From the reaction $e^- + p \leftrightarrow H$, we have $\mu_p + \mu_e = \mu_H$, which allows us to write,

$$n_H = g_h \left(\frac{m_H T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_p + \mu_e - m_H}{T}\right) = \frac{g_H}{g_p g_e} n_p n_e \left(\frac{m_e T}{2\pi}\right)^{-3/2} e^{B/T},] \tag{20}$$

where the binding energy $B \equiv m_p + m_e - m_H = 13.6$ eV, and we have used $m_p \simeq m_H$ in the prefactor. Defining an ionization fraction $X_e \equiv n_p/n_H$, and $g_p = g_e = 2$ and $g_H = 4$, and $n_B = \eta n_\gamma$, where $\eta = 5.1 \times 10^{-10}(\Omega_b h^2/0.019)$ is the baryon-to-photon ratio, and $n_p = n_e$ and $n_H = (1 - X_e)n_p$, we find (*this equation needs to be checked!*)

$$\frac{(1 - X_e^{\rm eq})}{(X_e^{\rm eq})^2} = \frac{4\sqrt{2}\zeta(3)}{\sqrt{\pi}}\eta \left(\frac{T}{m_e}\right)^{-3/2} e^{B/T}. \tag{21}$$

This is known as the Saha equation. Keep in mind that here that $T(z) = T_0(1 + z)$. At high temperatures, one gets an ionization fraction $X_e \to 1$, and $X_e \to 0$ at low temperatures. Evaluating the Saha equation numerically, one finds that at a redshift $z_{\rm rec} \simeq 1260$, at a temperature $T \simeq 0.3$ eV, about 90% of the electrons have become bound into hydrogen atoms. Of course, the Saha equation describes the ionization balance in equilibrium, and this will only be valid as long as the reactions $e^- + p \leftrightarrow H$ that maintain equilibrium occur rapidly compared with the expansion rate H. Toward the end of recombination, when $n_e = n_p$ drop exponentially, the reaction rate also drops exponentially until at some point this drops below the expansion rate. With the correct cross section for recombination, one finds that this happens at a redshift $z \simeq 1100$, resulting in a residual free-electron density $n_e \simeq 3 \times 10^{-4}$.

Photons decouple very shortly after recombination, as the free electrons disappear. The mean-free path for a photon is $\lambda = (n_e \sigma_T)^{-1}$, where $\sigma_T = 6.625 \times 10^{-25}$ cm^2 is the Thomson cross section. Photons decouple roughly when this exceeds the age of the Universe, which at these times, is $t = (2/3)(1 + z)^{-3/2}(\Omega_m H_0^2)^{-1/2}$. Plugging in numbers and evaluating the Saha equation numerically, one finds that CMB photons last scatter at $1+z_{ls} \simeq 1100$ when the age of the Universe is $t \simeq 380,000$ years (note that to get the correct age, you must take into account the fact that although the Universe is matter dominated at these redshifts, the radiation is not yet negligible). This is therefore where the CMB photons that we see were last scattered. Since the baryon-to-photon ratio is 5×10^{-10}, it means that only one photon per billion in the CMB is produced by an electron-proton recombination; the rest are just those in the thermal bath, and we therefore expect them to have a very precisely blackbody spectrum....as is observed.

8 Big-bang nucleosynthesis

The theory of big-bang nucleosynthesis is one of the major triumphs of cosmology in the 20th century. By simply applying the rates for nuclear reactions that are measured in the laboratory to a plasma of the correct baryon density in an expanding Universe roughly a few seconds to minutes after the big bang (at temperatures $T \sim 0.1 - 10$ MeV), we find that the neutrons and protons in the Universe organize themselves into roughly 75% hydrogen and 25% helium (by mass), with calculable trace abundances of deuterium and ^{7}Li. The predictions are in excellent agreement with the observations, and the success of the theory allows us to place important constraints to the content and evolution of the Universe just seconds after the big bang. The full theory is straightforward and quite involved, and we will not go into it in detail. Instead, we will try to understand in a simple way why the helium mass fraction is $\sim 25\%$.

Helium is the most tightly bound light ($A \lesssim 8$) nucleus; by "tightly bound", I mean largest binding energy per nucleon. More massive nuclei are more tightly bound. However, the nuclear reactions that might build heavier nuclei from helium nuclei, neutrons, and protons, freeze out (i.e., are outpaced by the expansion) shortly after helium nuclei are created, so these reactions never occur. Therefore, essentially all the neutrons around at the time of BBN wind up in helium nuclei, and this fact can be used to estimate the helium abundance. Initially, at temperatures $T \gg$MeV and $t \ll$sec, neutrons and protons are rapidly interconverting through reactions like $\nu_e + n \leftrightarrow p + e^-$ and $e^+ + n \leftrightarrow p + \bar{\nu}_e$. Strictly speaking, the reaction $n \leftrightarrow p + e^- + \bar{\nu}_e$ can also occur, but the timescale for this reaction is ~ 10 min, too long to take place at these early times. If proton-neutron conversion takes place faster than the expansion rate, then chemical equilibrium, $\mu_n + \mu_\nu = \mu_p + \mu_e$, will hold. If so, then the neutron-proton ratio will be

$$\frac{n}{p} \equiv \frac{n_n}{n_p} = \exp\left[-Q/T + (\mu_e - \mu_\nu)/T\right], \tag{22}$$

where $Q \equiv m_n - m_p = 1.293$ MeV is the neutron-proton mass difference. With the expressions for equilibrium abundances, one can show that at $T \gg m_e$, $\mu_e/T \sim (n_{e^-} - n_{e^+})/n_\gamma = n_p/n_\gamma \sim \eta \ll 1$, and similarly $\mu_\nu/T \ll 1$, if there is no large neutrino asymmetry. Therefore, if neutron-proton conversion occurs in equilibrium, then $(n/p)\mathrm{eq} = e^{-Q/T}$.

The neutron decay rate can be calculated in weak-interaction theory, and it is

$$\Gamma_{n\to pe\nu} = \tau_n^{-1} = \frac{G_F^2}{2\pi^3}(1 + 3g_A^2)m_e^5\lambda_0, \tag{23}$$

where G_F is Fermi's constant, g_A and axial-vector coupling, and $\lambda_0 = 1.6363$ is a phase-space factor that can be calculated from the theory. The constants G_F and g_A are determined from measurement of the neutron lifetime, the neutron decay spectrum, and can also be determined (and checked for self-consistency) from other weak-decay processes. The matrix elements for the other neutron-proton conversion processes are related to the neutron lifetime, and so the rate for proton-neutron interconversion in the thermal bath at temperature T can be shown to be

$$\Gamma_{pe\leftrightarrow\nu n} = \left\{ \begin{array}{cc} \tau_n^{-1}(T/m_e)^3 e^{-Q/T} & T \ll Q, m_e \\ \frac{7\pi^4\tau_n^{-1}}{30\lambda_0}\left(\frac{T}{m_e}\right)^5 \simeq G_F^2 T^5 & T \gg Q, m_e \end{array} \right. \tag{24}$$

Numerically, we find that at $T \gtrsim m_e$, $(\Gamma/H) \simeq (T/0.8\,\mathrm{MeV})^3$. Thus, at $T \gtrsim 0.8$ MeV, we expect the equilibrium neutron-proton ratio to hold: $n/p = (n/p)_{\mathrm{eq}}$. However, the neutron-proton mass ratio will then freeze out at $T \simeq 0.8$ MeV, and it will do so at a ratio,

$$\left(\frac{n}{p}\right)_{\mathrm{freezeout}} = e^{-Q/T_f} \simeq \frac{1}{6}. \tag{25}$$

Over the next $1-3$ minutes, the temperature drops to $T = 0.3-0.1$ MeV. Neutrons and protons will begin to combine, through a chain of nuclear reactions that proceeds through deuterium, to helium. The precise calculation is somewhat involved, and the "correct" answer cannot be obtained through any simple argument. Still, what happens, roughly speaking, is that neutrons and protons begin to collect into ^{4}He nuclei, and the helium abundance is roughly that determined by nuclear statistical equilibrium. The balance between free neutrons and neutrons in helium nuclei is determined by something like the Saha equation that we saw above that determines the balanced between free electrons and those bound in hydrogen atoms. As in that case, "recombination" takes place when the temperature is roughly 1/10 the ionization energy (in that case, 13.6 eV). In this case, the "ionization" energy is the binding energy per nucleon in a helium atom, roughly a few MeV per nucleon. It is thus reasonable to expect (and straightforward and detailed calculations bear out) that at a temperature $T \sim 0.1$ MeV, at a time $t \sim 3$ min, every neutron (to a first approximation) in the Universe finds itself in a helium nucleus. Although we calculated above that $ep \leftrightarrow \nu n$ freezeout results in $n/p \simeq 1/6$, over the three minutes since freezeout, roughly one in five of the remaining neutrons decayed (recalling that $\tau_n \simeq 15$ min). Thus, at $T \sim 0.1$ MeV, when neutrons fall into helium nuclei, the neutron-proton ratio has decreased to roughly $n/p \simeq 1/7$. Therefore, big-bang nucleosynthesis results in a helium mass fraction,

$$X_4 = \frac{4n_H}{n_N} = \frac{4(n_n/2)}{n_n + n_p} = \frac{2(n/p)}{1+(n/p)} \simeq 0.25, \qquad (26)$$

using $n/p \simeq 1/7$.

Detailed calculations evolve the entire network of nuclear reactions forward in time numerically with a series of rate equations ("Boltzmann equations") for all of the nuclear reactions. Detailed predictions are made for the abundances of ^{7}Li, ^{4}He, and deuterium as a function of the baryon-to-photon ratio $\eta \propto \Omega_b h^2$. The most interesting BBN prediction in recent years has been that for the D abundance, which has been measured in $z \sim 3$ Lyman-alpha absorption systems to be $D/H \simeq 2 \times 10^{-5}$ which suggests a baryon density $\Omega_b h^2 = 0.020 \pm 0.002$ (95% CL).

The agreement between the predicted and observed light-element abundances allows us to place some confidence in BBN, and therefore to use it to constrain alternatives to the standard cosmological model. In particular, the agreement suggests that the expansion rate at BBN is what we think it is. Suppose, for example, that there were several additional light-neutrino degrees of freedom, in addition to the three (e, μ, and τ) we know, in equilibrium at the time of BBN. Then, g_* would be bigger, and therefore the expansion rate $H^2 \propto g_*^{1/2} T^2$ larger at the same temperature. If the expansion rate were larger, then the neutron-proton conversion would freeze out at an earlier time and higher temperature, and thus at a higher neutron-to-proton mass ratio. This would then lead, according to the equation above, to a larger ^{4}He abundance. When people carry through this argument carefully, the conclusion is that (conservatively), there can be no more than the equivalent of one light additional neutrino at the time of BBN.

9 Thermal relics and dark matter

Many theories for new physics beyond the standard model, and particularly those that introduce supersymmetry (a symmetry between bosons and fermions), predict the existence of a new neutral and stable weakly-interacting massive particle (WIMP) χ that has a mass in the range 10s to 1000s of GeV and interactions with ordinary matter with strengths characteristic of electroweak interactions (e.g., like neutrinos). For example, in the minimal supersymmetric extension of the standard model (MSSM), the neutralino, a linear combination of the supersymmetric partners of the photon, Z^0 boson, and Higgs bosons (i.e., a linear combination of the photino, zino, and higgsino) is the lightest supersymmetric partner, it is neutral and stable, and can thus serve as a WIMP. Depending on the details of the supersymmetric model, these particles can have a mass m_χ in the 10s to 1000s of GeV range, and they have electroweak-strength interactions with quarks and leptons. The neutralino is a spin-1/2 particle, and it is a Majorana particle, meaning that it is its own antiparticle.

Recall that the baryon density is $\Omega_b \simeq 0.05$, while the nonrelativistic-matter density is $\Omega_m \simeq 0.3$. That means that $\sim 25\%$ of the critical density must be in the form of nonrelativistic nonbaryonic matter. This *dark matter* has been required by galactic dynamics for years. In our own Milky Way, the mismatch between the velocities that stars are observed to rotate around the center of the galaxy, and the velocities they should have if the stellar mass was all there were, has long suggested that dark matter outweighs the luminous matter in the Milky Way, and just about every other galaxy, by at least a factor of 10. We will now see that WIMPs may well be the dark matter, if they exist.

In the early Universe when the temperature $T \gg m_\chi$, these particles should be effectively massless and they should exist in thermal equilibrium with the standard-model particles (quarks, leptons, photons, gluons, gauge bosons) if, of course, the strength of their interactions with standard-model particles is large enough. We will consider the case of a neutralino, a Majorana spin-1/2 particle. The equilibrium abundance of χ particles is maintained by conversion of neutralinos to standard-model particles through reactions like $\chi\chi \leftrightarrow l\bar{l}$, where $l = (e^\pm, \mu^\pm, \tau^\pm, u, d, s, c, b, t, W^\pm, Z^0)$. The rate for annihilation of a pair of WIMPs to lighter particles is $\Gamma_{\chi\chi\to l\bar{l}} = \left\langle \sigma_{\chi\chi\to l\bar{l}|v|} \right\rangle n_\chi$, where $\sigma_{\rm ann} \equiv \sigma_{\chi\chi\to l\bar{l}}$ is the cross section for annihilation of WIMPs to lighter particles, $|v|$ is the relative velocity between the annihilating particles, the angle brackets denote an average over the thermal energy distribution, and n_χ is the WIMP number density.

The time evolution of the number density of WIMPs is given by the following *Boltzmann equation*,

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -\langle \sigma_{\rm ann}|v| \rangle \left[n_\chi^2 - (n_\chi^{\rm eq})^2 \right], \tag{27}$$

where $H = 1.66\, g_*^{1/2} T^2/m_{\rm Pl}$ is the expansion rate, n_χ the actual number density, and $n_\chi^{\rm eq}$ the number density the particles have in thermal equilibrium at a temperature T. This equation is derived from kinetic theory, but it can be understood heuristically in a simple way. The right-hand side of the equation describes the effect of annihilation and creation processes on the number density, while the left-hand side describes the effect of the expansion. Suppose, first, that the left-hand side were zero—i.e., particles were neither created nor destroyed. Then the equation would be $(dn_\chi/dt) + 3(\dot{a}/a)n_\chi = 0$, which has a solution $n_\chi \propto a^{-3}$, which is exactly what we expect if particle number is conserved. Now consider the right-hand side. The first term simply describes the rate per unit volume at which particles are removed by annihilation. (Note that there is a subtlety for Majorana particles here. This term should be multiplied by a factor of two, since two particles are removed in each annihilation process. However, the number of way of pairing N identical particles is $N(N-1)/2$, which implies that the rate per unit volume at which particles annihilate is $\langle \sigma_{\rm ann}|v| \rangle\, n_\chi^2/2$.) Now consider the second term, which describes creation of $\chi\chi$ pairs by the inverse process $l\bar{l} \to \chi\chi$. The principle of detailed balance tells us that in thermal equilibrium, the forward and reverse reactions occur at the same rate. Therefore, if the light particles $l\bar{l}$ have thermal distributions, the rate per unit volume for creation of χ particles is then $\langle \sigma_{\rm ann}|v| \rangle\, (n_\chi^{\rm eq})^2$.

In practice, the Boltzmann equation can be evolved numerically, and it is also fairly simple to come up with a reliable analytic approximation. Here, we work throug a very simple solution that is a bit less accurate than a more sophisticated solution, but describes the essential physics. We will see that at early times, when the annihilation rate exceeds the expansion rate, the relic density of particles tracks the equilibrium number density. Then, at some time, the annihilation rate gets outpaced by the expansion and at the comoving number density of particles remains constant thereafter.

The solution depends on whether we are dealing with a "hot relic," one that freezes out when it is relativistic ($T \gtrsim m_\chi$) like the example of neutrinos that we worked out before), or a "cold relic," one that freezes out when it is nonrelativistic ($T \lesssim m_\chi$). Let's first consider hot relics. Since the equilibrium number density $n_\chi \propto T^{-3}$, while the expansion rate $H \propto T^2$, the annihilation rate $\Gamma \propto n_\chi \sigma v$ will exceed the expansion rate at some sufficiently early time as long as $\sigma v \propto T^n$ with $n > -1$. This is certainly true for neutrinos ($n = 5$) as we saw before, and it is true for any particle candidate I've ever heard of. I'm not sure, but I think that it can be proved for pointlike particles that $n > -1$ always. So let us suppose that $n > -1$. Then, at early times, $\Gamma > H$, but then at some temperature T_f, defined by $\Gamma(T_f) = H(T_f)$, the annihilation rate drops below the expansion rate, and annihilations freeze out. The comoving number density of particles then scales like $n_\chi \propto a^{-3}$ thereafter. This means that the particle-to-entropy ratio, $Y \equiv n_\chi / s$ is therefore constant with time. From our expressions for n_χ (in the relativistic regime) and s, we can calculate that

$$Y = \frac{25\zeta(3)}{2\pi^4} \frac{g_{\text{eff}}}{g_{*s}} = 0.278 \frac{g_{\text{eff}}}{g_{*s}}, \tag{28}$$

for a particle that freezes out when it is relativistic, and where $g_{\text{eff}} = g$ for bosons, and $g_{\text{eff}} = (3/4)g$ for fermions. Since this remains constant with time, and since $s = 2970\,\text{cm}^{-3}$ today, $n_{\chi 0} = 825\,[g_{\text{eff}}/g_{*s}(T_f)]\,\text{cm}^{-3}$ today. Now suppose that although the relic freezes out when $T \gtrsim m_\chi$, but that $m_\chi \lesssim T_{\text{CMB}}$ today. Then, the relic will be nonrelatistic today and contribute a nonrelativistic-matter density (in units of critical) of $\Omega_\chi = m_\chi n_{\chi 0}/\rho_c = 7.83 \times 10^{-2}\, h^{-2} [g_{\text{eff}}/g_{*s}(T_f)](m_\chi/\text{eV})$. So let us now return to neutrinos, which decouple at a temperature $T \simeq$MeV when $g_s = 10.75$. For Majorana (2-component) neutrinos, $g_{\text{eff}} = 2(3/4)$, and so $\Omega_\nu h^2 = (m_\nu/91.5\,\text{eV})$. Since $\Omega_\chi h^2 < 0.122$ (from recent data from the CMB and other measurements) it implies that $m_\nu < 11$ eV. Considering that there are three neutrino species, what this really means is that if they are 2-component neutrinos, then there masses must sum to $\sum_i m_{\nu i} < 11$ eV. This is know as the *Cowsik-McLelland* bound, although it was in fact derived by Russian theorists (whose name I cannot remember now) a decade before, in the early 1960s.

As another application, suppose one postulated a left-right symmetric model with right-handed neutrinos that interact via exchange of some hypothetical and much more massive right-handed gauge bosons $W_R^\pm$ and Z_R^0 such that these right-handed neutrinos decoupled at a temperature $T \gtrsim 100$ GeV. Then g_{*s} would be about $10\times$ larger at decoupling and so the limit to the sums of these right-handed neutrino masses would be $\simeq 100$ eV.

Now let's move on to cold relics, which freeze out when $T \lesssim m_\chi$. Initially, at temperatures $T \gg m_\chi$, $n_\chi \propto T^3$, so under the same conditions for $\sigma_{\text{ann}} v$ as we considered for hot relics above, the annihilation rate Γ will at some sufficiently early time be $\Gamma \gtrsim H$. By assumption (or definition of cold relic), the annihilation rate $\Gamma \gtrsim H$ until T drops below m_χ. Subsequently, the number density $n_\chi \propto \exp(-m_\chi/T)$ drops exponentially with the temperature, and at some point, the equilibrium annihilation rate falls below the expansion rate. At this time, annihilations freeze out, and the comoving number density of WIMPs subsequently remains constant.

This behavior can be seen from the Boltzmann equation. If $\Gamma = n_\chi \langle \sigma_{\text{ann}} |v| \rangle \gg H$, then the right-hand side of the Boltzmann equation is larger than the $3Hn_\chi$ term, and the right-hand side drives $n_\chi \rightarrow n_\chi^{\text{eq}}$. If, however, $H \gg \Gamma$, then the right-hand side, which restores the density to the equilibrium density, becomes unable to keep pace with the dilution of the number density due to the expansion.

Freezeout is determined by $H(T_f) = \Gamma(T_f)$, or

$$1.66\, g_*^{1/2} \frac{T_f^2}{m_{\rm Pl}} = g \left(\frac{m_\chi T_f}{2\pi} \right)^{3/2} e^{-m_\chi/T_f} \left\langle \sigma_{\rm ann} |v| \right\rangle , \tag{29}$$

and $\langle \sigma_{\rm ann}|v|\rangle$ and g_* are both evaluated at a temperature T_f. This is rearranged algebraically to give

$$\frac{T_f}{m_\chi} = \left\{ \ln \left[\frac{g}{g_*} \frac{m_\chi^2 m_{\rm Pl} \left\langle \sigma_{\rm ann}|v| \right\rangle (T_f/m_\chi)^{1/2}}{1.66\,(2\pi)^{3/2}} \right] \right\}^{-1} . \tag{30}$$

This can be solved numerically for T_f, or evaluated recursively; defining $y = 0.038(g/g_*^{1/2}) m_\chi m_{\rm Pl} \left\langle \sigma_{\rm ann}|v| \right\rangle$, we have

$$\frac{m_\chi}{T_f} = \ln y - \frac{1}{2} \ln \ln y - \frac{1}{2} \ln \ln \ln y - \cdots . \tag{31}$$

For example, if $m_\chi = 100$ GeV, and the annihilation cross section is $\sigma v \simeq \alpha^2/m_\chi^2 \simeq 10^{-8}\,{\rm GeV}^{-2}$, then $m_\chi/T_f \simeq \ln[0.038(2/10)(100)(10^{19})(10^{-8})] \simeq 25$, and this depends only logarithmically on m_χ and $\langle \sigma_{\rm ann}|v|\rangle$. We thus learn that freezeout occurs when the temperature drops not to $T \simeq m_\chi$, but all the way down to $T \simeq m_\chi/20$, well into the nonrelativistic regime.

At this time, the WIMP abundance is

$$n_\chi^{\rm freezeout} = \frac{\Gamma_{\rm freezeout}}{\langle \sigma_{\rm ann}|v| \rangle} = \frac{H_{\rm freezeout}}{\langle \sigma_{\rm ann}|v| \rangle} = \frac{1.66\, g_*^{1/2} T_f^2 / m_{\rm Pl}}{\langle \sigma_{\rm ann}|v| \rangle} = \frac{1.66\, g_*^{1/2} m_\chi^2}{\langle \sigma_{\rm ann}|v| \rangle\, (m_\chi/T_f)^2 m_{\rm Pl}}, \tag{32}$$

and the entropy density is

$$s_{\rm freezeout} = \frac{2\pi^2}{45} g_{*s} \frac{m_\chi^3}{(m_\chi/T_f)^3}. \tag{33}$$

We then use the fact that n_χ/s remains constant after freezeout, the value $s_0 = 2970\,{\rm cm}^{-3}$, and then find that the WIMP mass density (in units of critical) is today

$$\Omega_\chi h^2 = \frac{m_\chi n_\chi}{(\rho_c/h^2)} \simeq \left(\frac{\langle \sigma_{\rm ann}|v| \rangle}{3 \times 10^{-27}\,{\rm cm}^3/{\rm sec}} \right)^{-1} . \tag{34}$$

Thus, to a first approximation, the relic density is simply inversely proportional to the annihilation cross section. If the cross section is larger, then WIMPs stay in equilibrium longer leading to a greater degree of exponential suppression, and thus have a lower relic density. Whereas if they have weaker interactions, they freeze out earlier, with a larger relic density. Note that if the annihilation cross section is $\langle \sigma_{\rm ann}|v| \rangle \sim 3 \times 10^{-26}\,{\rm cm}^3/{\rm sec}$, then $\Omega_\chi h^2 \simeq 0.1$, the value required to explain the nonbaryonic dark matter. Such a cross section is, in particle-physics units, $\langle \sigma_{\rm ann}|v| \rangle \sim 2.5 \times 10^{-9}$ ${\rm GeV}^{-2}$, remarkably close to our simple weak-scale estimate $\alpha^2/(100\,{\rm GeV})^2 = 10^{-8}\,{\rm GeV}^{-2}$. What this implies is that if there is new physics at the electroweak scale that involves the introduction of a new stable neutral particle, then that particle must have a relic density in the ballpark of that required to explain the nonbaryonic dark matter. This is truly a coincidence, as there is no *a priori* reason to expect the electroweak scale of elementary-particle interactions to have anything to do with the critical density, which is fixed by the expansion rate, something having to do with the largest-scale structure of the Universe. This coincidence has led a number of theorists and experimentalists to take the idea of WIMP dark matter very seriously.

Let's consider another application of this calculation. Consider a Universe that begins initially with no baryon-antibaryon asymmetry. In that case, baryons and antibaryons could annihilate to photons with a cross section $\langle\sigma_{\rm ann}|v|\rangle \simeq {\rm fm}^2$. With a baryon mass, one finds $T_f \simeq m_p/40 \simeq 22$ MeV and $(n_b/s) = (n_{\bar{b}}/s) \simeq 7\times10^{-20}$ (i.e., strongly-interacting particles annihilate very efficiently). Today, however, $n_b/s \simeq \eta/7 \sim 10^{-11} \gg 10^{-20}$. In other words, a baryon-symmetric Universe gets the baryon density wrong by 9 orders of magnitude. From this, we conclude that there must have been an initial baryon asymmetry, and that the abundance of antibaryons should be entirely negligible today.

As another example, suppose that there were a heavy (and stable!) fourth-generation (Dirac or Majorana) neutrino. Such a neutrino could annihilate through a Z^0 boson to the three light neutrinos, electrons, muons, tauons, and whatever quarks were lighter than this new neutrino. For Dirac particles, the cross section for annihilation of neutrino-antineutrino pairs to these lighter particles turns out to be $\langle\sigma_{\rm ann}|v|\rangle \simeq \alpha^2 m_\nu^2/m_Z^4$. If $g=2$ and $g_*=60$ (appropriate for particles in the $m_\nu \sim$GeV range), then $(m_\chi/T_f) \simeq 15{+}3\ \ln(m_\nu/{\rm GeV})$ and $(n_\chi/s) \simeq 6\times10^{-9}\,(m_\nu/{\rm GeV})^{-3}\,[1+(1/5)\ln(m_\nu/{\rm GeV})]$, and

$$\Omega_{\nu\bar{n}u}h^2 = 3\left(\frac{m_\nu}{\rm GeV}\right)^{-2}\left[1+\frac{1}{5}\ln\left(\frac{m_\nu}{\rm GeV}\right)\right]. \tag{35}$$

Therefore, we require $m_\nu \gtrsim 6$ GeV, or else the limit $\Omega_{\nu\bar{n}u}h^2] \lesssim 0.1$ is violated. This is known as the Lee-Weinberg bound (after particle theorists Steve Weinberg and Benjamin Lee), although it was derived by a number of other people beforehand, and probably first by Zeldovich in 1965.