PHY 475/375
Lecture 16
(May 21, 2012)

Big Bang Nucleosynthesis

One of the greatest successes of the Big Bang theory is its prediction of the relative abundances of the lightest elements (hydrogen, deuterium, helium, and some lithium and beryllium), which were formed when the Universe was very young (about a few seconds to minutes old).

Recall that at $T > 10^{13}\,\mathrm{K}$ (i.e., $kT > 1\,\mathrm{Gev}$, and $t < 10^{-5}\,\mathrm{s}$), the Universe contains protons and anti-protons, neutrons and anti-neutrons, electrons and positrons, all three flavors of neutrinos and anti-neutrinos, and photons, all of them in thermal equilibrium.

At 10^{13} K, protons and neutrons freeze out (decouple), whereupon protons and anti-protons annihilate, as do neutrons and anti-neutrons. Due to an as-yet unexplained excess of matter over anti-matter, some excess protons and neutrons are left behind in the Universe.

At $T \sim 10^{11}\,\mathrm{K}$ (i.e., $kT \sim 10$ MeV, and $t \sim 10^{-2}$ s), we are left with electrons and positrons, all neutrinos and their anti-neutrinos, and photons in thermal equilibrium. Note that in Lec 15 we mentioned two other leptons (μ^-, μ^+, and τ^-, τ^+), but they have frozen out by this time.

At this time, neutrons decay to protons and vice versa, via weak decays:

$$\begin{aligned} n &\rightarrow p + e^- + \overline{\nu}_e \\ n + e^+ &\rightarrow p + \overline{\nu}_e \\ p + e^- &\rightarrow n + \nu_e \end{aligned} \tag{10.7}$$

All of these occur rapidly, so they maintain the relative number densities of neutrons and protons at the equilibrium value:

$$\frac{n_n}{n_p} = \exp\left(-\frac{Q_n}{kT}\right) \tag{10.14}$$

where $Q_n = \left(m_n - m_p\right)c^2 = 1.29$ MeV.

By $T \sim 9 \times 10^9\,\mathrm{K}$ (i.e., $kT \sim 0.8$ MeV, and $t \sim 1$ s), all flavors of neutrinos have decoupled, and this decoupling of neutrinos means that the reactions (10.7) have become too slow to maintain the neutrons and protons at their equilibrium value (10.14). Therefore, the neutron-to-proton ratio is frozen at the value

$$\frac{n_n}{n_p} = \exp\left(-\frac{Q_n}{kT}\right) \approx \exp\left(-\frac{1.29\ \mathrm{MeV}}{0.8\ \mathrm{MeV}}\right) \approx 0.2 \tag{10.17}$$

at least until the neutrons start to decay. At these times, therefore, there was one neutron for every 5 protons in the Universe.

All along, ever since protons and neutrons have existed in the Universe, there has been a tendency for them to bind into nuclei (via interactions involving the strong nuclear force). However, they were blasted apart by energetic photons as soon as they formed.

Since the binding energy of deuterium $(1p, 1n)$ is 2.2 MeV, one might expect that now that we are down to $kT \sim 0.8$ MeV, there should be a lot of deuterium forming.

However, the Universe has more than a billion photons per nucleon (as we calculated in Lecture 14), so there are plenty of photons on the high-energy tail of the thermal distribution that can photodissociate deuterium nuclei. Therefore, we still don't have any stable deuterium nuclei.

Once $T \sim 10^9\,\mathrm{K}$ (i.e., $kT \sim 0.1$ MeV, and $t \sim 100$ s), however, the key reaction involving the strong nuclear force

$$p + n \rightleftharpoons D^+ + \gamma \tag{10.19}$$

where D^+ is the singly-charged deuterium ion (with one p and one n), proceeds only from left to right because there are few photons energetic enough to photodissociate the deuterium nucleus (the reaction going from right to left).

We now have stable deuterium nuclei, and the first step in nucleosynthesis is complete!

One might ask: what about proton-proton fusion, or neutron-neutron fusion?

Well, both of these processes involve the weak nuclear force, as can be seen by the involvement of a neutrino in the interactions:

$$p + p \rightarrow D^+ + e^+ + \nu_e \tag{10.18}$$

and

$$n + n \rightarrow D^+ + e^- + \overline{\nu}_e \tag{10.20}$$

The cross section for interactions involving the weak nuclear force is much smaller than the cross section for interactions involving the strong nuclear force. Moreover, since protons are all positively charged, they must surmount the Coulomb barrier in order to fuse.

Of course, proton-proton fusion is possible — in fact, it happens in the Sun and other stars. However, fusion in the Sun is a very slow process. Fortunately, the core of the Sun is a stable environment in hydrostatic equilibrium, with temperature and density changing very slowly over time.

In contrast, the temperature drops as $T \propto t^{-1/2}$ in the early Universe (because it is radiation-dominated; see Lecture 9), and the density of baryons drops as $n_{\mathrm{bary}} \propto t^{-3/2}$.

So, Big Bang Nucleosynthesis is a race against time! After less than an hour, the temperature and density have dropped too low for fusion to occur.

Given all this, neutron-proton fusion is favored. In fact, it proceeds until most free neutrons are locked into a nuclear bound state; some are left over after the temperature becomes too low for nucleosynthesis, and decay to protons. All leftover protons remain by themselves.

Once neutron-proton fusion has built up deuterium nuclei, several nucleon capture reactions can build heavier nuclei; for example:

$$D^+ + p \rightleftharpoons {}^3He^{++} + \gamma \tag{10.33}$$

$$D^+ + n \rightleftharpoons {}^3H^+ + \gamma \tag{10.34}$$

Soon after, these are converted to ${}^4He$ by reactions such as:

$$\begin{aligned} {}^3H^+ + p &\rightleftharpoons {}^4He^{++} + \gamma \\ {}^3He^{++} + n &\rightleftharpoons {}^4He^{++} + \gamma \end{aligned} \tag{10.38}$$

All such reactions involve the strong nuclear force, and have large cross sections and fast reaction rates. Therefore, once nucleosynthesis begins, $D, {}^3H$, and ${}^3He$ are all efficiently converted to ${}^4He$ nuclei.

Here, though, nucleosynthesis hits a roadblock. There are no stable nuclei with mass number $A = 5$. Fusing a p or n to ${}^4He$ won't work; ${}^5He$ and ${}^5Li$ are not stable nuclei.

Small amounts of ${}^6Li$ and ${}^7Li$ are made when ${}^4He$ nuclei combine with the much less abundant D or tritium (${}^3H$):

$${}^4He + D \rightleftharpoons {}^6Li + \gamma \tag{10.39}$$

$${}^4He + {}^3H \rightleftharpoons {}^7Li + \gamma \tag{10.40}$$

where we've suppressed writing the charges for convenience.

Big Bang Nucleosynthesis could not produce heavier nuclei with $A > 7$, because there is no stable $A = 8$ element. If ${}^8Be$ had been formed, it would have almost immediately disintegrated into two ${}^4He$'s (with a decay time of only 3×10^{-16} s). Stars do it, of course, via the *triple-α* reaction ${}^4He + {}^4He + {}^4He \rightarrow {}^{12}C$, but this requires high temperature and density.

By $T \sim 3 \times 10^8$ K (i.e., $kT \sim 0.03$ MeV, and $t \sim 1000$ s $\equiv$ 20 min), the temperature is low enough that the nuclei can no longer easily overcome the Coulomb barrier that separates them, and nucleosynthesis is effectively over. So, what do we have at the end of nucleosynthesis?

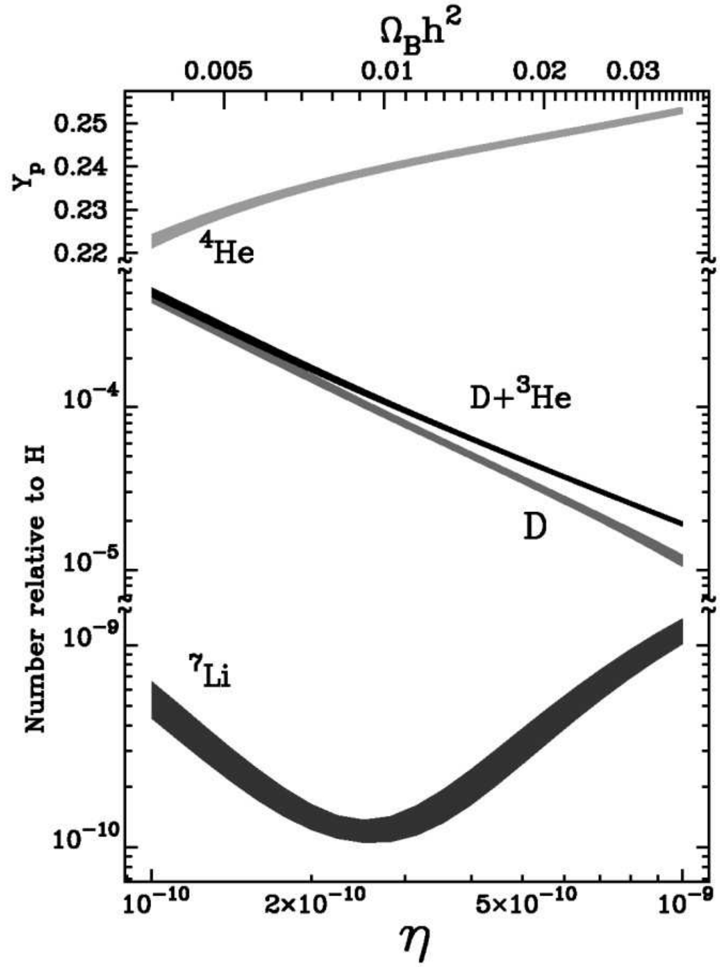
- Nearly all the baryons left over are in the form of free protons or ${}^4He$ nuclei.
- The lifetime of a free neutron is a little over 10 min, so any residual free neutrons quickly decay into protons.
- Small amounts of $D, {}^3H$, and ${}^3He$ are left over, the ${}^3H$ is unstable and decays to ${}^3He$
- Very small amounts of ${}^6Li$, ${}^7Li$, and ${}^7Be$ are synthesized. The ${}^7Be$ is later converted to ${}^7Li$ by electron capture (${}^7Be + e^- \rightarrow {}^7Li + \nu_e$).

While the rates and yields come from nuclear physics data, the key point of interest for cosmologists is that the baryon-to-photon ratio η determines the final relative abundances of protons and ^4He nuclei, and the much smaller amounts of D, Li, and Be nuclei, etc.

A higher η allows stable deuterium to form at a higher temperature, thereby giving an earlier start to nucleosynthesis. This means that more neutrons can interact and become part of helium nuclei, leaving less D and ^3He as leftovers. This would increase the fraction of helium nuclei relative to H-nuclei.

The relations are not quite as obvious for other species, but the calculations that need to be carried out have been done, even though they are extensive due to the large number of pathways for forming nuclei after deuterium. An example of such a calculation is shown in Figure 10.5 in your text (taken from Burles, Nollett, & Turner (2001), ApJ, 552, L1), and reproduced on the right.

For an example of a species where the relation may not be quite as obvious, notice how the production of ^7Li decreases with increasing η up to a minimum in the predicted density of ^7Li at $\eta \approx 3 \times 10^{-10}$, then begins to increase again. This is because the direct production of ^7Li by the fusion of ^4He and ^3H (equation 10.40) is a decreasing function of η, while its indirect production from the electron capture by ^7Be is an increasing function of η.



In the figure above, the helium fraction is denoted by Y_p (see the top left); Y_p is used to designate the primordial helium fraction of the Universe (that is, the helium fraction before nucleosynthesis begins in stars), and is a dimensionless number:

$$Y_p \equiv \frac{\rho(^4He)}{\rho_{\text{bary}}} \tag{10.5}$$

That is, Y_p is the mass density of primordial ^4He divided by the mass density of all the baryonic matter.

Everything we have discussed so far, of course, has been just predictions from calculations. This turns out to be the easy part, since nuclear models/atomic data are now excellent, so the calculations are robust.

How do the observations match up?

This turns out to be the hard part, in particular, trying to get at primordial abundances from observations. That is, we need to measure the primordial densities of the light elements before nucleosynthesis in stars started to alter the chemical composition of the Universe.

The easiest test is to go after the primordial ^4He abundance (Y_p).

That is because although Y_p should go up with increasing baryon density, it has a pretty weak dependence on baryon density (as we saw in the graph on the previous page).

So, if we can find that the primordial ^4He abundance is about 25%, we have a reasonable validation of our model.

As we discussed in class, this requires careful spectroscopic observations, supported by good models that track the ionization in stellar atmospheres and gas clouds, and incorporate precise and accurate values of atomic and nuclear parameters.

By observing many stars, and nearby and distant gas clouds, we do get $Y_p \sim 0.25$. But what if we want to be precise and accurate?

- The general strategy is to determine the Y-values of a sample of objects, where X, Y, and Z are the abundances per unit mass of H, He, and the rest of the elements respectively, with the usual normalization $X + Y + Z = 1$.

- Recall that astronomers collectively club everything higher than H or He as "metals", so the fraction Z would give the metallicity of the object.

- Having determined the Y values of a number of objects with varying metallicities Z, one then searches the (Y, Z)-space to find a slope $\Delta Y/\Delta Z$.

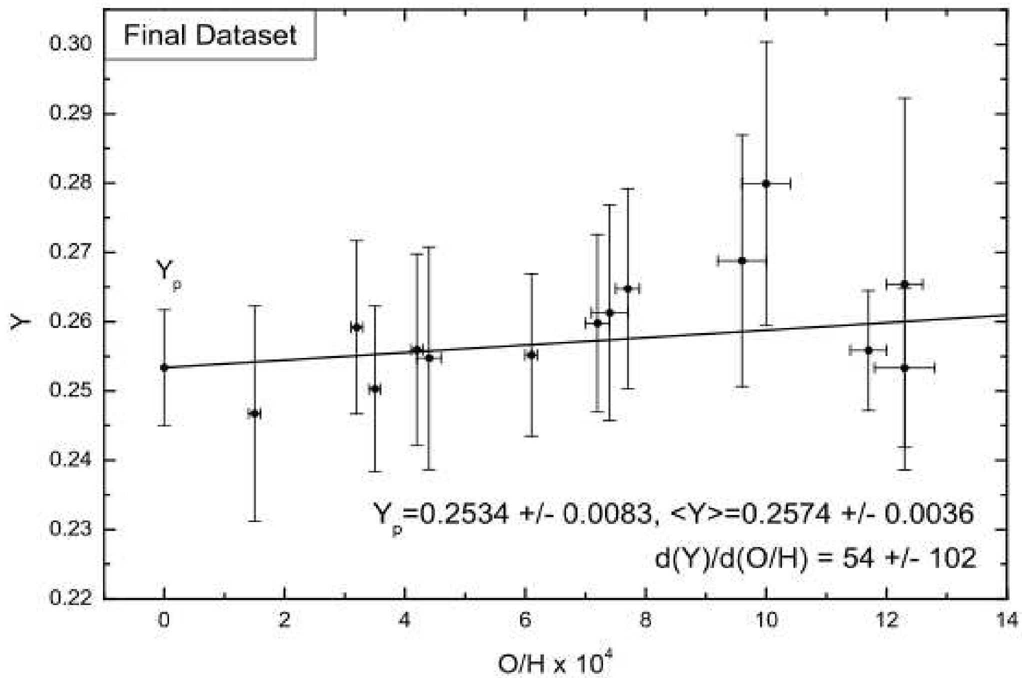
- Since it is difficult to count all varieties that contribute to the metallicity Z, recent work has focussed on using the oxygen (O) abundance as a proxy for the overall metallicity, so one would find $\Delta Y/\Delta O$.

- One then extrapolates this slope back to $Z = 0$. Since $Z = 0$ is the state of having almost nothing higher than H and He, it is called zero metallicity.

- Therefore, the extrapolated value of the He abundance at $Z = 0$ gives an estimate of the primordial helium abundance Y_p.

One such object of choice appears to be low-metallicity ionized hydrogen (H II) regions in dwarf galaxies. The reason for looking in dwarf galaxies is that such galaxies have very low star formation, so their material has not been processed as much by generations of stars as in, e.g., our own Galaxy. This is borne out by the low metallicity in the observed H II regions.

Recently, Aver, Olive, & Skillman of the University of Minnesota published a determination of $Y_p = 0.2534 \pm 0.0083$ from a large dataset of spectroscopic observations in low metallicity H II regions in dwarf galaxies. Their plot of the helium fraction Y vs. metallicity (with the O-to-H ratio standing in as proxy for metallicity), and the extrapolated $dY/d(O/H)$ back to zero metallicity to get the primordial helium abundance is shown below.



The error bars in this method are quite large. A very precise listing of the error budget in such work is given in a paper by Peimbert, Luridiana, and Peimbert (2007, *Astrophysical Journal*, v. 666, pp. 636). They list thirteen sources of error; see their Table 7 for details.

- Some of these errors are related to fundamental atomic physics (e.g., recombination coefficients of atomic hydrogen (H I), and singly ionized helium (He I) lines, and a correction factor for helium ionization).
- Some of the errors have to do with level populations in the environment (e.g., collisional excitation of H I and He I lines, and recombination coefficients of H I and He I lines).
- Some of the errors are related to the measurements themselves (e.g., correction for reddening, optical depth of He I triplet lines, and H I and He I line intensities).
- Some errors are related to the properties of the gas clouds (e.g., their temperature and density structure).
- Finally, there is the error in the extrapolation to zero metallicity (i.e., knowing the correct slope dY/dO).

The uncertainties in these are between ± 0.0005 to ± 0.0015, i.e., within a factor a 3 of each other, so a substantial improvement in any one would not significantly improve the determination of Y_p.

The D/H Abundance Ratio

The cleanest way to go after Y_p is likely by finding the deuterium abundance ratio, D/H.

- This is because primordial deuterium production has a strong dependence on the baryon density, and it seems impossible to make deuterium anywhere other than in the early stages of the Universe.
- In fact, deuterium is very easily destroyed in stars. In one of the first fusion reactions that occur when a gas cloud collapses gravitationally to form a star, deuterium is fused into helium. Therefore, the deuterium fraction in the Universe decreases with time.
- So, the observed abundance of deuterium in stars is a lower limit to the primordial abundance.

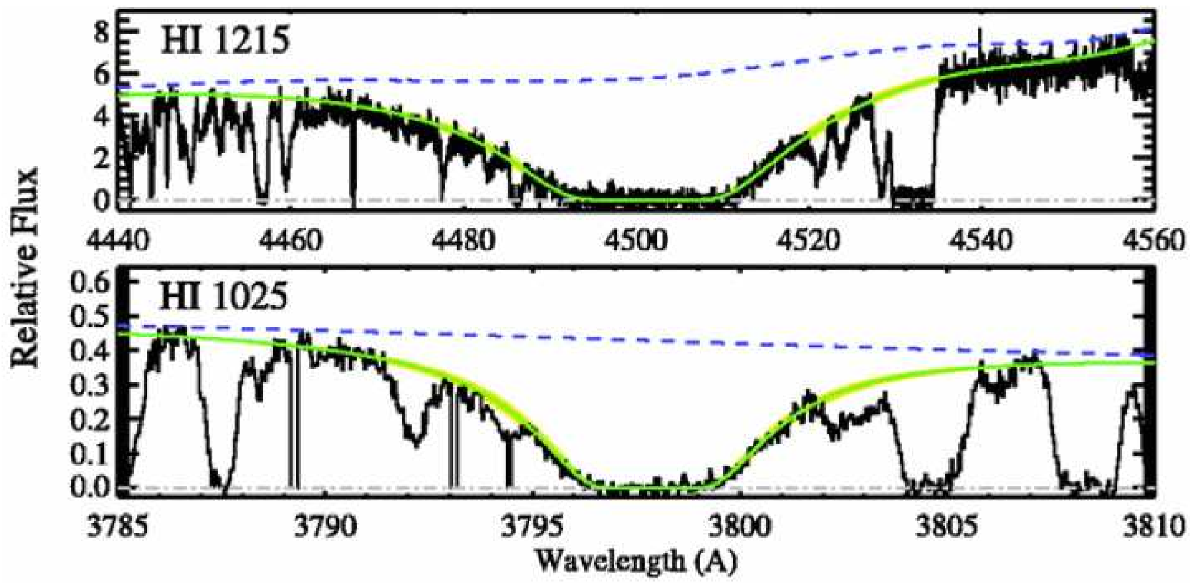
In the Solar System and local Interstellar Medium of our Galaxy, astronomers find that the helium abundance is $D/H \sim (1-2) \times 10^{-5}$, meaning that the primordial value

$$(D/H)_p > 1 \times 10^{-5}$$

A better way to find the primordial D/H is to look at the spectra of high-redshift quasars.

- Quasars are extremely bright sources that lie at great distances.
- Of course, we don't care about the deuterium content of the quasar itself; instead, we just want to use it to illuminate the gas clouds which lie between it and us.
- The strategy is to find intergalactic gas clouds that contain no detectable stars, and have very low levels of elements heavier than lithium. These would then have claims to a D/H content close to primordial.
- Neutral hydrogen atoms in these gas clouds will absorb photons whose energy corresponds to the Lyman-α transition from the atom's electronic ground state ($n = 1$) to ($n = 2$) with $\lambda_{\rm H} = 121.57$ nm.
- In the deuterium atom, the greater mass of the nucleus shifts the electron's energy levels, so that the Lyman-α transition in deuterium corresponds to a slightly shorter wavelength $\lambda_{\rm D} = 121.54$ nm.
- When we observe the light from a quasar that has passed through an intergalactic cloud at redshift $z_{\rm cl}$, we will see a strong absorption line at $\lambda_{\rm H}(1 + z_{\rm cl})$ due to absorption from ordinary hydrogen, and a much weaker absorption line at $\lambda_{\rm D}(1 + z_{\rm cl})$ due to absorption from deuterium.

An example is shown below for H I and D I Lyman series absorption in a gas cloud at redshift $z = 2.70262$ lying along the line of sight to the quasar SDSS 1558-0031; the plot is taken from Figure 1 of O'Meara et al. (2006, *Astrophysical Journal*, v. 649, L61).



The top panel in the figure above is the Lyman-α transition taken with the 6.5 m Magellan Clay telescope at Las Campanas, and the bottom panel shows the Lyman-β transition taken with the 10 m Keck telescope in Hawaii.

The data shown in the figure above have not been continuum subtracted, but the best estimate of the local continuum level is shown by the dashed blue line. What this means is that the blue dashed line is what you would see if the quasar light came to you directly without passing through the intervening gas cloud.

Note that the x-axis is labeled in Angstroms (A).

At $z_{\rm cl} = 2.70262$, the difference between the Lyman-α wavelengths for H I and D I is only 1.1 A, so the H I and D I lines will be blended together. This is why you don't see two dips, one for H I and the other for D I. Instead, the two have blended together to produce a broad absorption trough.

The authors have assumed that the gas cloud has a single absorption dip for the H line, and another single absorption dip for the D line, and fitted to the observed data. The solid green line shows their best single-component fit to the H I and D I absorption.

In other words, the velocity structure of the intervening gas cloud is very simple, with only one component.

They measure $\log_{10}(D/H) = -4.48 \pm 0.06$, corresponding to $D/H = (3.3 \pm 0.5) \times 10^{-5}$.

More importantly, O'Meara et al. (2006) point out that several criteria must be met for a quasar absorption system to give a reasonable measurement of D/H.

- First, you need a large column of hydrogen gas along the line of sight in the intervening gas cloud where the absorption is taking place (the official terminology for this is that the column density of hydrogen gas must be high in the gas cloud). That is because the typical D/H ratio is only about 10^{-5}, so you need a lot of ordinary hydrogen to get enough deuterium to detect.
- The velocity structure of the gas cloud must be simple. Ideally, you would like to have only a single component of absorbing gas. If there are multiple absorbing complexes, they must be well separated in frequency (or, equivalently through the Doppler shift, velocity). This is because H I and D I have almost the same spectra, separated by 0.1 nm at $z = 2.7$, as mentioned above.
- There cannot be other interloping line structure, including metal lines.
- The background quasar must be bright, so that one can do high signal-to-noise, high-resolution spectroscopy with a reasonable allotment of telescope time.

Each one of the above criteria acts to decrease the probability that a D/H measurement can be made toward any given quasar. Since all the criteria must be met, this means that only about 1% of quasars at $z \sim 3$ are suitable for measuring D/H.

Like the O'Meara et al. (2006) results, measurements from several high-redshift quasar absorption systems give D/H values in the range $(2-4) \times 10^{-5}$. A weighted average of these measurements gives $D/H = (3.0 \pm 0.4) \times 10^{-5}$ (Burles, Nollett, & Turner (2001), *Physical Review D*, v. 63, 6.063512).

To obtain the density parameter for baryons (Ω_b), we note that papers in journals usually quote values for $\Omega_b h^2$, where $h = H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$, because the errors in Ω_b and H_0 are degenerate.

Burles et al. (2001) find $\Omega_b h^2 = 0.020 \pm 0.002$. With our adopted value of $H_0 = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, this translates to $\Omega_b = 0.041 \pm 0.004$.

This compares well with the value obtained from the measurements of CMB anisotropies; recall that the 7-yr WMAP result gives a tighter constraint of $\Omega_b = 0.046 \pm 0.001$.

So what is the take-away lesson from Big Bang Nucleosynthesis? There is no doubt that it is extremely successful.

The $\sim$25% ^4He abundance is a key piece of evidence for the Big Bang model.

The best constraints come from measurements of the D/H abundance, the state of the art implies $\Omega_b h^2 = 0.020 \pm 0.002$; with $h = 0.7$, this corresponds to $\Omega_b = 0.041 \pm 0.004$.

These measurements are consistent with $\Omega_b h^2$ measured from CMB anisotropies, an extremely important consistency check, since CMB anisotropies come from completely different physics.

Inflation in the Early Universe

The Standard Hot Big Bang model that we have studied so far is an excellent fit to many observed properties of the Universe. For example, it explains

- the expansion of the Universe according to Hubble's law
- the existence of a Cosmic Microwave Background with a blackbody spectrum
- the primordial abundance of the light elements (H, He, D, Li, etc.)

However, the standard Big Bang theory needs some enhancements because it does not explain the following:

- the "flatness" problem, discussed below
- the "horizon" problem, discussed below
- other important matters like the baryon asymmetry and the origin of large scale structure

The Flatness Problem

We know that the Universe is very close to being spatially flat at the current epoch.

More formally, we can write this in the following way. The Friedmann equation in the form written in equation (4.29) in Lecture 6 relates the spatial curvature of the Universe to the density parameter Ω:

$$1 - \Omega(t) = -\frac{\kappa c^2}{R_0^2} \frac{1}{a(t)^2 H(t)^2} \tag{11.1}$$

At the present moment, the density parameter and curvature are therefore linked by the equation:

$$1 - \Omega_0 = -\frac{\kappa c^2}{R_0^2 H_0^2} \tag{11.2}$$

The results of Type Ia supernova observations and CMB anisotropy measurements are consistent with the value

$$\left|1 - \Omega_0\right| \leq 0.2 \tag{11.3}$$

with even stronger constraints of 0.02 from WMAP+BAO+H_0 (where BAO=Baryon Acoustic Oscillation) data (Komatsu et al. (2011), *Astrophysical Journal Supplement Series*, v. 192, pp. 18).

The question arises as to why the Universe should be so close to flat today, especially because when you extrapolate back into the past, this means that the Universe was even closer to flatness, and this is known as the Flatness Problem.

To see this aspect of approaching near-flatness in the past, let us combine equations (11.1) and (11.2) to get the density parameter as a function of time:

$$1 - \Omega(t) = \left[-\frac{\kappa c^2}{R_0^2} \right] \frac{1}{a(t)^2 H(t)^2}$$

$$\Rightarrow \quad 1 - \Omega(t) = \left[H_0^2 \Big(1 - \Omega_0 \Big) \right] \frac{1}{a(t)^2 H(t)^2} \tag{11.4}$$

For a spatially flat universe dominated by radiation and matter (i.e., before the onset of dark energy domination), we get from equation (6.6) that

$$\frac{H(t)^2}{H_0^2} = \frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} \tag{11.5}$$

so that (suppressing t's, for convenience)

$$\left(\frac{H^2}{H_0^2} \right) a^4 = \Omega_{r,0} + a\Omega_{m,0}$$

or

$$\frac{H_0^2}{a(t)^2 H(t)^2} = \frac{a^2}{\Omega_{r,0} + a\Omega_{m,0}}$$

Substituting this in equation (11.4), we get

$$1 - \Omega(t) = \frac{(1 - \Omega_0)\, a^2}{\Omega_{r,0} + a\Omega_{m,0}} \tag{11.6}$$

In the early Universe, when it was radiation dominated (i.e., when $\Omega_{r,0} \approx 1 - \Omega_0$), we get from equation (11.6) that

$$\Big| 1 - \Omega \Big|_r = \frac{(1 - \Omega_0)\, a^2}{\Omega_{r,0}} = \frac{\Omega_{r,0}\, a^2}{\Omega_{r,0}}$$

so that, in the epoch of radiation-dominance

$$\Big| 1 - \Omega \Big|_r \propto a^2 \tag{11.7}$$

Meanwhile, using the same procedure for the epoch when the Universe was dominated by matter, we get

$$\Big| 1 - \Omega \Big|_m = \frac{(1 - \Omega_0)\, a^2}{a\, \Omega_{m,0}} = \frac{\Omega_{m,0}\, a}{\Omega_{m,0}}$$

so that, in the epoch of matter-dominance

$$\Big| 1 - \Omega \Big|_m \propto a \tag{11.8}$$

Recall from Lecture 10 that we calculated $a_{rm} \equiv \Omega_{r,0}/\Omega_{m,0}$ in the epoch of radiation-matter equality. Using the Benchmark model with $\Omega_{r,0} = 8.4 \times 10^{-5}, \Omega_{m,0} = 0.3$, and $\Omega_{\Lambda,0} = 0.7$, this gives

$$a_{rm} \equiv \frac{\Omega_{r,0}}{\Omega_{m,0}} \approx 2.8 \times 10^{-4}$$

Therefore, at the moment of radiation-matter equality, the Universe was flat to a accuracy of

$$\left|1 - \Omega_{rm}\right| \leq 2 \times 10^{-4} \tag{11.9}$$

Recall from Lecture 9 that in a spatially flat, radiation-dominated universe $a \propto t^{1/2}$. So, around the time of Big Bang nucleosynthesis ($t = 100$ s), we would get

$$a \sim a(t_0)\left(\frac{t}{t_0}\right)^{1/2} \sim \left[\frac{100 \text{ s}}{13.6 \times 10^9\ (365 \times 24 \times 3600)}\right]^{1/2} \sim 10^{-8}$$

Note that we aren't assuming a flat universe to calculate the deviation from flatness, merely doing an order of magnitude estimate for a at an early time in a radiation dominated universe.

Using the order of magnitude estimate above, we get that during the time of Big Bang Nucleosynthesis, the Universe was flat to an accuracy of

$$\left|1 - \Omega_{\text{BBN}}\right| \leq 10^{-16} \tag{11.10}$$

Finally, from equation (6.38), we can get an order-of-magnitude estimate for the scale factor near the Planck time ($t_P \approx 5 \times 10^{-44}$ s):

$$a \sim \left(2\sqrt{\Omega_{r,0}}\, H_0 t\right)^{1/2} \sim 10^{-32}$$

so that, at the Planck time, the Universe was flat to an accuracy of

$$\left|1 - \Omega_{\text{P}}\right| \leq 10^{-60} \tag{11.11}$$

This remarkable closeness of Ω to unity (hence, spatial flatness) in the early Universe is known as the flatness problem. While it can be dismissed as a coincidence, one would feel more satisfied having a physical mechanism to flatten the Universe early in its history.

The Horizon Problem

The horizon problem, simply put, arises from the fact that the Universe appears to be homogenous and isotropic on the largest scales.

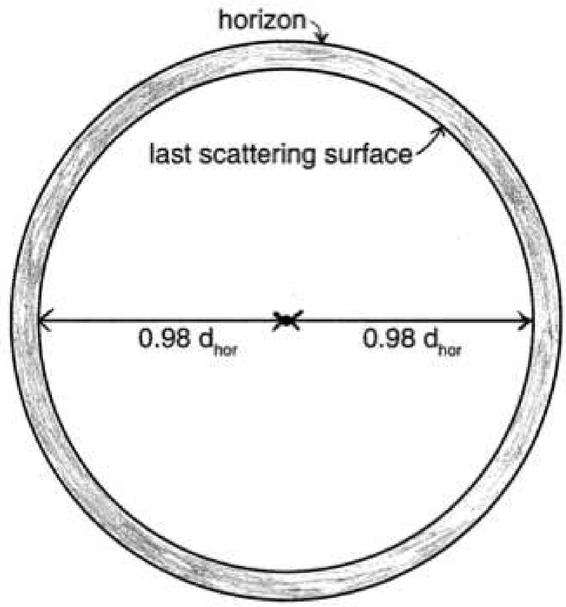
One might question why this is a problem. After all, the very fact that it is homogenous and isotropic, allowed us to describe its curvature by the relatively simple Robertson-Walker (*FLRW*) metric, with an expansion described by the Friedmann equation. Things would have been a lot more messy if it weren't homogenous and isotropic.

However, to see why the large scale homogeneity and isotropy of the Universe is so troubling in the standard Big Bang scenario, consider two antipodal points on the last scattering surface, as shown in Figure 11.1 in your text and reproduced below.

The current proper distance to the last scattering surface is

$$d_p(t_0) = c\int_{t_{\rm ls}}^{t_0} \frac{dt}{a(t)} \qquad (11.12)$$

Since the last scattering of the CMB photons occurred a long time ago ($t_{\rm ls} \ll t_0$), the current proper distance to the last scattering surface is only slightly smaller than the current horizon distance. For the Benchmark model, the current proper distance to the last scattering surface is $d_p(t_0) = 0.98\, d_{\rm hor}(t_0)$.



Thus, two antipodal points on the last scattering surface, separated by 180° as seen by an observer on Earth, are currently separated by a proper distance of $1.96\, d_{\rm hor}(t_0)$.

Since these two points are farther apart than the horizon distance, they are causally disconnected. That is, they haven't had time to exchange information. So, how is it that they have the same temperature to within one part in 10^5 (once the dipole distortion is subtracted from the CMB)?

Moreover, consider the following. In the standard Hot Big Bang scenario, the universe was matter-dominated at the time of last scattering, so the horizon distance at that time can be approximated by the relation we wrote in equation (5.58) in Lecture 9 for a flat, matter-only universe:

$$d_{\rm hor} = \frac{2c}{H(t_{\rm ls})} \qquad (11.13)$$

Since the Hubble distance at the time of last scattering was $c/H(t_{\rm ls})$, which we calculated to be equal to 0.2 Mpc in the last lecture, we get that the horizon distance at the time of last scattering was:

$$d_{\rm hor} \approx 0.4 \text{ Mpc}$$

This means that points more that 0.4 Mpc apart at the time of last scattering were not in causal contact in the standard Hot Big Bang scenario.

In the last lecture, we calculated that the angular diameter distance to the last scattering surface is $d_A \approx 13$ Mpc. Thus, points on the last scattering surface that were separated by a horizon distance will have an angular separation equal to

$$\theta_{\rm hor} = \frac{d_{\rm hor}(t_{\rm ls})}{d_A} \approx \frac{0.4 \text{ Mpc}}{13 \text{ Mpc}} \approx 0.03 \text{ rad} \approx 2^\circ \qquad (11.14)$$

as seen from the Earth today.

Equation (11.14) tells us that points on the last scattering surface separated by an angle as small as $\sim 2^\circ$ were out of contact with each other at the time the temperature fluctuations were imprinted on the CMB.

Yet we find that the temperature fluctuations $\delta T/T$ are as small as 10^{-5} on angular scales $\theta > 2^\circ$, corresponding to $l < 100$ in the CMB power spectrum discussed in the last lecture.

This is a remarkable coincidence. How can points that are not in causal contact know that they should synchronize?

In the next class, we will discuss the inflationary model and see how it can solve the flatness and horizon problems.