Physics 432/532: Cosmology
Winter 2021-2022
Prof. Michael S. Vogeley

Lecture Notes #3: Hot Big Bang

Concepts for this section of course:

- thermodynamics
- recombination
- microwave background
- primordial nucleosynthesis

We've now established that the universe is indeed expanding, that it looks roughly homogeneous on large enough scales, and that the age predicted by models that are solutions to Einstein's equations. So, what are the implications of this expansion for very early times in the universe?

1 Thermodynamics

To a high degree of precision, we can treat the expansion of the universe as being adiabatic. As we'll see, the entropy of the universe is dominated by the entropy of the background radiation, which is isentropic. For an adiabatic process, $TV^{\gamma-1} = \mathrm{const}$, thus for the universe,

$$T \propto R^{-3(\gamma-1)}$$

For radiation, the ratio of specific heats $\gamma = 4/3$, thus $T \propto 1/R$.

We divide the energy content of the universe into matter that behaves like "pressureless dust" and relativistic particles (photons, neutrinos). Following the previous equation, the energy density of the relativistic content then evolves as

$$\rho_r \propto R^{-4}$$

Earlier we derived this relation simply by scaling photon energies by R^{-1}. Obviously, the matter density evolves as

$$\rho_m \propto R^{-3}$$

We discussed earlier in the course that there is a crucial epoch in the history of universe, when the density of matter and radiation are equal. Before this time radiation dominates, after it matter dominates and structure can begin to form. This occurs at redshifts

$$1 + z_{eq} = 23900\Omega_{matter}h^2(T/2.73K)^{-4}$$

where T is the current temperature of the CMB. Start thinking about how to compute this; you'll have it as a homework or exam problem.

1.1 Quantum gravity limit

Extrapolating backwards to $T \to \infty$ as $R \to 0$, we anticipate that there is some epoch when classical physics breaks down and our general relativistic solution for the history of the universe is no longer valid. The energy density becomes so large that quantum fluctuations in this energy cause fluctuations in space curvature that make a mess of the spacetime metric (mini black holes winking in and out of existence, etc.). This happens when the effective wavelength (the de Broglie wavelength) of particles

$$\lambda_Q = \frac{2\pi\hbar}{mc}$$

becomes comparable to their Schwarzschild radius,

$$R_S = \frac{2Gm}{c^2}$$

Roughly equating these scales yields the mass scale on which quantum gravity becomes important, which we call the Planck mass,

$$m_P \equiv \sqrt{\frac{\hbar c}{G}} \approx 10^{19}\,\mathrm{Gev}$$

(Reminder of mass scale in energy units: electron mass is 0.511 MeV. As a temperature, GeV$\sim 10^{13}$K.) Related to this are the Planck length

$$l_P \equiv \sqrt{\frac{\hbar G}{c^3}} \approx 10^{-35}\,\mathrm{m}$$

and Planck time

$$t_P \equiv \sqrt{\frac{\hbar G}{c^5}} \approx 10^{-43}\,\mathrm{s}$$

This is a course in "classical" cosmology insofar as we will not even attempt to deal with the Planck epoch.

1.2 Early thermal equilibrium?

Much of cosmology relies on the assumption that the universe was in thermal equilibrium at early epochs. Densities were high, but the universe was also expanding at a higher rate. The rate for two-body processes is proportional to $n_1 n_2$ and densities scale as $n \propto R^{-3}$, thus the rate for two-body processes scales as

$$\Gamma \propto n_1 n_2 \propto R^{-6}$$

This strong increase in rate implies that thermal equilibrium is a good approximation at early times.

1.3 Entropy of the background

Recall the first law of thermodynamics,

$$dE = TdS - PdV$$

The entropy

$$S = \frac{E + PV}{T}$$

of the radiation background has entropy density (recall that radiation has equation of state $p = \rho c^2/3 = \epsilon/3$)

$$s = \frac{\epsilon + P}{T} = \frac{4\epsilon}{3T}$$

where for blackbody radiation

$$\epsilon = aT^4$$

with $a = 7.56 \times 10^{-16}\,\mathrm{J\,K^{-1}\,kg^{-1}}$ and $T = 2.728$ K for the Cosmic Microwave Background. (Note: a is related to the Stefan-Boltzman constant σ by $\sigma \equiv ac/4$.) Comparing the entropy density to the matter density in terms of Ω and h, we see that the entropy per unit mass is

$$\frac{S}{M} = 1.09 \times 10^{12} (\Omega_{matter} h^2)^{-1}\,\mathrm{J\,K^{-1}\,kg^{-1}}$$

By comparison $S/M = 14$ for water at $T = 300$ warmed by one degree. This huge entropy in the universe means that non-reversible thermodynamic processes are practically negligible and we can think of the expansion as being isentropic, thus adiabatic.

1.4 Neutrino background and massive neutrinos

In addition to a thermal background of photons, there is an isotropic flux of neutrinos left over from the early universe. At a redshift $z \sim 10^{10}$, neutrinos decoupled from other components of the mass-energy, as the rate of weak force interactions dropped below the expansion rate $\Gamma_{weak} < H$. Scaling from the observed photon background, we infer that this neutrinos background must have temperature

$$T_\nu = (4/11)^{1/3} T_\gamma = 1.95\,\mathrm{K}$$

These neutrinos were once ultrarelativistic, but are now too low in energy to allow direct detection. But if neutrinos have mass, we would see their gravitational effects. The cosmic density of neutrinos is $133\,\mathrm{cm}^{-3}$ per species. These neutrinos would have a contribution to the density parameter

$$\Omega_\nu h^2 = \frac{\sum m_i}{93.5\,\mathrm{eV}}$$

For $h = 0.68$ neutrinos would close the universe ($\Omega_\nu = 1$) if the sum of the masses were about 43 eV – this could be a single massive species or a combination of masses of the three species. Recall that there is a neutrino for each lepton, e, μ, τ, thus ν_e, ν_μ, ν_τ. For three equal neutrinos masses, $\bar{m}_\nu \approx 14$ eV would close the universe.

There is evidence for oscillation of neutrinos between different "flavor states" $(\nu_e, \nu_\mu, \nu_\tau)$, each of which is a superposition of mass eigenstates (m_1, m_2, m_3). A "holy grail" of neutrino physics is to measure the matrix that describes the mixing between flavor and mass eigenstates. Neutrino oscillation observations are sensitive to differences of the squares of the masses, which provides constraints on the possible spectrum of masses; a mass difference implies a non-zero mass for the two species that mix states. Atmospheric neutrino oscillation experiments provide evidence for a mass difference $(m_3^2 - m_2^2)c^2 \approx 2.4 times 10^{-3}\,\mathrm{eV}^2$. This result is consistent with oscillation between ν_μ and ν_τ. Solar neutrino data show a mass difference $(m_2^2 - m_1^2)c^2 \sim 7.5 \times 10^{-5}\,\mathrm{eV}^2$. Combining this information yields a lower bound on the sum of neutrino masses

$$(m_1 + m_2 + m_3)^2 \geq 0.057 eV$$

assuming that $m_1 \geq 0$.

For a short time in the early 1980's, there was great excitement when one experiment showed $m \sim 10$ eV for the electron neutrino. Many cosmologists considered what the implications of this "hot dark matter" (HDM) would be. "Hot" because it was relativistic at the epoch of structure formation. "Dark" because it is only weakly interacting thus causes no detectable radiation. Observations of structure in the universe rule out a large contribution of neutrinos to the mass-energy density because free-streaming of relativistic neutrinos suppresses early formation of galaxies, thus even a modest density of relic neutrinos would reduce the number of high-redshift galaxies below what we observe. We'll discuss this later in the course. The current upper bound from large-scale structure is

$$(m_1 + m_2 + m_3)c^2 \leq 0.3 eV$$

which is far below the mass derived above required to close the universe. At the maximum value, $\Omega_\nu = 0.007$, which makes neutrinos only a tiny fraction of the mass density (compare to $\Omega_{matter} = 0.3$).

2 Recombination

When the universe cools to a few thousand degrees $T \approx 3000$ K, hydrogen can form because the rate of photoionization drops below the formation rate. The rate equations for this are quite complicated because the hydrogen must cool to its ground state by a forbidden transition, 2S $\rightarrow$ 1S, which involves two-photon emission (remember why from the electric dipole selection rules you learned in QM).

The formation of hydrogen does not occur all at once. The ionization fraction decreases over a finite length of time, which corresponds to a finite range in the size of the universe and finite thickness to the shell of matter from which the photons emerge. The optical depth to Thomson scattering, which coupled the photons and electrons, is

$$\tau(z) = 0.37(z/1000)^{14.25}$$

Recall the definition of optical depth: $\tau \equiv \alpha ds$, where α is the absorption coefficient and τ is the e-folding factor for absorption. This is a very steep dependence of optical depth on redshift. Thus, the optical depth dramatically increases just beyond a redshift of $z = 1000$. If we work out the where most of the photons come from, they emerge from a mean redshift $z = 1065$ (where $\tau = 1$), with standard deviation of 80.

[Draw schematic of shell and radiation with different T from different R(t).]

Thus, the photons emerge from a shell with finite thickness, the "last-scattering surface" which is not a surface at all. This shell lies at fixed mean redshift, but the physical location of this region moves away from us as the universe expands, thus over time the region of the universe within the "shell" of last scattering changes (though certainly too slowly for us to observe).

The emergent radiation must have a blackbody spectrum, since the matter was in thermal equilibrium. Note that the radiation emitted at larger redshift had higher temperature. But the expansion of the universe causes the energies of those photons to shift downwards to exactly match the temperature at later times and smaller redshifts. Thus, even though the CMB is a superposition of blackbody spectra, those spectra exactly line up. This is important: the radiation must come from a shell in which the temperature scales exactly as $T \propto 1/R(t)$ or we would not observe a perfect blackbody.

[Draw nested $B_\nu(T)$.]

Notes on blackbody spectrum: Rayleigh-Jeans limit $B_\nu \propto \nu^2 kT$, Wien tail $B_\nu \propto \nu^3 \exp(-h\nu/kT)$.

The finite thickness of the shell also means that surface brightness fluctuations in this radiation will be suppressed on small scales because the radiation we see is an average over regions within the shell. On the other hand, fluctuations comparable to the thickness of the shell will be observable, with the strong imprint of the inhomogeneity of the universe at that epoch.

[Draw shells with inhomogeneities of different size, angle.]

As we discussed in detail with respect to galaxy observations, the relationship between proper size and angle on the sky depends on the cosmological parameters. A bit later in the course we will discuss in detail how the anisotropy of the CMB can be used to constrain cosmological models.

2.1 Detection of the microwave background

The existence of the microwave background had been predicted by Gamow, Alpher, and Bethe based on nucleosynthesis arguments in 1946. They predicted $T \sim 4-5$ K. In the 1960's, a group at Princeton also predicted this background and set to work building a detector on the roof of the geology building. But it was Arno Penzias and Robert Wilson at Bell Labs in 1965 who accidently discovered this radiation while searching the sky for possible sources of radio emission that would interfere with communications, and later won the Nobel Prize for this discovery. To quote Bob Dicke, leader of the Princeton group, "boys, we've been scooped."

In 1989, NASA launched a satellite to examine this background in greater detail. The COBE measured the CMB to have a temperature of

$$T = 2.728 \pm 0.004\,\mathrm{K}$$

When COBE pointed at regions in the sky that are not contaminated by signal from our own Galaxy, it was shown to be a perfect thermal source. The uncertainty in this measurement is limited by the ability to make a comparison source that is a perfect blackbody.

[Figures 6.1, 6.2 from Peebles]

The precision to which the CMB is thermal rules out many alternative sources for the CMB (other than the model described above) and rules out certain models for structure formation. For example, if there are hot electrons from ionized gas along the path of the CMB photons, then the shape of the spectrum will be altered as the photons Compton scatter off the electrons, thus transfering photons from low to high energies in the spectrum (depopulate the Rayleigh-Jeans regime into the Wien tail). The fractional energy change of the photons is characterized by the "Compton y-parameter,"

$$y \equiv \int \sigma_T n_e \frac{kT}{m_e c^2} dl$$

which is integrated over the path of the photons. The COBE observations set a limit $y < 1.5 \times 10^{-5}$.

Another effect on the spectrum is due to the Doppler shift cauesd by motion of the Earth relative to the CMB. Note that the CMB does seem to provide an absolute reference frame for motion in the universe. It was emitted at a particular epoch in time, when the universe had a particular scale size. Thus, sitting still in comoving space, the radiation appears to have the same temperature everywhere – it must because only the expansion factor of the universe has affected it. But, if we are moving with respect to comoving coordinates, in other words if we have a "peculiar velocity," then we will see a Doppler shift of the CMB: hotter in the direction of our motion, cooler behind.

Motion relative to the CMB simply shifts the observed temperature of the blackbody radiation: every photon shifts in frequency but the thermal spectrum is maintained. (You worked this out in the second problem set.) If we move at velocity v with Lorentz factor $\gamma = (1-(v/c)^2)^{-1/2}$, then the observed temperature on the sky at angle θ from the direction of our motion is

$$T_{obs} = \frac{T_0}{\gamma[1-(v/c)\cos\theta]}$$

We can expand this in multipole moments

$$T_{obs} = T_0[1 + \frac{v}{c} + \frac{1}{2}\left(\frac{v}{c}\right)^2 + (v^3)]$$

respectively monopole, dipole, quadrupole... The dipole has been detected. The quadrupole effect is difficult to disentangle from other possible cosmological effects.

The observed dipole implies a motion of the Earth with velocity

$$v = 371 \pm 1\,\mathrm{km\,s^{-1}}$$

in the direction $(l, b) = (264^\circ, 48^\circ)$, where (l, b) are Galactic coordinates. This is the velocity at which Earth is speeding through this "absolute" rest frame! If we take into account that the Earth is going around the Galaxy, and that our Galaxy is moving relative to our local group of galaxies, then we infer that the whole local group is moving with respect to the CMB at an even higher velocity

$$v_{LG} \sim 600\,\mathrm{km\,s^{-1}}$$

toward $(l, b) = (270^\circ, 30^\circ)$. This higher velocity results because the Earth's motion with respect to the local group is in nearly the opposite direction, and so partially cancels, the motion of the local group.

[Draw schematic of Earth's motion within LG, LG motion wrt CMB.]

Define Galactic coordinates: l is axial coordinate, with $l = 0$ toward center of Galaxy, b is elevation above/below Galactic plane.

How do we explain such a large velocity? It must be due to the gravitational attraction of large scale structure in the universe. Finding the source or sources of this attraction is a key part of our mapping of the distribution of galaxies in the universe, which we'll discuss later in the course.

3 Nucleosynthesis

At early enough times, the weak force allows conversions between protons and neutrons with reactions

$$p + e \leftrightarrow n + \nu$$

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

In equilibrium, the relative abundance of protons and neutrons varies as

$$\frac{N_n}{N_p} = e^{-\Delta mc^2/kT}$$

Detailed examination of the nuclear reactions shows that neutrons "freeze out," i.e. are no longer transmuted into protons and vice versa, at temperature $T = 10^{10.142}$ K. This yields a neutron to proton ratio $N_n/N_p \approx 0.34$.

If all the neutrons go into forming He, then the density of hydrogen is set by putting all the remaining protons into H. Assuming only H and He, we thus obtain an equation for the *mass* fraction Y of He,

$$Y = \frac{\rho_{He}}{\rho_{baryons}} = 1 - \frac{n_p - n_n}{n_p + n_n} = 2\left(1 + \frac{n_p}{n_n}\right)^{-1}$$

For $n_n/n_p = 0.34$, this gives $Y = 0.28$, very close to what is observed.

The next significant nucleus left over from the big bang is deuterium, which forms from the neutrons that fail to form He. The relevant reactions are

$$\mathrm{D} + p \rightarrow {}^3\mathrm{He}$$

followed by

$${}^3\mathrm{He} + n \rightarrow {}^4\mathrm{He}$$

But some D escapes, plus a small amount of $^3\mathrm{He}$. There are also trace quantities of Lithium ($^7\mathrm{Li}$) and Berylium ($^7\mathrm{Be}$).

The sense of the dependence on baryon density is that the *fraction* of mass in deuterium and $^3\mathrm{He}$ goes down as the baryon density increases. In other words, a high $\mathrm{D/H}$ ratio implies a low baryon density. Observers typically try to measure D/H by looking at absorption lines of these isotopes. Theoretical predictions for the ratio $(\mathrm{D} +^3 \mathrm{He})/\mathrm{H}$ are usually parametrized by the baryon to photon ratio, which is the inverse of the "entropy per baryon,"

$$\eta \equiv (n_p + n_n)/n_\gamma = 2.74 \times 10^{-8}(T/2.73K)^{-3}\Omega_b h^2$$

Thus, if we measure the relative abundance of deuterium, we can infer the baryon density.

Recent work on fitting the observed abundances of H, He, D, and Li yields a baryon density

$$\Omega_B h^2 = 0.020 \pm 0.002$$

which implies $\Omega_B \sim 0.04$ for $h \sim 0.7$. *The density of baryons is clearly quite far from critical density.*

4 Baryogenesis

Why is there more matter than antimatter? At early times, the thermal background is hot enough to produce copious number of particles and antiparticles in exact numerical balance. But at lower energies, these annihilate. Unless there are anti-matter galaxies hidden in the universe (which seems unlikely, since there is non-zero density of IGM gas that would collide with these galaxies), we need to explain why the early universe produce roughly $1 + \mathcal{O}(10^{-9})$ protons for every antiproton. Similar difficulties arise for the other particles.

Briefly, Grand Unified Theories of particle physics might provide an answer by allowing decay processes that do not conserve baryon or lepton number. The gauge bosons of GUTs mediate interchange between baryons and leptons. These bosons might decay to produce extra leptons or baryons, e.g.,

$$X^{4/3} \to e^+ + \bar{d}$$

$$X^{4/3} \to 2u$$

Such a decay fulfills one of three conditions necessary for particle-antiparticle assymetry to arise

- $\Delta B \neq 0$ (non-conservation of baryons)
- CP violation (different rates for particle-antiparticle and mirroring, e.g. right vs. left hand particles)
- non-equilibrium conditions

The second condition is necessary so that antiparticles are not produced at same rate as particles ($e + \bar{d}$ vs $e^+ + d$) The last condition is required so that the reverse reactions do not immediately erase the baryon asymetry.

Again, similar concerns arise for the asymmetry of leptons, e.g., e vs. e^+. And the asymmetry in leptons must be *exactly* the same as that in baryons, else there would be more/less protons than electrons and the universe would have a net charge. Again, GUTs provide such a mechanism because, although B and L are not individually conserved, the difference $B - L$ is conserved.