Lecture notes for Cosmology (ns-tp430m)

by Tomislav Prokopec

Part II: The Standard Cosmological Model

A. The Friedmann-Lamaître-Robertson-Walker (FLRW) cosmology

1. The metric tensor

We now consider the most general homogeneous space times, the Friedmann-Lemaître-Robertson-Walker (FLRW) expanding universes, and their ramifications. When written in spherical coordinates, the corresponding line element is of the form,

$$ds^2 = c^2dt^2 - a^2\frac{dr^2}{1-kr^2} - a^2r^2(d\theta^2 + \sin^2(\theta)d\varphi^2)\,, \tag{1}$$

where $a = a(t)$ is the scale factor, and r, θ and φ are the spherical coordinates.

When $k > 0$, the spatial sections of the space-time (1) are positively curved (spherical geometry), when $k < 0$ the spatial sections are negatively curved (hyperboloidal geometry), and finally when $k = 0$, the spatial-sections are flat (flat geometry).

In order illustrate the meaning of k, we now consider a two dimensional sphere S^2. When embedded into a 3-dimensional flat space, the equation characterising S^2 (with the origin placed at $\vec{r}_0 = (x_0, y_0, z_0) = 0$) is,

$$x^2 + y^2 + z^2 = R^2_{\rm curv}\,, \tag{2}$$

where $R_{\rm curv}$ denotes the radius (of curvature) of the sphere. The line element in this three dimensional Euclidean flat space is simply,

$$d\vec{\ell}^2 = dx^2 + dy^2 + dz^2\,. \tag{3}$$

Imagine now that we live on the surface of the sphere, and we would like to describe our position on the sphere. To this purpose, it is convenient to express dz in terms of dx and dy. By taking a differential of Eq. (2), we immediately arrive at,

$$dz^2 = \frac{(xdx + ydy)^2}{R^2_{\rm curv} - x^2 - y^2}\,. \tag{4}$$

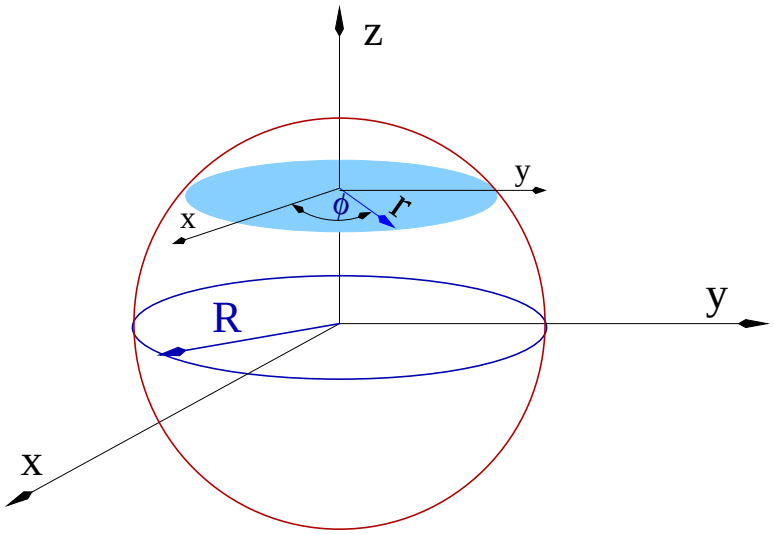


FIG. 1: The sphere S^2 embedded in a 3-dimensional flat Euclidean space, used to illustrate the curvature k of the FLRW space-time, whose line element is given in (1).

Next, it is convenient to define the spherical coordinates, which characterise the section of the sphere whose radius is, $r = (R^2_{\rm curv} - z^2)^{1/2}$, as shown in figure 1,

$$\begin{aligned} x &= r\cos(\theta) \\ y &= r\sin(\theta)\,. \end{aligned} \tag{5}$$

Taking a differential of these equations, we find,

$$x dx + y dy = r dr\,, \tag{6}$$

which, when used in (4), allows us to recast the line element (3) in spherical coordinates as,

$$d\vec{\ell}^2 = \frac{dr^2}{1 - r^2/R^2_{\rm curv}} + r^2 d\theta^2\,. \tag{7}$$

Elevating this discussion to one dimension higher ($S^2 \to S^3$), we conclude that $k > 0$ in the FLRW line element (1) has the following simple interpretation in terms of the curvature radius of the space,

$$k = \frac{1}{R^2_{\rm curv}}\,. \tag{8}$$

This simple interpretation holds of course only for spaces of constant spatial curvature. More generally, when $k = k(x)$, k signifies the local curvature of spatial sections of the space.

For negatively curved spaces, one should exact the replacement $z^2 \to -z^2$ in Eq. (2), which replaces S^2 by a 2-dimensional hyperboloid. After repeating the above procedure, one arrives at the conclusion that, in space-times with negatively curved spatial sections ($k < 0$), k and the radius of curvature $R_{\rm curv}$ are related as,

$$k = -\frac{1}{R^2_{\rm curv}}\,. \tag{9}$$

2. The Levi-Cività connection

The nonvanishing elements of the metric tensor corresponding to the FLRW line element (1) are,

$$g_{00} = 1\,, \qquad g_{rr} = -\frac{a^2}{1-kr^2}\,, \qquad g_{\theta\theta} = -a^2 r^2\,, \qquad g_{\varphi\varphi} = -a^2 r^2 \sin^2(\theta)\,. \tag{10}$$

Recall that the Levi-Cività connection is defined as,

$$\Gamma^{\mu}_{\alpha\beta} = \frac{1}{2} g^{\mu\nu}\Big(\partial_\alpha g_{\nu\beta} + \partial_\beta g_{\alpha\nu} - \partial_\nu g_{\alpha\beta}\Big)\,. \tag{11}$$

For the moment we shall distinguish the timelike and the spacelike elements of $\Gamma^{\mu}_{\alpha\beta}$. One easily checks that, $\Gamma^0_{00} = 0$, $\Gamma^i_{00} = 0$, and $\Gamma^0_{0i} = 0$, and the nonvanishing elements are,

$$\Gamma^i_{j0} \;=\; \frac{1}{c}\frac{\dot a}{a}\delta^i_{\ j} \tag{12}$$

$$\Gamma^0_{ij} \;=\; -\frac{1}{c}\frac{\dot a}{a} g_{ij} \tag{13}$$

$$\Gamma^i_{jl} \;=\; \frac{1}{2} g^{ik}\Big(\partial_j g_{kl} + \partial_l g_{jk} - \partial_k g_{jl}\Big)\,, \tag{14}$$

where $\partial_0 a = c^{-1} da/dt \equiv \dot a/c$.

The nonvanishing spatial components of the Levi-Cività connection (14) are,

$$\Gamma^r_{rr} \;=\; \frac{1}{2} g^{rr}\partial_r(g_{rr}) = \frac{kr}{1-kr^2} \tag{15}$$

$$\Gamma^r_{\theta\theta} \;=\; -r(1-kr^2) \tag{16}$$

$$\Gamma^r_{\varphi\varphi} \;=\; -r\sin^2(\theta)(1-kr^2) \tag{17}$$

$$\Gamma^\theta_{\theta r} \;=\; \frac{1}{r} \tag{18}$$

$$\Gamma^\varphi_{\varphi r} \;=\; \frac{1}{r} \tag{19}$$

$$\Gamma^\theta_{\varphi\varphi} \;=\; -\sin(\theta)\cos(\theta) \tag{20}$$

$$\Gamma^\varphi_{\varphi\theta} \;=\; \frac{\cos(\theta)}{\sin(\theta)}\,. \tag{21}$$

3. The curvature

The Ricci curvature tensor is the following contraction of the Riemann curvature tensor,

$$\mathcal{R}_{\alpha\beta} \equiv \mathcal{R}^{\mu}_{\ \alpha\mu\beta} = \partial_\mu \Gamma^\mu_{\alpha\beta} - \partial_\alpha \Gamma^\mu_{\mu\beta} + \Gamma^\mu_{\rho\mu}\Gamma^\rho_{\alpha\beta} - \Gamma^\mu_{\rho\beta}\Gamma^\rho_{\alpha\mu}\,. \tag{22}$$

The (00) component is then,

$$\begin{aligned}
\mathcal{R}_{00} \;&=\; \partial_\mu \Gamma^\mu_{00} - \partial_0 \Gamma^\mu_{0\mu} + \Gamma^\mu_{\rho\mu}\Gamma^\rho_{00} - \Gamma^\mu_{\rho 0}\Gamma^\rho_{0\mu} \\
&=\; -\partial_0\Big(\frac{1}{c}\frac{\dot a}{a}\delta^i_{\ i}\Big) - \Gamma^i_{j0}\Gamma^j_{0i} \\
&=\; -\frac{3}{c^2}\frac{\ddot a}{a}\,.
\end{aligned} \tag{23}$$

Similarly, one can show that the off-diagonal components vanish,

$$\mathcal{R}_{0i} = 0\,. \tag{24}$$

Next, the spatial components of the Ricci curvature tensor are,

$$\begin{aligned}
\mathcal{R}_{ij} &= \partial_\mu \Gamma^\mu_{ij} - \partial_j \Gamma^\mu_{i\mu} + \Gamma^\mu_{\rho\mu}\Gamma^\rho_{ij} - \Gamma^\mu_{\rho i}\Gamma^\rho_{j\mu} \\
&= \partial_0\Big(-\frac{1}{c}\frac{\dot a}{a}g_{ij}\Big) + \Big(-\frac{1}{c}\frac{\dot a}{a}g_{ij}\Big)\Big(\frac{1}{c}\frac{\dot a}{a}\delta^l_l\Big) - \Big(-\frac{1}{c}\frac{\dot a}{a}g_{lj}\Big)\Big(\frac{1}{c}\frac{\dot a}{a}\delta^l_i\Big) - \Big(\frac{1}{c}\frac{\dot a}{a}\delta^l_j\Big)\Big(-\frac{1}{c}\frac{\dot a}{a}g_{il}\Big) + {}^3\mathcal{R}_{ij} \\
&= -\frac{1}{c^2}\Big[\frac{\ddot a}{a} + 2\Big(\frac{\dot a}{a}\Big)^2\Big]g_{ij} + {}^3\mathcal{R}_{ij}\,,
\end{aligned} \tag{25}$$

where ${}^3\mathcal{R}_{ij}$ denotes the curvature of the spatial section of the FLRW space-time,

$${}^3\mathcal{R}_{ij} = \partial_l \Gamma^l_{ij} - \partial_i \Gamma^l_{jl} + \Gamma^l_{kl}\Gamma^k_{ij} - \Gamma^l_{ki}\Gamma^k_{jl}\,. \tag{26}$$

We now easily find for the Ricci scalar,

$$\mathcal{R} = \mathcal{R}_{00} + g^{ij}\mathcal{R}_{ij} = -\frac{6}{c^2}\Big[\frac{\ddot a}{a} + \Big(\frac{\dot a}{a}\Big)^2\Big] + {}^3\mathcal{R}\,, \tag{27}$$

where

$${}^3\mathcal{R} = g^{ij}\,{}^3\mathcal{R}_{ij} \tag{28}$$

denotes the Ricci curvature scalar of the spatial section of the space-time.

Next we need the nonvanishing elements of the spatial curvature (26) of the FLRW space-time. The (rr) component reads,

$$\begin{aligned}
{}^3\mathcal{R}_{rr} &= \partial_l \Gamma^l_{rr} - \partial_r \Gamma^l_{rl} + \Gamma^l_{kl}\Gamma^k_{rr} - \Gamma^l_{kr}\Gamma^k_{rl} \\
&= \partial_r\Big(\frac{kr}{1-kr^2}\Big) - \partial_r\Big(\frac{kr}{1-kr^2} + \frac{1}{r} + \frac{1}{r}\Big) + \frac{kr}{1-kr^2}\Big(\frac{kr}{1-kr^2} + \frac{1}{r} + \frac{1}{r}\Big) - \Big(\frac{kr}{1-kr^2}\Big)^2 - 2\Big(\frac{1}{r}\Big)^2 \\
&= \frac{2k}{1-kr^2}\,.
\end{aligned} \tag{29}$$

Similarly, the $(\theta\theta)$ and $(\varphi\varphi)$ components are found to be,

$$\begin{aligned}
{}^3\mathcal{R}_{\theta\theta} &= 2kr^2 \\
{}^3\mathcal{R}_{\varphi\varphi} &= 2kr^2\sin^2(\theta)\,,
\end{aligned} \tag{30}$$

and the off-diagonal elements vanish, ${}^3\mathcal{R}_{r\theta} = 0 = {}^3\mathcal{R}_{r\varphi} = {}^3\mathcal{R}_{\theta\varphi}$. All of the elements of the Ricci curvature tensor of spatial sections of the FLRW space-time can be compactly written as follows,

$${}^3\mathcal{R}_{ij} = -\frac{2k}{a^2}g_{ij}\,, \tag{31}$$

such that the corresponding Ricci scalar reads,

$$ {}^3\mathcal{R} = -\frac{6k}{a^2}\,. \tag{32} $$

With this and Eqs. (23–25), we can finally write the components of the Ricci curvature and Ricci scalar,

$$ \begin{aligned} \mathcal{R}_{00} &= -\frac{3}{c^2}\frac{\ddot{a}}{a} && (33)\\ \mathcal{R}_{ij} &= -\frac{1}{c^2}\Big[\frac{\ddot{a}}{a} + 2\Big(\frac{\dot{a}}{a}\Big)^2 + 2\frac{c^2k}{a^2}\Big]g_{ij} && (34)\\ \mathcal{R} &= -\frac{6}{c^2}\Big[\frac{\ddot{a}}{a} + \Big(\frac{\dot{a}}{a}\Big)^2 + \frac{c^2k}{a^2}\Big]\,. && (35) \end{aligned} $$

4. The FLRW equations

We now make use of the definition of the Einstein curvature tensor,

$$ G_{\mu\nu} = \mathcal{R}_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R}\,, \tag{36} $$

to find

$$ \begin{aligned} G_{00} &= \frac{3}{c^2}\Big[\Big(\frac{\dot{a}}{a}\Big)^2 + \frac{c^2k}{a^2}\Big]\\ G_{0i} &= 0\\ G_{ij} &= \frac{1}{c^2}\Big[2\frac{\ddot{a}}{a} + \Big(\frac{\dot{a}}{a}\Big)^2 + \frac{c^2k}{a^2}\Big]g_{ij}\,. && (37) \end{aligned} $$

The Einstein equations relate the geometry (curvature) of the space-time with the matter content,

$$ G_{\mu\nu} - g_{\mu\nu}\frac{\Lambda}{c^2} = \frac{8\pi G_N}{c^4}T_{\mu\nu}\,, \tag{38} $$

where $T_{\mu\nu}$ denotes the stress-energy tensor of matter fields. For an ideal fluid – which is characterised by spatial isotropy – stress energy tensor is of the form,

$$ T_{\mu\nu} = (\rho + \mathcal{P})\frac{u_\mu u_\nu}{c^2} - g_{\mu\nu}\mathcal{P}\,. \tag{39} $$

In the fluid rest frame, in which $u_\mu = c\delta_\mu{}^0$, this reduces to,

$$ T_{00} = \rho\,, \qquad T_{ij} = -\mathcal{P}g_{ij}\,, \tag{40} $$

where ρ and $\mathcal{P}$ denote the energy density and pressure of the fluid, respectively. The (00) and (ij) components of Eq. (38) are,

$$ \begin{aligned} 3\Big(\frac{\dot{a}}{a}\Big)^2 + 3\frac{c^2k}{a^2} - \Lambda &= \frac{8\pi G_N}{c^2}\rho\\ \Big[2\frac{\ddot{a}}{a} + \Big(\frac{\dot{a}}{a}\Big)^2 + \frac{c^2k}{a^2} - \Lambda\Big]g_{ij} &= \frac{8\pi G_N}{c^2}(-\mathcal{P}g_{ij})\,, && (41) \end{aligned} $$

which can be recast as the following FLRW equations,

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_N}{3c^2}\rho + \frac{\Lambda}{3} - \frac{c^2 k}{a^2} \tag{42}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3c^2}(\rho + 3\mathcal{P}) + \frac{\Lambda}{3}\,. \tag{43}$$

These equations govern the evolution of the Universe in the Standard Cosmological Model. The meaning and dimensions of the symbols are as follows: a denotes the scale factor (dimensionless), H [$\mathtt{s}^{-1}$] is the Hubble parameter, G_N [$\mathtt{m}^3\mathtt{kg}^{-1}\mathtt{s}^{-2}$] the Newton constant, Λ [$\mathtt{s}^{-2}$] the cosmological term, c [$\mathtt{ms}^{-1}$] the speed of light, ρ [$\mathtt{kgm}^{-1}\mathtt{s}^{-2}$] the energy density, $\mathcal{P}$ [$\mathtt{kgm}^{-1}\mathtt{s}^{-2}$] the pressure, and k [$\mathtt{m}^{-2}$] is the curvature of the spatial section of the space-time. Note that the only difference between the dynamics of a flat and curved FLRW universe is the curvature term, $-c^2k/a^2$, in the constraint equation (42).

The covariant conservation of the stress-energy tensor, $\nabla^\mu T_{\mu\nu} = 0$, yields in a FLRW background,

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + \mathcal{P}) = 0\,, \tag{44}$$

which is a consistency condition that can be also derived from Eqs. (42–43).

Finally, we note that equations (42–43) can be generalised to inhomogeneous space-times, simply by exacting the replacements,

$$a = a(x)\,, \quad H = H(x)\,, \quad k = k(x)\,, \quad \rho = \rho(x)\,, \quad p = p(x)\,, \tag{45}$$

and by adding the term $-(c^2/3)(\nabla^2\delta\mathcal{P})/(\rho+\mathcal{P})$ on the right hand side of Eq. (43), where $\delta\mathcal{P}/(\rho+\mathcal{P})$ is a presure deviation, which is related to the local spatial curvature, $\mathcal{R}$ (not to be confused with the Ricci scalar) as follows, $\dot{\mathcal{R}} = -H\delta\mathcal{P}/(\rho+\mathcal{P})$, $k = -(2/3)a^2\nabla^2\mathcal{R}$ (see *e.g.* Lyth and Riotto, Phys. Rept. **314** (1999) 1-146, [e-Print Archive: hep-ph/9807278]).

5. Matter inventory of the Universe

Depending on scaling properties with the expansion and the composition, the matter of the Universe can be conveniently divided into,

$$\rho = \sum_i \rho_i = \rho_{\rm dm} + \rho_{\rm b} + \rho_\gamma + \rho_{\rm Q} + \rho_\nu + \cdots\,, \tag{46}$$

such that Eq. (42) can be recast at present as,

$$1 = \Omega_{\rm dm} + \Omega_{\rm b} + \Omega_\gamma + \Omega_{\rm Q} + \Omega_\nu + \cdots + \Omega_\Lambda + \cdots + \Omega_k\,, \tag{47}$$

where

$$\Omega_i = \frac{\rho_i}{\rho_{\rm cr}}\,, \qquad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}\,, \qquad \Omega_k = -\frac{c^2 k}{H_0^2} \tag{48}$$

and

$$\rho_{\rm cr} = \frac{3c^2}{8\pi G_N} H_0^2 \tag{49}$$

denotes the critical energy density. One then defines matter energy density, $\Omega_{\rm m} = \Omega_{\rm dm} + \Omega_{\rm b}$ as a sum of dark matter and baryonic matter. Furthermore, one defines dark energy as, $\Omega_{\rm DE} = \Omega_\Lambda + \Omega_Q + \cdots$, as a sum of the energy density in the cosmological term, in Q-matter, *etc.* The energy density in neutrinos, Ω_ν and in photons, Ω_γ, are negligible today, when compared with the critical energy density, but they played an important role in the early Universe.

Since the uncertaintly in the Hubble parameter is relatively large, in order to reduce the uncertaintly, energy densities are ofen expressed multiplied by a suitable power of the Hubble parameter, whose value today is (WMAP and HST Key Project),

$$h = 0.71 \pm 0.02, \,, \tag{50}$$

where h is defined by, $H_0 = 100h$ kms^{-1}Mpc^{-1}, see Eq. (124).

In the following we quote densities as they are constrained by the WMAP measurements. The energy density in the baryonic matter of the Universe is,

$$\Omega_{\rm b} h^2 = 0.02265 \pm 0.00059 \,, \tag{51}$$

implying that protons and neutrons make up about 4-5% of the critical density of the Universe, $\Omega_{\rm b} = 0.045 \pm 0.003$, which also translates into a constraint on the baryonic density, $n_{\rm b} = 0.25 \pm 0.01$ m^{-3}, and on the baryon-to-photon ratio, $\eta \equiv n_{\rm b}/n_\gamma = 6.1 \pm 0.2 \times 10^{-10}$. The dark matter density $\Omega_{\rm dm}$, whose composition is not known today (except that it is not baryonic) is about 5 times larger, such that the total matter density,

$$\Omega_{\rm m} h^2 \equiv \Omega_{\rm dm} h^2 + \Omega_{\rm b} h^2 = 0.1369 \pm 0.0037 \,, \tag{52}$$

makes up about 27% of the critical energy density, $\Omega_{\rm m} = 0.27 \pm 0.02$. For neutrinos there is an upper bound,

$$\Omega_\nu h^2 < 0.0076 \ (95\% \, {\rm CL}) \,. \tag{53}$$

The density in the cosmic microwave background (CMB) photons today is given by,

$$\Omega_\gamma h^2 = \frac{\Omega_{\rm m} h^2}{1 + z_{\rm eq}} \simeq 4.3 \times 10^{-5} \,, \tag{54}$$

which implies a number density of photons today, $n_\gamma = 410.4 \pm 0.9$ cm^{-3}. From this and Eq. (52), we infer that the redshift at the matter and radiation equality is about, $z_{\rm eq} \simeq 3230 \pm 210$. Assuming that the temperature of cosmic background neutrinos is $T_\nu \simeq 1.96$ K, which comes out by a standard

calculation from thermal history of the Universe, one can calculate the density of massless neutrinos per species, $n_\nu \simeq 112\ cm^{-3}$.

Finally, the dark energy density, which dominates the Universe's energy today,

$$\Omega_{\rm DE} = 0.72 \pm 0.02\,. \tag{55}$$

While the precise composition of the dark energy is not known, it is known that its equation of state is consistent with the equation of state of the cosmological term, $w_\Lambda = p_\Lambda/\rho_\Lambda = -1$. More precisely, the bound is, $-1.3 < w_{\rm DE} < -0.8$ (95% CL).

6. The Universe is spatially flat

When all energy densities are summed up, one gets for the total energy density,

$$\Omega_{\rm tot} = \Omega_{\rm m} + \Omega_{\rm DE} = 1.01 \pm 0.01\,, \tag{56}$$

such that the current Universe is nearly spatially flat. To be more quantitative, we have,

$$1 - \Omega_{\rm tot} = \Omega_k = -\frac{c^2 k}{H_0^2}\,, \qquad |\Omega_k| < 0.02\,, \tag{57}$$

which then implies for the radius of curvature of the Universe, $R_{\rm curv} = |k|^{-1/2}$,

$$R_{\rm curv} > \frac{1}{\sqrt{|1-\Omega_{\rm tot}|}}\frac{c}{H_0}\,. \tag{58}$$

Taking account of the current Hubble radius, $R_H(t_0) = c/H_0 \simeq 4200 \pm 200$ Mpc, we get the following bound on the curvature radius of the Universe,

$$R_{\rm curv} \geq 30\ {\rm Gpc} \quad (R_{\rm curv} > 0) \qquad R_{\rm curv} \geq 40\ {\rm Gpc} \quad (R_{\rm curv} < 0)\,, \tag{59}$$

implying that $R_{\rm curv}$ is at least *six* times larger than the Hubble radius.

The radius of curvature of the FLRW metric (1) does not change with time, and yet its importance does chance bacause scale factor $a = a(t)$ depends on time. As a consequence the curvature term scales with time. Let us consider this question a little more closely. In matter era, when $a = (t/t_0)^{2/3}$ ($a_0 = 1$), $H = 2/(3t)$, we have

$$Ha = H_0\Big(\frac{t_0}{t}\Big)^{1/3} = \frac{H_0}{\sqrt{a(t)}}\,, \tag{60}$$

such that

$$\Omega_{\rm tot}(t) - 1 = -\Omega_k(t) = \frac{c^2 k}{H_0^2} a(t)\,. \tag{61}$$

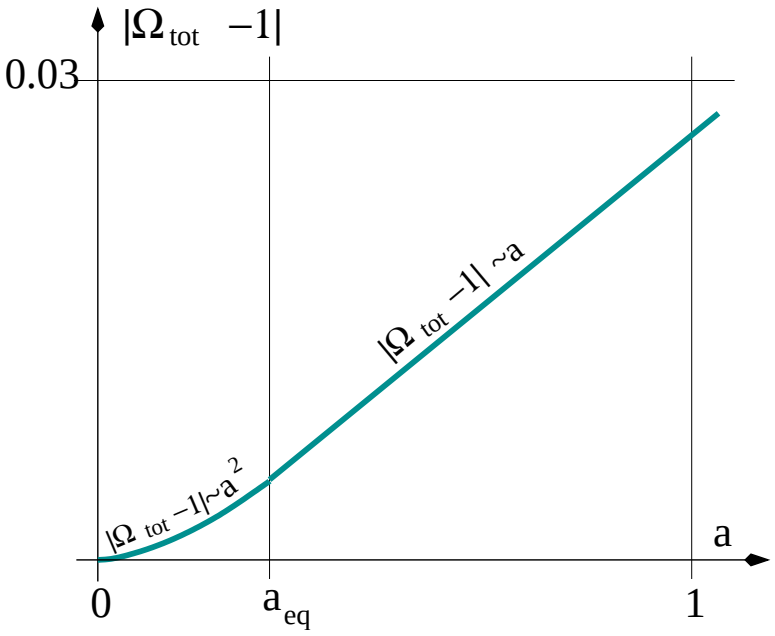


FIG. 2: The evolution of the curvature density $|\Omega_k| = |\Omega_{\rm tot} - 1|$ as a function of scale factor a. In matter era $|\Omega_k|$ evolves linearly with a, while in radiation era it evolves quadratically with a, approaching zero as $a \to 0$. The bound $|\Omega_{\rm tot} - 1| < 0.03$ at $a = a_0 = 1$ is also indicated.

Similarly, in radiation era, $a = a_{\rm eq}(t/t_{\rm eq})^{1/2}$, $H = 1/(2t)$, we have,

$$\Omega_{\rm tot}(t) - 1 = \frac{c^2 k}{H_{\rm eq}^2}\left(\frac{a}{a_{\rm eq}}\right)^2, \tag{62}$$

where $a_{\rm eq}$ and $H_{\rm eq}$ denote the scale factor and Hubble parameter at the matter-radiation equality, respectively. This implies that the deviation of the total energy density (curvature density) from unity was much smaller in the past than it is today. This is illustrated in figure 2. As one goes back in time towards the radiation-matter equality, $|\Omega_{\rm tot} - 1|$ decreases linearly with time. On the other hand in radiation era $|\Omega_{\rm tot} - 1|$ decreases quadratically with time, reaching $\Omega_{\rm tot} - 1 \to 0$ as $a \to 0$. This means that the deviation of $\Omega_{\rm tot}$ from unity is very small in the early Universe. This is known as the flatness problem, which is elegantly solved by cosmic inflation. Indeed, during inflation, $a = a_I {\rm e}^{H_I t}$, $H = H_I =$ const., such that the curvature term decreases exponentially with time during inflation,

$$\Omega_{\rm tot}(t) - 1 = \frac{c^2 k}{H_I^2}{\rm e}^{-2H_I t}, \tag{63}$$

explaning thus its smallness at the beginning of radiation era.

7. *Coincidence problems*

Let us consider once more the Friedmann equation (42), which on the right-hand-side contains several, at a first sight unrelated, quantities, the matter energy density ρ (which consists of several

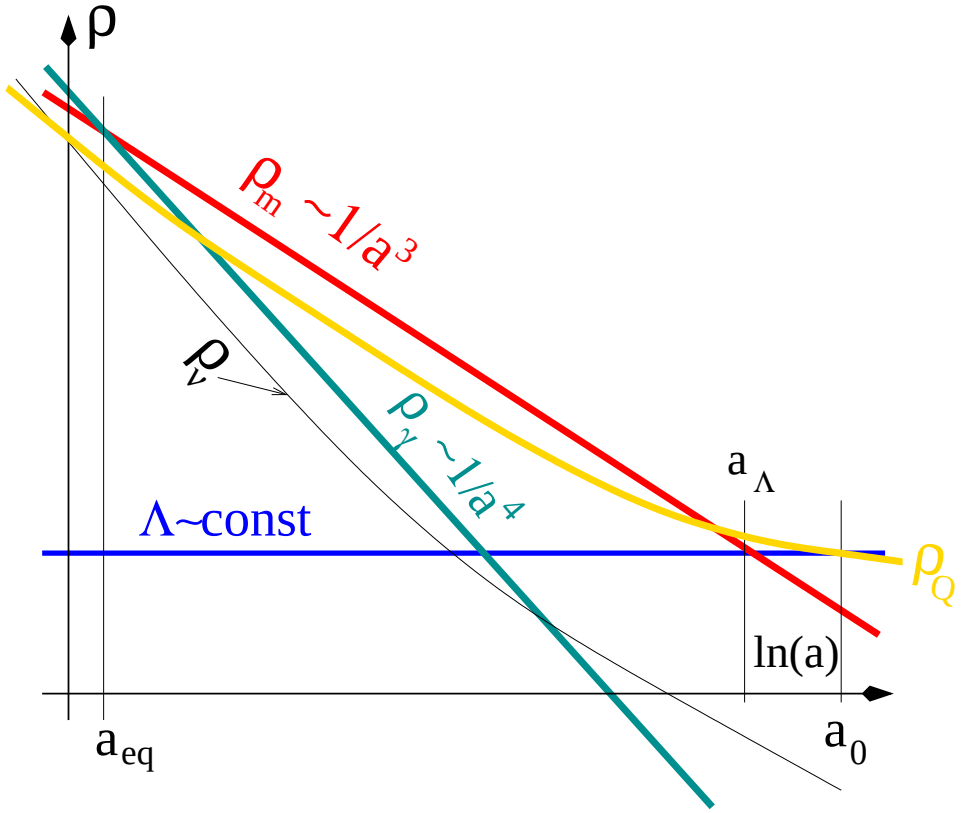


FIG. 3: The different components contributing to the Universe's energy density as a function of scale factor. a_0 denotes the scale factor today, $a_{\rm eq} = a_0/(3230 \pm 210)$ denotes the scale factor at the matter-radiation equality and $a_\Lambda \simeq 0.7 a_0$ is the scale factor at which $\rho_\Lambda = \rho_m$.

distinct contributions), the cosmological term Λ, and the curvature term, $-c^2k/a^2$. The scaling of these terms with the expansion of the Universe is very different. The matter density, which can be written as, $\rho = \rho_{\rm m} + \rho_\gamma + \rho_Q$, each of which scales as, $\rho_{\rm m} \propto 1/a^3$, $\rho_\gamma \propto 1/a^4$, $\rho_Q \propto 1/a^{3+w_Q}$, the cosmological term remains constant, $\Lambda =$ const., and the curvature term scales as $1/a^2$. The scalings of different terms contributing to the energy density of the Universe is shown in figure 3.

The first coincidence that may appear puzzling is, why is the total energy density, $\Omega_{\rm tot} = 1.01 \pm 0.02$, so close to the critical energy density. In other words, why is the radius of curvature of the Universe so big? This is elegantly explained by an epoch of primordial inflation, and hence it is not any more a source of great discomfort. This is sometimes referred to as the Dicke coincidence (1961).

The second coincidence, which is why $\Omega_{\rm DE} \sim \Omega_{\rm m}$ at present, has no natural explanation. Before the discovery that the Universe is accelerating, it was largely considered that $\Omega_\Lambda = 0$. The lack of explanation for why the matter vacuum fluctuations do not contribute to Λ was termed the cosmological constant problem. The situation at present is even more discomforting. Not only that we need an explanation for why the matter vacuum fluctuations do not contribute to Λ. We also need an explanation for what makes up more than 70% of the energy density of the Universe. Moreover, why is $\Omega_{\rm DE}$ so close

to $\Omega_{\rm m}$ today, a seemingly two completely unrelated quantities. The problem is further aggravated when one considers the scaling with the scale factor, as shown in figure 3. Let us assume that the dark energy is composed of the cosmological term. While the energy density in nonrelativistic matter and the cosmological term was equal relatively recently, $a_\Lambda \simeq 0.7$ ($z_\Lambda \simeq 0.4$), at equality, when $z_{\rm eq} \simeq 3230$, $\rho_{\rm m}/\rho_\Lambda \sim 10^{10}$; at even earlier epochs the ratio was only bigger.

A third coincidence, why is $\rho_{\rm dm} \sim 5\rho_{\rm b}$, is perhaps explained by the same origin of dark matter and baryonic matter. While neither the origin of dark matter, nor baryonic matter is known with certainly, there are competing explanations, some of them testable by near future experiments. A plausible explanation for the dark matter of the Universe are weakly interacting massive nonrelativistic particles. A popular candidate is the lightest supersymmetric particle, which is in many realisations of supersymmetry the neutralino, which is the common name for the supersymmetric partners of neutral Higgs particles and of neutral massive gauge bosons. On the other hand, it is known that the neutralino may participate in creating the baryonic matter of the Universe at the electroweak phase transition, representing thus the sought-for link. The future experiments will show whether the Nature has chosen to utilise this link.

There are further seeming coincidences, which are most likely just coincidences, requiring no deeper physical explanation. One of the curious facts is the time of equal matter and radiation, $z_{\rm eq} = 3230 \pm 210$, which is so close to decoupling, $z_{\rm dec} \simeq 1089$, which is the time when the photons decoupled from the rest of the matter, when the electrons and protons largely recombined into neutral hydrogen. Furthermore, one may wonder why have neutrinos become nonrelativistic only recently, *i.e.* why is the mass scale of the neutrinos, $m_\nu c^2$, so close to the temperature scale today, $k_B T_\nu$.

There are other curious coincidences in the physics of the early Universe, and we shall mention them in passing in the course of these lectures. The coincidences in cosmology have lead B. Carter to postulate the anthropic principle, according to which our existence has something to do with the choice of the fundamental constants in Nature, and hence with the way the Universe is. Its philosophical implications have captured a lot of attention in popular writings.

8. Horizons

Let us now consider a light ray moving along a geodesic, such that

$$ds = 0\,. \tag{64}$$

Next, let us choose a geodesic which, in spherical coordinates, passes through $r = 0$, along $\theta = \text{const.}$ and $\varphi = \text{const.}$ A useful example for such a geodesic in curved space is a great circle on a sphere, which

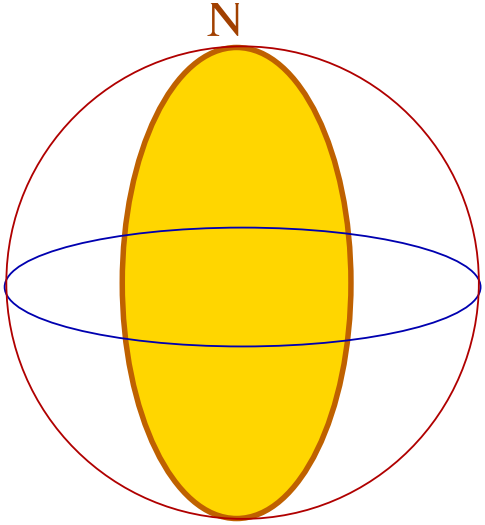


FIG. 4: An example of a geodesic on a positively curved space: a great circle of a sphere passing through the North Pole.

passes through the North Pole, an example of which is shown in figure 4.

For these types of geodesics we then have the following definition of horizon,

$$\ell_{\rm phys} = \int_{r_0}^{r_H} \sqrt{g_{rr}(r')}dr' \,. \tag{65}$$

Eq. (65) represents the *event horizon*, when the upper limit of integration, $r = r_H$, corresponds to some final time, $t_{\rm f}$, which is either $t_{\rm f} \to \infty$, or the time at which the geodesic stops, which may happen when the geodesic hits a curvature singularity. For example, for a closed universe, which eventually recollapses, the event horizon is determined by Eq. (65) evaluated at the time of collapse. On the other hand, Eq. (65) represents the *particle horizon*, when the upper limit of integration corresponds to the present time.

To get a clearer physical picture of particle horizons, let us consider the particle horizon of the FLRW space-time (1),

$$\ell_{\rm phys}(t) = a\int_0^{r_H} \frac{dr'}{\sqrt{1-kr'^2}} = a\int_0^t \frac{cdt'}{a(t')}\,, \tag{66}$$

where we chose $t_{\rm in} = 0$. One often defines the comoving particle horizon,

$$\ell_c(t) = \frac{\ell_{\rm phys}}{a} = c(\eta - \eta_{\rm in})\,, \tag{67}$$

where η denotes conformal time, defined by, $ad\eta = dt$.

Upon rewriting, $dt = da/\dot{a}$, and making use of the Friedmann equation (42), we get for the particle horizon (66) at time t ($a = a(t)$),

$$\ell_{\rm phys} = a\int_0^a \frac{cda'}{a'\dot{a}'} = \frac{ca}{H_0}\int_0^a \frac{da'}{\sqrt{\Omega_\gamma + \Omega_{\rm m}a' + \Omega_\Lambda a'^4 + \Omega_Q a'^{1-3w_Q} + \Omega_k a'^2}}\,, \tag{68}$$

where

$$\Omega_\gamma = \frac{\rho_\gamma(t_0)}{\rho_{\rm cr}}\,, \qquad \Omega_{\rm m} = \frac{\rho_m(t_0)}{\rho_{\rm cr}}\,, \qquad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}\,, \qquad \Omega_Q = \frac{\rho_Q(t_0)}{\rho_{\rm cr}}\,, \qquad \Omega_k = -\frac{c^2 k}{H_0^2}\,, \tag{69}$$

and $\rho_{\rm cr} = (3c^2H_0^2)/(8\pi G_N)$ denotes the critical energy density of the Universe at present, and we took account of the scalings of the different matter components,

$$\rho_\gamma(t) = \frac{\rho_\gamma(t_0)}{a^4}\,, \qquad \rho_m(t) = \frac{\rho_m(t_0)}{a^3}\,, \qquad \rho_Q(t) = \frac{\rho_Q(t_0)}{a^{3+3w_Q}}\,, \qquad \Lambda = \text{const.}\,, \tag{70}$$

and $w_Q = p_Q/\rho_Q$.

Eq. (68) represents a general expression for the particle horizon in FLRW space-times. We shall now evaluate it for some simple cases.

Consider first a matter dominated era, in which $\Omega_{\rm m} = 1$, and all other components vanish (recall that Eq. (42) implies that at present, $1 = \Omega_\gamma + \Omega_{\rm m} + \Omega_Q + \Omega_\Lambda + \Omega_k$). In this case Eq. (68) is easily integrated to give,

$$\ell_{\rm phys} = \frac{2ca^{3/2}}{H_0}\,. \tag{71}$$

Since in a matter dominated universe, $a \propto t^{2/3}$, such that the Hubble parameter at present, $H_0 = 2/(3t_0)$, the particle horizon (71) at present epoch equals

$$\ell_{\rm phys}(t_0) = 3ct_0 \qquad (\texttt{matter era})\,, \tag{72}$$

which is comparable to the Hubble radius,

$$R_H(t_0) \equiv cH_0^{-1} = \frac{1}{2}\ell_{\rm phys}(t_0)\,. \tag{73}$$

On the other hand, in radiation era, in which $\Omega_\gamma = 1$ and other components vanish, Eq. (68) integrates to,

$$\ell_{\rm phys} = \frac{c}{H_{\rm eq}}\Big(\frac{a}{a_{\rm eq}}\Big)^2 \qquad (\texttt{radiation era})\,, \tag{74}$$

where now $H_{\rm eq} = H(t_{\rm eq})$ is the Hubble parameter at matter-radiation equality, $\rho_m = \rho_\gamma$. Since during radiation era, $a = a_{\rm eq}\,(t/t_{\rm eq})^{1/2}$, which implies, $H_{\rm eq} = 1/(2t_{\rm eq})$, the particle horizon (74) at equality is equal to the Hubble radius,

$$\ell_{\rm phys} = 2ct_{\rm eq} = R_{\rm H}(t_{\rm eq})\,. \tag{75}$$

Consider now an epoch of cosmic inflation, during which the energy of the Universe is dominated by a cosmological term Λ (de Sitter inflation), such that the Hubble parameter (42) is given by $H_I = \sqrt{\Lambda/3}$ and scale factor expands exponentially, $a = a_I \mathrm{e}^{H_I t}$. In this case particle horizon (68) reads,

$$\ell_{\rm phys} = \frac{ac}{H_I}\int_{a_{\rm in}}^{a}\frac{da'}{a'^2} = \frac{c}{H_I}\Big(\frac{a}{a_{\rm in}} - 1\Big) \qquad (\text{de Sitter inflation})\,, \tag{76}$$

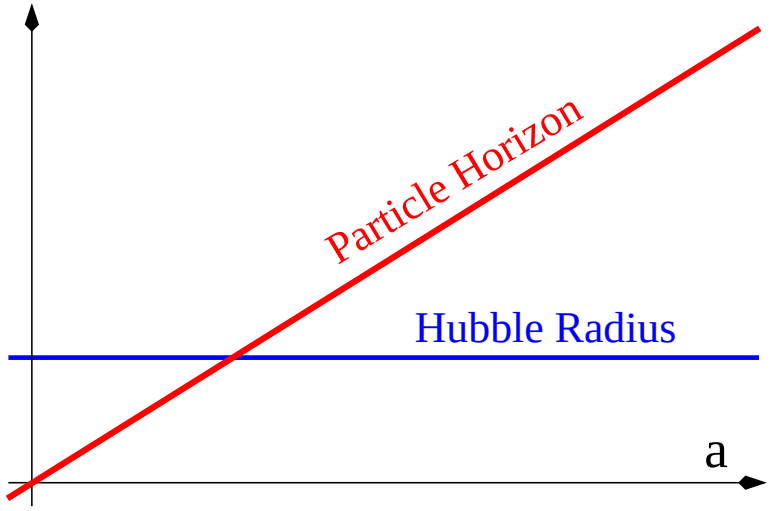


FIG. 5: Hubble radius and particle horizon during de Sitter inflation. While Hubble radius is constant $R_H = c/H_I$, particle horizon grows exponentially with time (linearly with scale factor $a = \mathrm{e}^{H_I(t-t_{\rm in})}$).

where $a_{\rm in} = a(t_{\rm in}) = a_I \mathrm{e}^{H_I t_{\rm in}}$. When compared with Hubble radius $R_H = c/H_I$, particle horizon grows exponentially large,

$$\frac{\ell_{\rm phys}}{R_H} = \mathrm{e}^{H_I t} - 1. \tag{77}$$

The growth of particle horizon as a function of scale factor during de Sitter inflation is shown in figure 5. Since particle horizon is defined as the distance traversed by photons, the regions emcompassed by particle horizon are in causal contact. Inflation is a special space-time in which causally related regions extend much beyond the Hubble radius. This property of inflation is used to resolve the flatness problem of the Universe, to be discussed in chapter on cosmic inflation. Note that particle horizon in eternal inflation ($t_{\rm in} \to -\infty$) diverges.

Alternatively, cosmic inflation may be driven by Q-matter (quintessence) with a negative equation of state, $w_Q = p_Q/\rho_Q < -1/3$, In this case, $H_Q = \dot{a}/a \propto a^{-(3/2)(1+w_Q)}$, such that $a \propto t^{2/(3+3w_Q)}$, and $H_Q = 2/[(3+3w_Q)t]$. With $-1 < w_Q < -1/3$, the Universe undergoes an accelerated expansion, $\ddot{a}/a = -2(1+3w_Q)/[9(1+w_Q)^2 t^2] > 0$. The particle horizon (68) of a universe dominated by quintessence reads,

$$\ell_{\rm phys} = \frac{ac}{H_Q a_{\rm in}^{(3+3w_Q)/2}} \int_{a_{\rm in}}^{a} da' a'^{-\frac{1}{2}+\frac{3}{2}w_Q} = -\frac{2c}{H_Q(1+3w_Q)}\left[\frac{a}{a_{\rm in}} - \left(\frac{a}{a_{\rm in}}\right)^{\frac{3(1+w_Q)}{2}}\right] \qquad \text{(Q inflation)}, \tag{78}$$

such that it grows linearly with scale factor and as a power of time t.

To a good approximation our Universe is spatially flat ($\Omega_k \approx 0$) and filled with nonrelativistic matter with $\Omega_{\rm m} \approx 0.26$ and dark energy, whose composition resembles that of the cosmological term, $\Omega_\Lambda \approx 0.74$. In this case particle horizon (68) reduces to,

$$\ell_{\rm phys}(t_0) = \frac{ac}{H_0} \int_0^1 \frac{da'}{\sqrt{\Omega_{\rm m} a' + \Omega_\Lambda a'^4}} \tag{79}$$

which can be evaluated in terms of elliptical integrals.

In an open universe, with $\Omega_k = 1 - \Omega_{\rm m}$, we have

$$\ell_{\rm phys} = \frac{ac}{H_0}\int_0^a \frac{da'}{\sqrt{\Omega_{\rm m}a' + \Omega_k a'^2}} \approx \frac{ac}{H_0}\frac{2}{\sqrt{\Omega_k}}\ln\Big(\frac{\sqrt{\Omega_k a} + \sqrt{\Omega_{\rm m} + \Omega_k a}}{\sqrt{\Omega_{\rm m}}}\Big)\,. \tag{80}$$

At the present epoch $a = a_0 = 1$ and when expanded in powers of Ω_k, Eq. (80) simplifies to,

$$\ell_{\rm phys}(t_0) = \frac{c}{H_0}\frac{1}{\sqrt{\Omega_k}}\ln\Big(\frac{1+\sqrt{\Omega_k}}{1-\sqrt{\Omega_k}}\Big) \simeq \frac{2c}{H_0}\Big(1 + \frac{1}{3}(1-\Omega_{\rm m}) + \mathcal{O}(\Omega_k^2)\Big)\,. \tag{81}$$

With $\Omega_{\rm m} \simeq 0.3$, $\Omega_k \simeq 0.7$, we get, $\ell(t_0) \simeq 2.89c/H_0$.

9. Do we live in an open, closed, or flat universe?

The short answer to this question is: we do not know, but we do know that the Universe is awfully close to being spatially flat, since its curvature radius (59) is greater than about 23 Gpc. In short,

$$\Omega_{\rm tot}\begin{cases} > 1 : \texttt{CLOSED} \\ = 1 : \texttt{FLAT} \\ < 1 : \texttt{OPEN} \end{cases}\,. \tag{82}$$

The three different geometries are illustrated in figure 6. In a closed (positively curved) universe the sum of the angles in a (large) triangle is greater than 180^o, in a flat Universe the sum equals 180^o, and in an open universe the sum is smaller than 180^o.

A nearly flat Universe is beneficial for formation of (intelligent) life, since it lasts long. In order to see this, let us integrate the Friedmann equation (42),

$$\int_0^a \frac{da'}{\sqrt{\frac{8\pi G_N}{3c^2}\rho a^2 + \frac{\Lambda}{3}a^2 - c^2 k}} = t\,, \tag{83}$$

where we took $t_{\rm in} = 0$, $a_{\rm in} = a(t_{\rm in}) = 0$. Equivalently, Eq. (83) can be recast as ($a_0 = 1$),

$$\int_0^a \frac{da'}{\sqrt{\Omega_{\rm m}a'^{-1} + \Omega_\gamma a'^{-2} + \Omega_Q a'^{-1-3w_Q} + \Omega_\Lambda a'^2 + \Omega_k}} = H_0 t\,. \tag{84}$$

This can be integrated in some simple cases. For example, in a universe dominated by nonrelativistic matter we have $\rho = \rho_m/a^3$, $\Lambda = 0$, $\Omega_{\rm m} + \Omega_k = 1$. In an open Universe, $\Omega_k > 0$ ($k < 0$), and we find

$$\int_0^a \frac{da'}{\sqrt{\Omega_{\rm m}a'^{-1} + \Omega_k}} = \frac{1}{\Omega_k}\left\{a\sqrt{\Omega_{\rm m}a^{-1} + \Omega_k} - \frac{\Omega_{\rm m}}{2\sqrt{\Omega_k}}\ln\left(\frac{\sqrt{\Omega_{\rm m}a^{-1} + \Omega_k} + \sqrt{\Omega_k}}{\sqrt{\Omega_{\rm m}a^{-1} + \Omega_k} - \sqrt{\Omega_k}}\right)\right\} = H_0 t\,, \tag{85}$$

which is shown in figure 7. The two curves represent: (a) $\Omega_{\rm m} = 0.26$ (upper red solid curve) and (b) $\Omega_{\rm m} = 1$ (lower green dashed curve). Note that the age of the Universe today is, $t_0 \simeq 0.83H_0^{-1}$ (Case (a)) and $t_0 = (2/3)H_0^{-1}$ (Case (b)).

FIG. 6: The spatial geometry of a homogeneous open, flat and closed Friedmann-Lemaître-Robertson-Walker (FLRW) Universe.

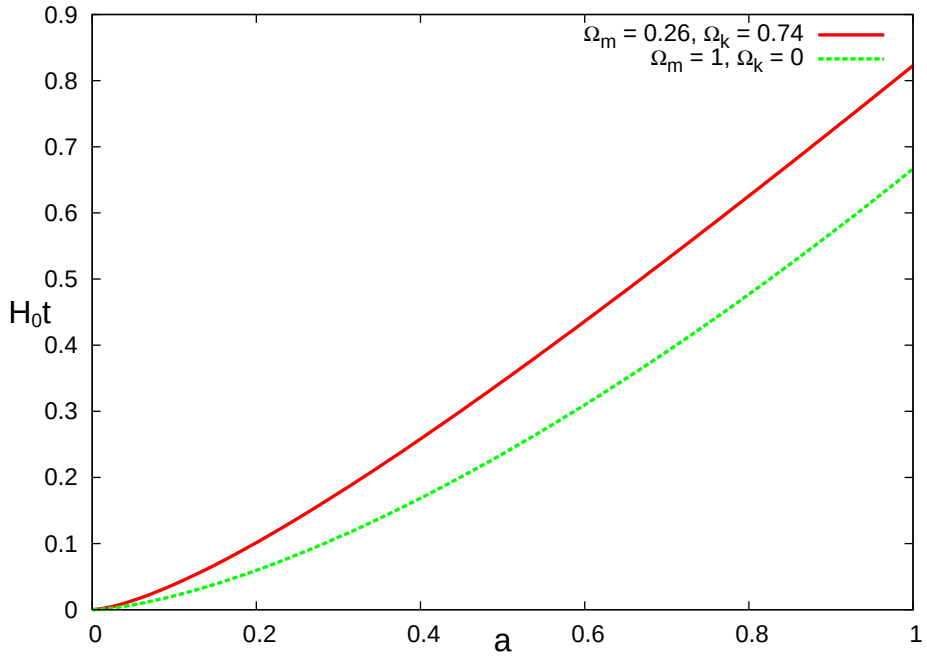


FIG. 7: The evolution of scale factor a *vs.* cosmological time H_0t in an open universe dominated by nonrelativistic matter ($\Omega_k = 1 - \Omega_{\rm m}$) with (a) $\Omega_{\rm m} = 0.26$ (upper red solid curve) and (b) $\Omega_{\rm m} = 1$ (lower green dashed curve).

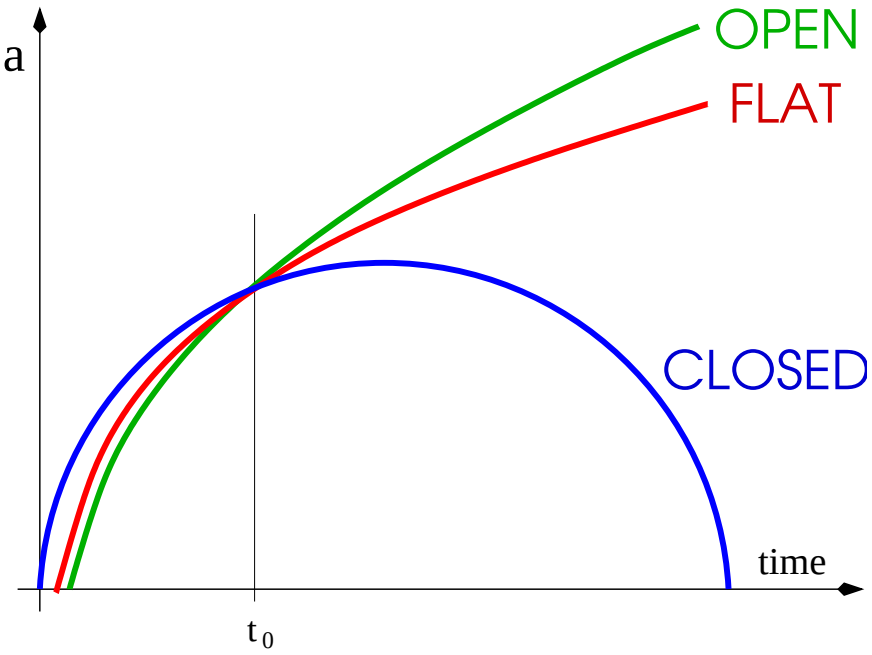


FIG. 8: Scale factor a as a function of cosmic time t in an open (green line), flat (red line) and closed (blue line) universe. Our Universe is very close to being spatially flat. t_0 denotes the present time.

To study the properties of the general equation (84) for the age of the Universe is quite delicate. In the case when $\Lambda = 0$ and $\Omega_Q = 0$, the Universe will expand forever if $\Omega_k \geq 0$ ($k \leq 0$), and it will recollapse if $\Omega_k < 0$ ($k > 0$). This is illustrated in figure 8. We expect that the universe recollapses when the denominator in Eq. (84) vanishes at some instant in time. This is however, not true in general. One can imagine a situation in which $\Omega_\Lambda > 0$ and/or $\Omega_Q > 0$ today, such that at some time in the future the denominator in (84) vanishes, and the Universe starts recollapsing. Nevertheless, before the Universe reaches zero size, it may start expanding again. A universe in which there was or will be a prolonged period during which the Universe expands or collapses at a slow pace is called a loitering phase, and used to be invoked in order to resolve the age crisis of the Universe of the 1990's. Namely, a loitering phase tends to prolong the age of the Universe, for the Universe of a given size and matter composition.

10. The age of the Earth, the Solar System and the Universe

A first estimate of the age of the Earth is due to Rutherford, and it was based on the Uranium dating. Uranium appears in Nature in two isotopes, ^{235}U and ^{238}U, and whose concentrations are related as, $^{235}U = 0.00725^{238}U$. ^{238}U decays predominately through the radium series, whose final element is an isotope of lead,

$$^{238}U \overset{radium\ series}{\longrightarrow} {}^{206}Pb\,, \tag{86}$$

with the lifetime, $t_{238} \equiv \lambda_{238}^{-1} = 6.45$ Gy. ^{235}U on the other hand, decays predominately through the actinium series,

$$^{235}U \quad \stackrel{actinium\ series}{\longrightarrow} \quad ^{207}Pb\,, \tag{87}$$

with the lifetime, $t_{235} \equiv \lambda_{235}^{-1} = 1.015$ Gy. The assumption that the abundance of the uranium isotopes, ^{235}U and ^{238}U, in the proto-Earth was equal, lead Rutherford to the following estimate of the age of the Earth (which also represents an estimate for the age of the Solar System),

$$t_{\rm Earth} \sim 3 \times 10^9 \text{ year}\,. \tag{88}$$

A more modern technique for estimation of the age of the Solar System is due to Patterson (1956), and it is based on dating of meteorites. Since ^{204}Pb has no long-lived parent isotopes, we can safely assume that its concentration was identical at the time of creation as it is today. What is then measured is the following two ratios,

$$\begin{aligned} R_{235} &= \frac{^{207}Pb}{^{204}Pb} = \frac{(^{207}Pb)_{\rm in}}{^{204}Pb} + \frac{^{235}U}{^{204}Pb}\Big(\mathrm{e}^{\lambda_{235}t_\odot} - 1\Big) \\ R_{238} &= \frac{^{206}Pb}{^{204}Pb} = \frac{(^{206}Pb)_{\rm in}}{^{204}Pb} + \frac{^{238}U}{^{204}Pb}\Big(\mathrm{e}^{\lambda_{238}t_\odot} - 1\Big)\,. \end{aligned} \tag{89}$$

Let us consider two meteorites (a and b), with different ratios, R_{235a}, R_{235b}, R_{238a} and R_{238b}. Upon taking the ratio of the subtracted ratios for the different meteorites, the dependence on the initial concentrations drops out,

$$\frac{R_{235a} - R_{235b}}{R_{238a} - R_{238b}} = \frac{(^{235}U)_a - (^{235}U)_b}{(^{238}U)_a - (^{238}U)_b}\,\frac{\mathrm{e}^{\lambda_{235}t_\odot} - 1}{\mathrm{e}^{\lambda_{238}t_\odot} - 1}\,, \tag{90}$$

which gives the following estimate for the age of the Solar System,

$$t_\odot \simeq 4.6 \times 10^9 \text{ year}\,. \tag{91}$$

To estimate the age of our galaxy (Milky Way) is complicated by the fact that heavy elements are constantly replenished by supernova explosions. From astrophysical considerations we know that there were many more supernovae in the first billion years after the creation of the galaxy, whereby afterwards the rate of supernova explosions has been relatively low and stable. Upon taking this into consideration, and depending on precisely what one assumes for the rate of supernova explosions, one arrives at different estimates of the age of the Milky Way, which fall in the range,

$$t_{\rm Milky\ Way} \sim 6 - 14 \text{ Gy}\,. \tag{92}$$

More modern methods are based on the estimate of the age of oldest stars in globular clusters. The current quoted estimate is,

$$t_{\rm globular\ clusters} = 13 \pm 2 \text{ Gy}\,, \tag{93}$$

which is to be compared with the WMAP estimate for the age of the Universe,

$$t_0 = 13.7 \pm 0.2 \text{ Gy}\,. \tag{94}$$

While this is certainly the most precise estimate of the age of the Universe, one should keep in mind that it is a derived quantity. The age (94) is obtained by making certain assumptions on the energy composition of the Universe, and then integrating the Friedmann equation (84), based on which one arrives at the estimate (94). Given this, it is still a good idea to continue using direct methods to improve age estimates.

In the 1990s an age crisis was declared, which was based on the fact that the Universe was thought to be matter dominated at present, from which one can estimate the age to be between $(2/3)H_0^{-1}$ (for the Einstein-de Sitter Universe with $\Omega_{\rm m} = 1$ at present) and $0.83H_0^{-1}$ (for an open Universe with $\Omega_{\rm m} = 0.26$ at present), as can be seen from figure 7. Moreover, the age of the globular clusters was overestimated by about 10%. The age of globular clusters was corrected by the Hyparcos satellite, which measured the paralaxes of nearby stars, and thus allowed for a more precise estimate of the distance to globular clusters, which in turn affected the estimated age. The age crisis was eventually resolved by the observation that the Universe's energy density is dominated by a dark energy, which makes the Universe of the same size older, $t_0 \simeq H_0^{-1}$.

B. Basic facts about the Universe

1. *The Mach Principle and the Cosmological Principle*

In the 1890's Ernst Mach argued that inertial reference frames of the Newton theory, and likewise an unobservable aether, cannot have absolute meaning. In 1893 he postulated the following *Mach Principle*,

Intertial frames are determined by the distribution and motion of the matter in the Universe.

Even though Einstein accepted the Mach Principle as a guiding principle in constructing his general theory of relativity, the theory is *not Machian*. This can be seen as follows. Since the Einstein equation is of second order in derivatives, for any distribution of matter and at each point of space and time, one can find a coordinate transformation into a locally Minkowsky coordinate system, such that the metric tensor reduces to the Minkowski metric, and the Levi-Cività connection vanishes,

$$g_{\mu\nu}(x) \to \eta_{\mu\nu} = \text{diag}(1,-1,-1,-1)\,, \qquad \Gamma^{\mu}_{\alpha\beta} = 0\,. \tag{95}$$

The invariance of the Minkowski line element in this frame,

$$ds_M^2 = dt^2 - d\vec{x}^2 \tag{96}$$

generalises to the invariance of the line element for a general metric tensor, $g_{\mu\nu} = g_{\mu\nu}(x)$,

$$ds^2 = g_{\mu\nu}(x) dx^\mu dx^\nu \, . \tag{97}$$

The coordinate system (96) is equivalent to local inertial coordinates of the Newton mechanics. In the absence of any matter, $g_{\mu\nu} = \eta_{\mu\nu}$ everywhere is the unique solution to the Einstein field equations. This allows us to postulate that a particle can move arbitrarily far from any distribution of matter, to regions where matter contributes unmeasurably little to $g_{\mu\nu}$, such that the solution for $g_{\mu\nu}$ is arbitrarily close to $\eta_{\mu\nu}$. Since this defines a solution to the Einstein field equation *far* from any distribution of matter, this is contrary to the Mach principle, which states that inertial reference frames can be defined only with respect to a local matter distribution.

From a modern point of view, it is natural to contend that the rest frame of the Universe corresponds to the rest frame of the cosmic microwave background (CMB) photons. All other inertial frames are defined by the velocity with respect to that frame.

In the 1930's Einstein postulated that,

The Universe is homogeneous and isotropic, when averaged over large scales.

This idea was named in 1935 by Milne the *Einstein Cosmological Principle*. At that time there was no evidence supporting homogeneity on large scales. As we will see in a moment, today there is ample evidence supporting the Cosmological Principle.

Figure 9 shows the APM Galaxy Survey. The depth of the Survey (corresponding to the distance of the faintest galaxies in the Survey) and the width of the Survey are about 600 Mpc, which corresponds to about 15% is the Hubble radius, $cH_0^{-1} \simeq 4300$ Mpc. The survey represents a two dimensional projection of galaxies (no direct information about the distance to the galaxies is given), and covers a large portion of the sky.

One clearly sees homogeneity at the scale of the survey. The lagest structures seen in figure 9 are voids and filamental structures, which can be estimated to be about 20% of the size of the survey, corresponding to largest structures in the Universe of about 100 Mpc.

In order to provide a more complete information about the three dimensional structures of the Universe, the 2dF (Two Degree Field) and SDSS (Sloan Digital Sky Survey) surveys have been conducted,

FIG. 9: The APM Galaxy Survey (Automated Plate Measuring Machine, completed in 1990) is a catalogue of several millions of galaxies.

which in addition measure the redshift $z = a_0/a - 1$ of galaxies, providing thus an (incomplete) information about the distance, where $a = a(t)$ denotes the scale factor at time t and $a_0 = a(t_0)$ today.

According to the Hubble law,

$$cz \simeq v_H \simeq H_0 r_{\rm ph} \,, \tag{98}$$

the Universe expands uniformly. Here v_H denotes the Hubble speed, H_0 the Hubble parameter today, and $r_{\rm ph}$ the physical distance. In reality, one must add to (98) the (peculiar) velocity caused by the local dynamics inside clusters of galaxies, which introduces a noisy component to the Hubble law. In general, as distances become greater, peculiar velocity becomes a smaller fraction of the Hubble velocity, v_H, such that one can obtain the actual distance from measured redshifts to a higher accuracy.

The 2dF survey (`http://www.mso.anu.edu.au/2dFGRS/`) was completed in 2002, and contains redshifts of about 220,000 galaxies (and about 22,000 quasars), selected from the Southern Sky part of the APM Galaxy Catalog, as shown in figure 10. The SDSS (`http://www.sdss.org/`), which was

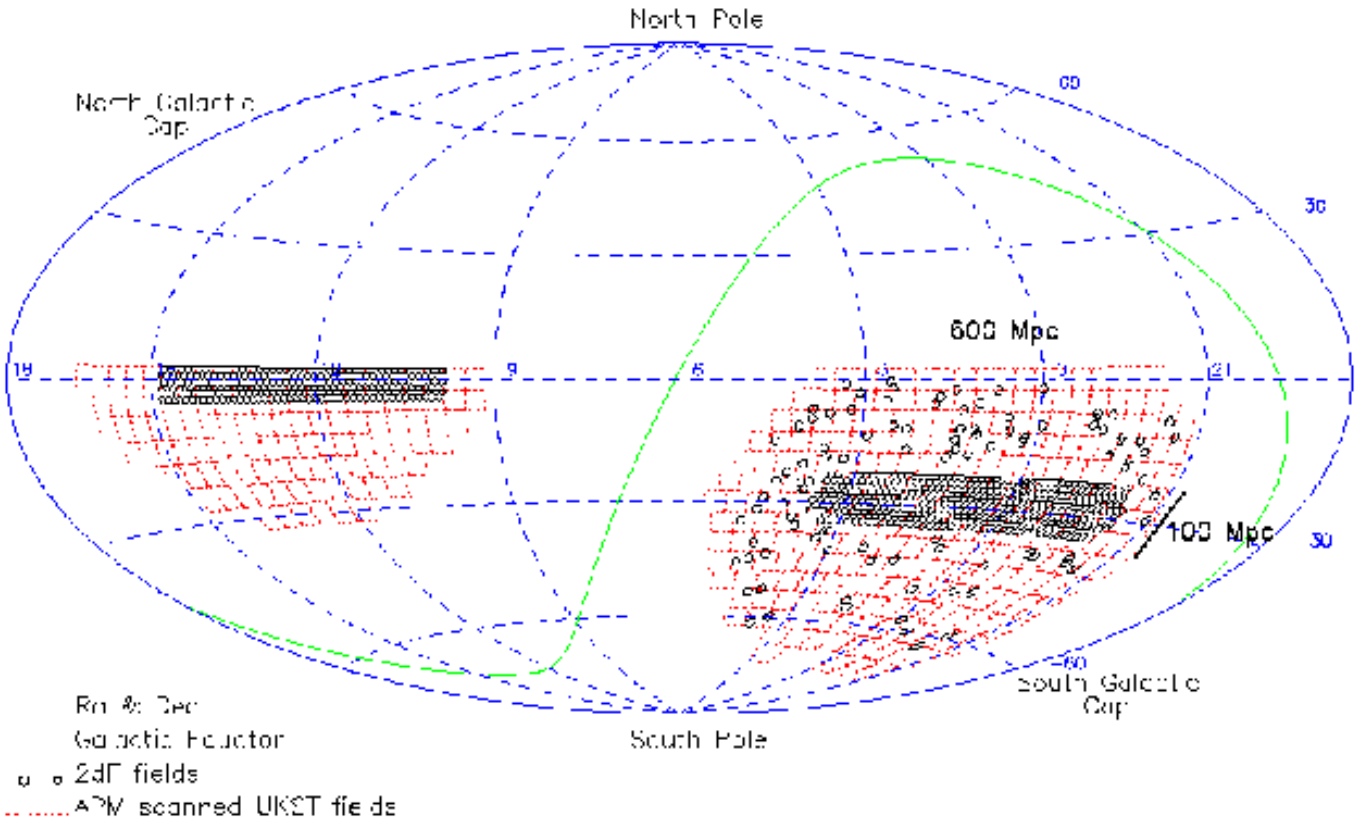


FIG. 10: The regions of the sky covered by the 2dF Survey of galaxies, one located at the equatorial plane, and another about 30 degrees south of the equator. In addition, there are small circles, which represent a part of the 'pencil' survey, which goes even deeper in redshift space.

completed in June 2006, contains a publicly available database, with information on about 150 million galaxies, among them about one million with recorded spectra and redshifts, and about 10^5 quasars with redshifts.

In figure 11 the two slices of the 2dF survey are shown. The radial distance from the center is proportional to the redshift, the most distant galaxies correspond to $z \simeq 0.25$. At large distances one can clearly identify the selection effect, according to which, at large distances only bright galaxies are selected from the APM catalogue (they appear faint due to their large distance). At a first sight, the density of galaxies appear larger at intermediate distances (red color) than in the intermediate neighbourhood. This impression is caused by the fact that the slice becomes thicker as one goes to larger distances.

A more careful analysis of the catalogue confirms the basic picture obtained by analysing the APM catalogue. The bigest structures correspond to voids, fillametal structures, and possibly walls whose size is not greater than about 100 Mpc.

FIG. 11: The two slices from figure 10 of the galaxies covered by the 2dF Survey. The size of the survey is about 1000 Mpc wide and about 1000 Mpc deep, or in redshift space up to about $z \simeq 0.25$.

A new Large Synoptic Survey Telescope is under construction in Cerro Pachón in Chile to become operational in 2012 (for more information see `http://www.lsst.org/lsst_home.shtml`). The telescope's primary mirror is 8.4m in diameter and is designed to measure 3 dimensional maps of the mass distribution of the Universe, and in particular to provide accurate information about the dark matter and dark energy of the Universe. Furthermore the telescope will provide a large data base of various astronomical data.

In order to get a more quantitative estimate of galaxy clustering, one needs to study correlations between the positions of galaxies. Assuming galaxies have a certain definite distribution of sizes, one can estimate the size of largest structures, by making use of the reduced two-point correlation function,

$\xi(r)$, defined as follows,

$$dP = n_0^2[1 + \xi(r/r_0)]dV_1 dV_2\,, \tag{99}$$

where n_0 denotes the average density of galaxies, dV_1 and dV_2 are small volume elements placed at points $\vec{r}_1$ and $\vec{r}_2$, respectively, and P is the joint probability for finding galaxies in the volume element, $dV_1 dV_2$. Because of homogeneity, ξ depends only on the magnitude of the relative distance, $r = \|\vec{r}_1 - \vec{r}_2\|$; r_0 is defined at the relative distance at which $\xi = 1$. Note that by definition, $\int \xi(r/r_0)dV_1 dV_2 = 0$. For a random stationary Poisson process, $\xi = 0$. Any deviations from $\xi = 0$ are caused by gravitational clustering. Observationally,

$$\xi|_{\rm obs} = \left(\frac{r}{r_0}\right)^{-\gamma}, \qquad \gamma = 1.77 \pm 0.04\,, \qquad r_0 = 7.4 \pm 1.4 \text{ Mpc}\,. \tag{100}$$

While this behaviour is reproduced by N-body numerical simulations of structure formation from initial adiabatic scale-invariant density perturbations (adiabatic scale-invariant perturbations are defined roughly as equal relative matter density perturbations on all scales, which are gaussian distributed), there is currently no analytical understanding of this empirical law. Based on the Abel survey of clusters, one infers that an analogous law is observed for clusters of galaxies, $\xi_{\rm cl} \sim (r/r_{\rm cl})^{-\gamma_{\rm cl}}$, with $\gamma_{\rm cl} \simeq 2$, suprisingly close to the correlation function of galaxies, and $r_{\rm cl} \simeq 20 \pm 5$ Mpc. Since clusters are not yet fully virialised, the correlation function need not have the same slope.

The form of the correlation function for galaxies and clusters clearly indicates that on large scales the Universe becomes more and more homogeneous. The question is then whether this law is obeyed on even larger scales, and in precisely what way is homogeneity reached.

This question can be answered by studying the reduced two-point correlation function as a function of scale. In figure 12 we show the observed reduced two-point correlation function, ξ, based on about 200,000 2dF galaxies, 26,000 SDSS galaxies, and the APM survey as a function of scale. The figure shows that no simple power law can be fitted to correlations on all scales covered by the surveys, hence Eq. (100) holds only approximately. Note that the two point correlation function becomes smaller than the correlation function on scales larger than 10 Mpc, and it is much smaller than the one on scales of about 100 Mpc. To see how homogeneity is approached on even larger scales, in figure 13 we show a compilation of data, which shows the amplitude of relative energy density fluctuations, as a function of scale, from a sub-megaparsec scale, which is estimated by the intergalactic hydrogen clumping, up to the Hubble scale, which is measured by the cosmic microwave background radiation. Note that the amplitude of density fluctuations, measured by cosmic microwave background radiation, is of the order 10^{-4} on the largest (Hubble) scales, implying that the Universe is homogeneous to one part in 10,000 on the largest observable scales.

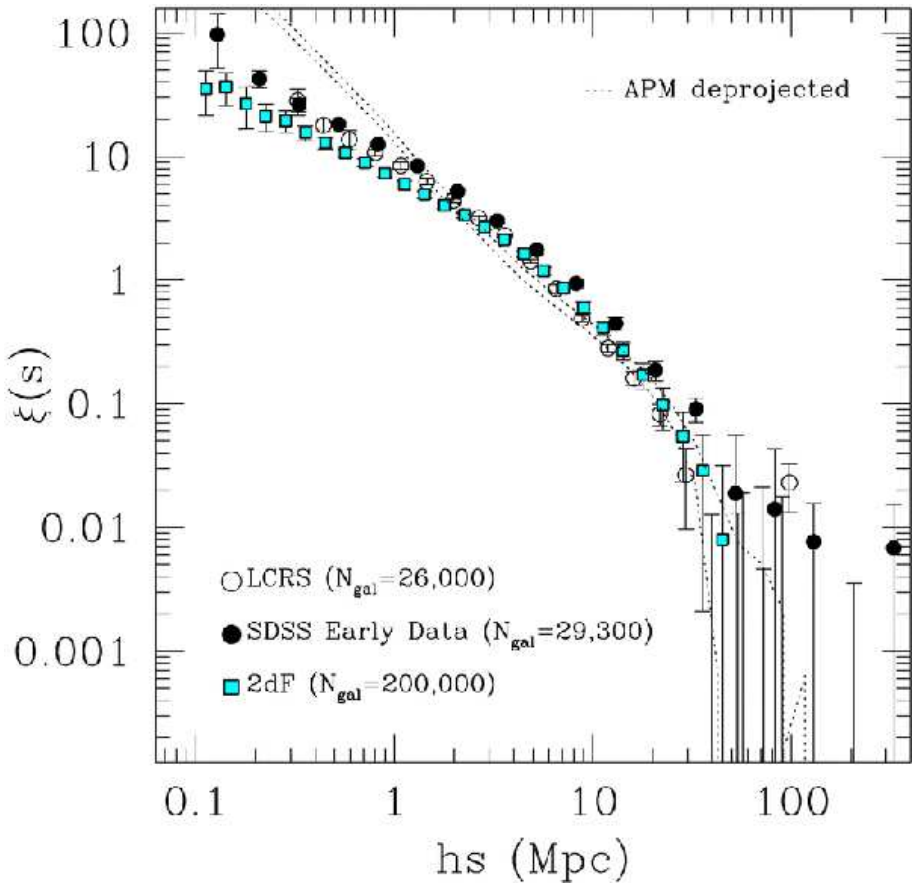


FIG. 12: The reduced two-point correlation function ξ as a function of scale (in the units of $h^{-1}Mpc$, with $h = 0.73 \pm 0.03$). Note that the scales are logarithmic, such that the correlations are displayed from about 0.1 Mpc up to more than 200 Mpc.

Further evidence for homogeneity comes from the foregrounds observed by the MWAP satellite (`http://map.gsfc.nasa.gov/`, see also `http://map.gsfc.nasa.gov/m_or.html`). In figure 14 we show the 3-color foreground signal maps (red, green, blue) of the Ka-Band Map (33 GHz). Synchrotron radiation is red, free-free is green, and thermal dust is blue. Figure 15 shows the W-Band Map (94 GHz) of the same foreground signals. Note that the signal from the Milky Way galaxy shows prominently as a bright horizontal streak across the middle of each map. The foreground signals (primarily our galaxy) are weakest in the W-Band map (94 GHz). Apart from the Milky Way, the maps clearly indicate homogeneity of the dust distribution on large scales. Furthermore, in figure 16 we show the microwave foreground point sources observed by the WMAP satellite, which are also distributed homogeneously over the sky.

As a final evidence of homogeneity, we look at the cosmic microwave background radiation (CMBR). Since the photons of the CMB radiation were emitted at the time of last scattering at $z \simeq 1089$, the

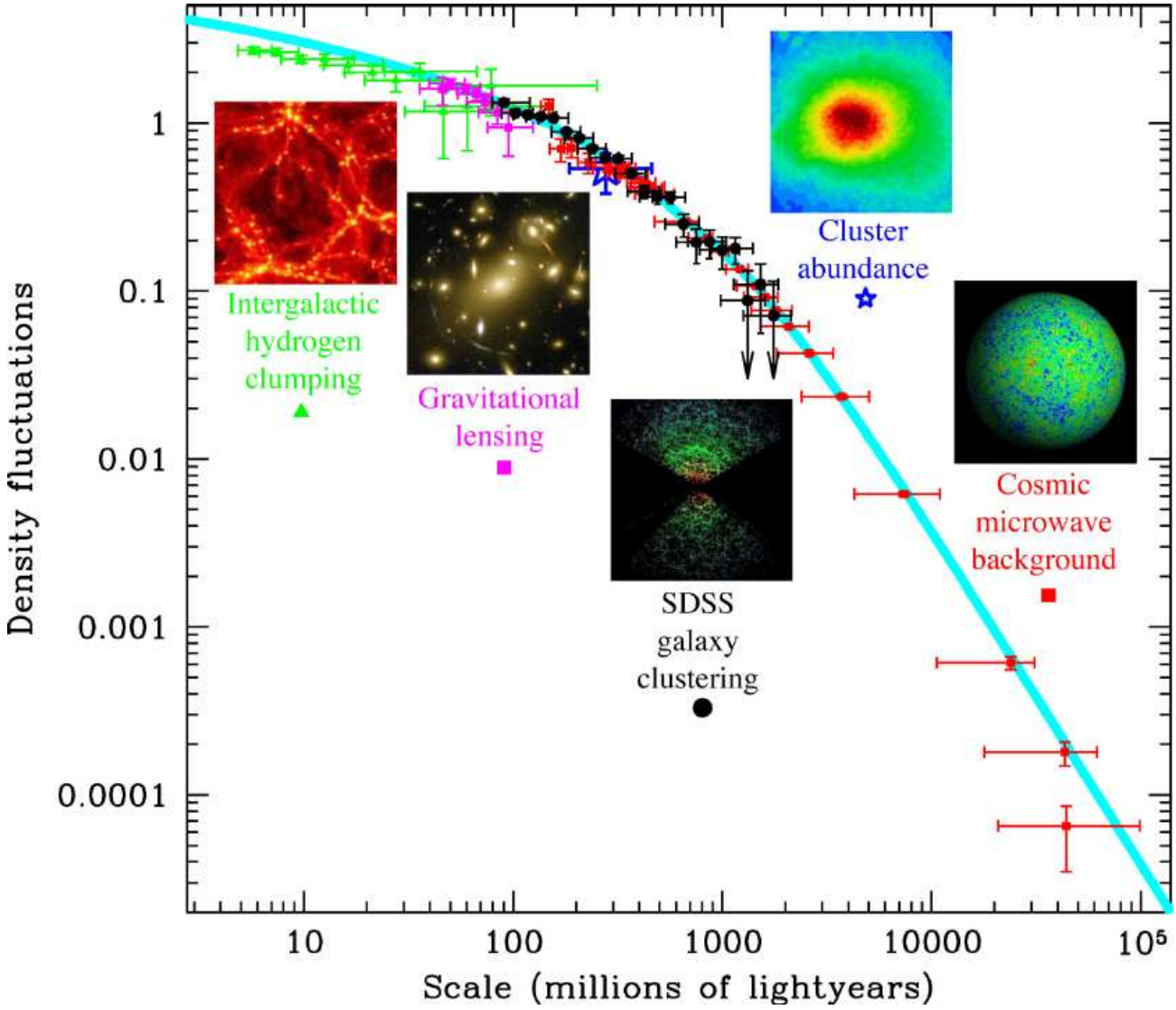


FIG. 13: The energy density fluctuations as a function of scale (represented in the units of light-years). The spectrum streches over almost four decades of scales.

temperature at which electrons and protons recombined into the neutral hydrogen, by observing the CMB we mostly see how the Universe looked at the time of last scattering, albeit redshifted.

The expansion of the Universe redshifts the temperature of the photon fluid, but does not change the nature of the photon distribution function. For example, let us consider a photon fluid with a thermal distribution function (defined as the photon occupancy per unit phase space volume), $n_\gamma = 1/(\mathrm{e}^{E_\gamma/(k_B T_\gamma)} - 1)$, where $E_\gamma = pc = \hbar\nu$ denotes photons energy, and $T_\gamma = T_\gamma(x)$ a local temperature and $k_B = 8.617342 \pm 0.000015 \times 10^{-5}$ eV/K is the Boltzmann constant. Since the photon energy redshifts with the expansion as, $E_\gamma \to E_\gamma/a$, and in absence of interactions (free streaming) the photon number is an invariant, the photon temperature redshifts according to,

$$T_\gamma \to \frac{T_\gamma}{a}\,. \tag{101}$$

FIG. 14: The infrared foregrounds observed by the WMAP satellite. The map represents the K-Band (23 GHz) foregrounds. Synchrotron radiation is red, free-free is green, and thermal dust is blue.

FIG. 15: The infrared foregrounds observed by the WMAP satellite. The map represents the W-Band (94 GHz) foregrounds. Synchrotron radiation is red, free-free is green, and thermal dust is blue.

Spatial variations in the photon temperature are a measure of homogeneity of the Universe at recombination. In figure 17 we show the spectrum measured by the WMAP of the CMB photons, onto which a perfect black body spectrum is superimposed (red curve). The spectrum is so perfect black body that in no place deviations are bigger than the thickness of the red curve. On average, the error bars are

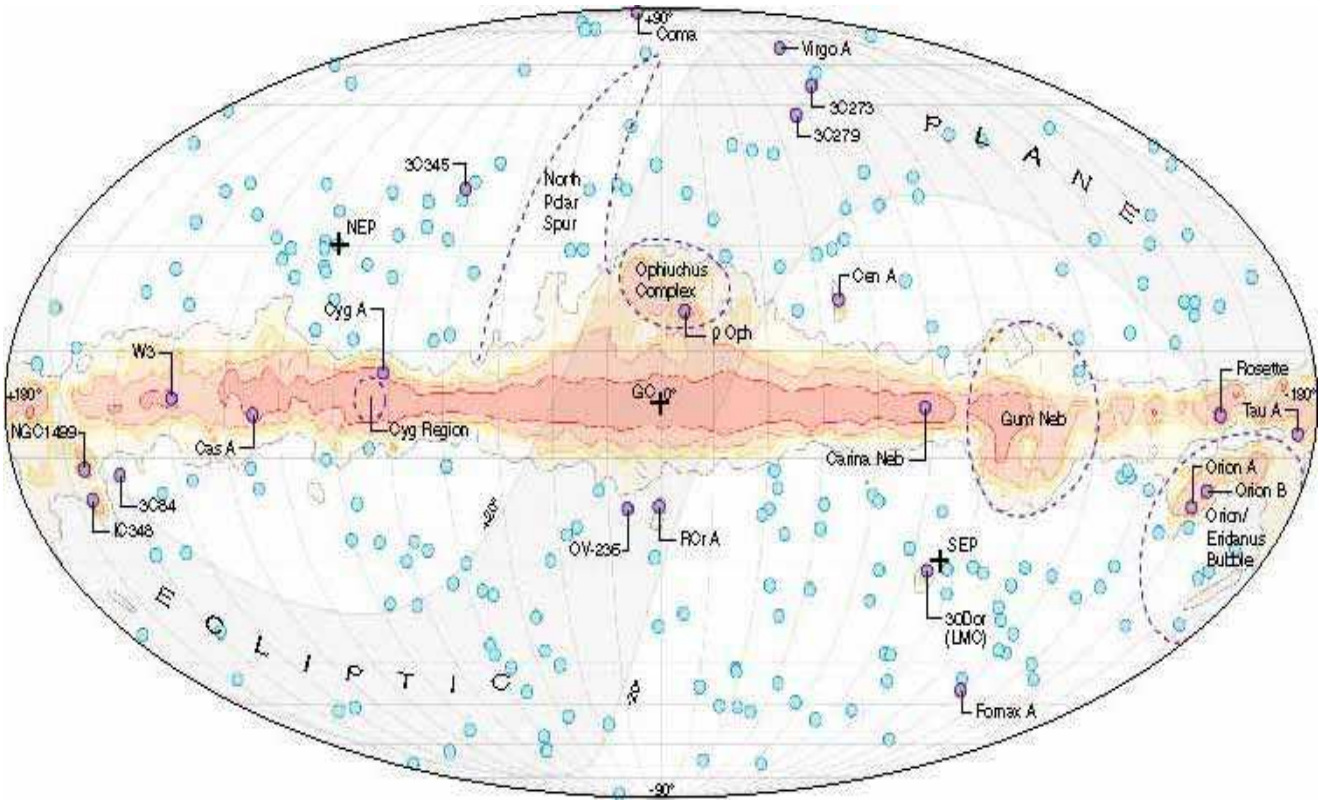


FIG. 16: The microwave foreground point sources observed by the WMAP satellite.

about 20% of the curve thickness. This is a very strong evidences that the early Universe was in thermal equilibrium, and one of the stronger supporting evidences for the Standard Big-Bang model. In figure 18 we show the full sky map of the CMBR, with the monopole $T_0 = 2.725 \pm 0.001$ K being subtracted. The red color corresponds to the hottest region of the sky, $T = 2.729$ K while the violet spot is the coldest region of the sky, $T = 2.721$ K. The plot is completely dominated by the dipole, whose amplitude is, $T_1 = 3.358 \pm 0.0017$ mK in the direction on the sky, $(\ell, b) = (263^o.86 \pm 0^o.04, 48^o.24 \pm 0^o.10)$, and it does not represent an intrinsic anisotropy, but it is a consequence of our motion with respect to the rest frame of the CMB which, as we already mentioned, can be considered as the rest frame of the Universe. The implied motion of the Solar System with respect to the CMB reference frame is, $v = 368 \pm 2$ km/s, and of the Local Group has the amplitude, $v_{\rm LG} = 627 \pm 22$ km/s, directed towards the galactic longitude, $l = 273^\circ \pm 3^\circ$ and lattitude, $b = 27^\circ \pm 3^\circ$.

The intrinsic anisotropies in the CMB occur at the relative level of 10^{-5} (tens of micro-kelvins), and were first observed by the COBE satellite in 1992 (for which John Mather and George Smoot were awared Nobel Prize for Physics in 2006), and more recently by several baloon experiments (*e.g.* BOOMERanG and MAXIMA), and by the WMAP satellite. The root-mean-square value of the WMAP

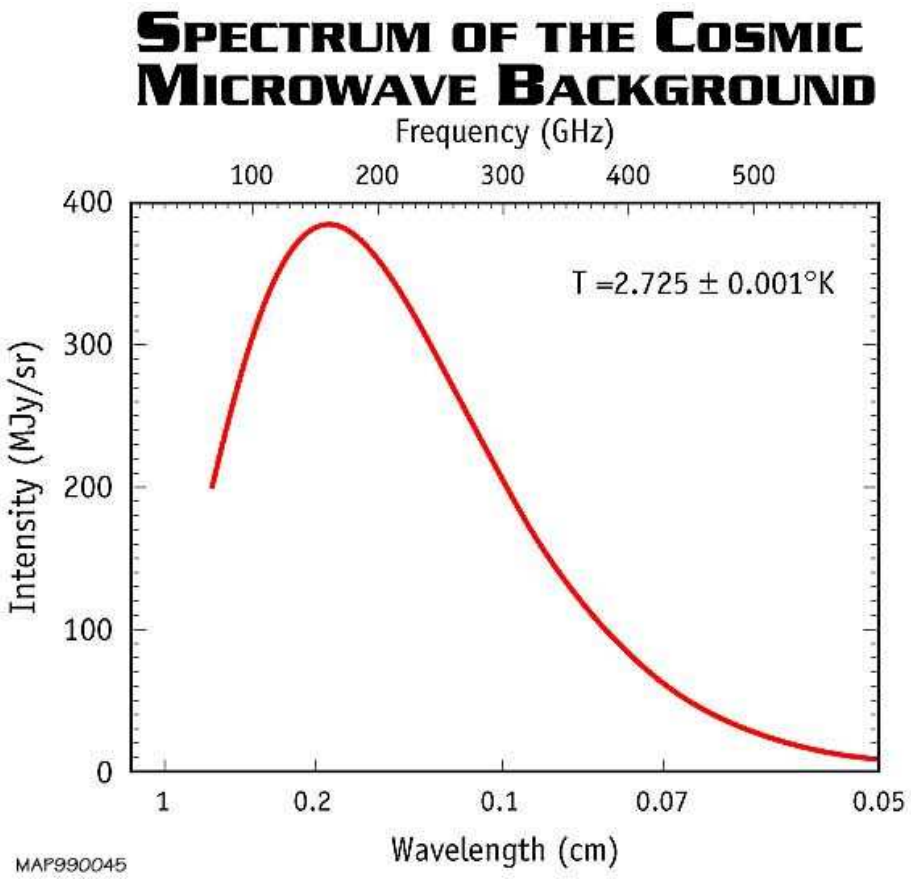


FIG. 17: Black Body Spectrum as seen by the WMAP satellite. The error bars are about 20% of the line thickness.

quadrupole temperature anisotropy is by about a factor 2 lower than expected, $Q_{\rm rms} = 8 \pm 2$ μK, or $\Delta T_2 = 12.4 \pm 2.8$ μK. The WMAP team measured the anisotropies in five frequency bands, K-Band Map (23 GHz), Ka-Band Map (33 GHz), Q-Band Map (41 GHz) V-Band Map (61 GHz), and W-Band Map (94 GHz). The resulting CMB maps of the whole sky are shown in figures 19, 20 and 21. Note that the vertical red stripes in the figures are not primordial, but represent mostly the foreground microwave sources in the Milky Way, see figures 14–16. Since the galactic microwave foregrounds have different spectral characteristics from the back body of the primordial photons, to a large extent they can be subtracted by combining the maps at different frequencies. One such map is shown as the last image in figure 21, in which, apart from a few bright spots, the galactic foregrounds have been almost completely subtracted, and the remaining fluctuations are considered to be primordial.

In conclusion, we have argued that the today's evidence for homogeneity of the Universe on large scales is very convincing, hence the Einstein Cosmological Principle can be considered proven. Moreover, the ultimate piece of evidence, the cosmic microwave background radiation, strongly suggests that the

FIG. 18: The dipole in CMB as measured by the COBE satellite. The temperature range is T=2.721K (violet) to 2.729K (red). The inferred dipole velocity of the Solar System is $v = 368 \pm 2$ km/s and of the Local Group, $v_{\rm LG} = 627 \pm 22$ km/s.

young Universe was much smaller and in thermal equilibrium. As we will see in the course of these lectures, a self-consistent picture of an expanding universe has emerged, which is often dubbed the Big Bang Paradigm.

2. The Universe is expanding

The main evidence supporting the fact that the Universe is expanding is *the Hubble law*, which is the observational fact that more distant galaxies are receeding away from us with a speed which is proportional to their distance. The recession speed is determined by the Doppler redshift of spectral lines. The original Hubble diagram (1929), which was used as the evidence for a linear Hubble law, is shown in figure 22.

It is not a well known fact that an evidence for expansion is also contained in galaxy catalogues, as shown for example in figure 9. The evidence is made apparent as follows. Let us define apparent magnitude of a light source in the sky,

$$m = -2.5 \log(\mathcal{F}) + \text{const.}\,, \tag{102}$$

where $\mathcal{F} = dE/(dtdAd\nu)$ denotes the measured energy flux on the Earth, in the units of energy (dE) per unit time (dt) per unit area (dA) per unit frequency interval ($d\nu$), such that the change in apparent

FIG. 19: The intrinsic anisotropies in the CMB photons as obserbed by the WMAP satellite, first year data. The WMAP team measured the anisotropies in five frequency bands. The figure shows the K-Band (23 GHz) and Ka-Band (33 GHz) maps.

magnitude, $\Delta m = m_1 - m_2 = 5$, corresponds to the ratio of the measured fluxes, $\mathcal{F}_1/\mathcal{F}_2 = 100$. Next we define the absolute magnitude as,

$$M = -2.5 \log(\mathcal{L}) + \text{const.}\,, \tag{103}$$

where $\mathcal{L}$ denotes the absolute luminosity of the source. The constant in Eqs. (102–103) is fixed such that at the distance 10 pc, $M = m$, which then implies,

$$m - M = 5 \log\left(\frac{r}{\text{Mpc}}\right) + 25\,. \tag{104}$$

FIG. 20: The figure shows the Q-Band (41 GHz) and the V-Band (61 GHz) map.

Recall that the luminosity distance d_L is defined as,

$$\mathcal{F} = \frac{\mathcal{L}}{4\pi d_L^2}\,, \tag{105}$$

such that in a static spatially flat universe, the luminosity distance corresponds to the coordinate distance, $d_L = r$, which is in this case also equal to the physical distance. In an expanding and curved universe, the relation between d_L and the matter content of the Universe is more complex, and it is the subject of Problem 2.5. Assume for simplicity that all galaxies have identical luminosity. Then the number of galaxies with an observed flux, which is greater than some $\mathcal{F}$, equals to,

$$N(>\mathcal{F}) = n_0 V = n_0 \frac{4\pi r^3}{3} = \frac{4\pi n_0}{3}\Big(\frac{\mathcal{L}}{4\pi\mathcal{F}}\Big)^{\frac{3}{2}}\,, \tag{106}$$

where r denotes the radius of a sphere of volume V, and n_0 an average density of galaxies. If there is a spread in the intrinsic luminosity of galaxies, which is characterised by a probability distribution, we

FIG. 21: The figure shows the W-Band (94 GHz) and a combined map, which is a linear combination of the 5 maps (K, KA, Q, V and W band maps), designed such that the foreground effect of the galaxy is minimised.

can bin them, $\{n_i, \mathcal{F}_i\}$, where n_i denotes average density of galaxies in bin i, and Eq. (106) generalises to,

$$N(>\mathcal{F}) = \sum_i \frac{4\pi n_i}{3}\Big(\frac{\mathcal{L}_i}{4\pi\mathcal{F}}\Big)^{\frac{3}{2}}\,, \tag{107}$$

such that we expect that the law, $N(>\mathcal{F}) \propto \mathcal{F}^{-3/2}$ holds universally. From Eq. (102) it follows, $\mathcal{F} \propto 10^{-0.4m}$, which then implies,

$$N(>\mathcal{F}) = N(<m) \propto 10^{0.6m}\,. \tag{108}$$

Taking a derivative of the differential form of this expression,

$$N(<m) = \int_0^m \frac{dN}{dm'}dm' \tag{109}$$

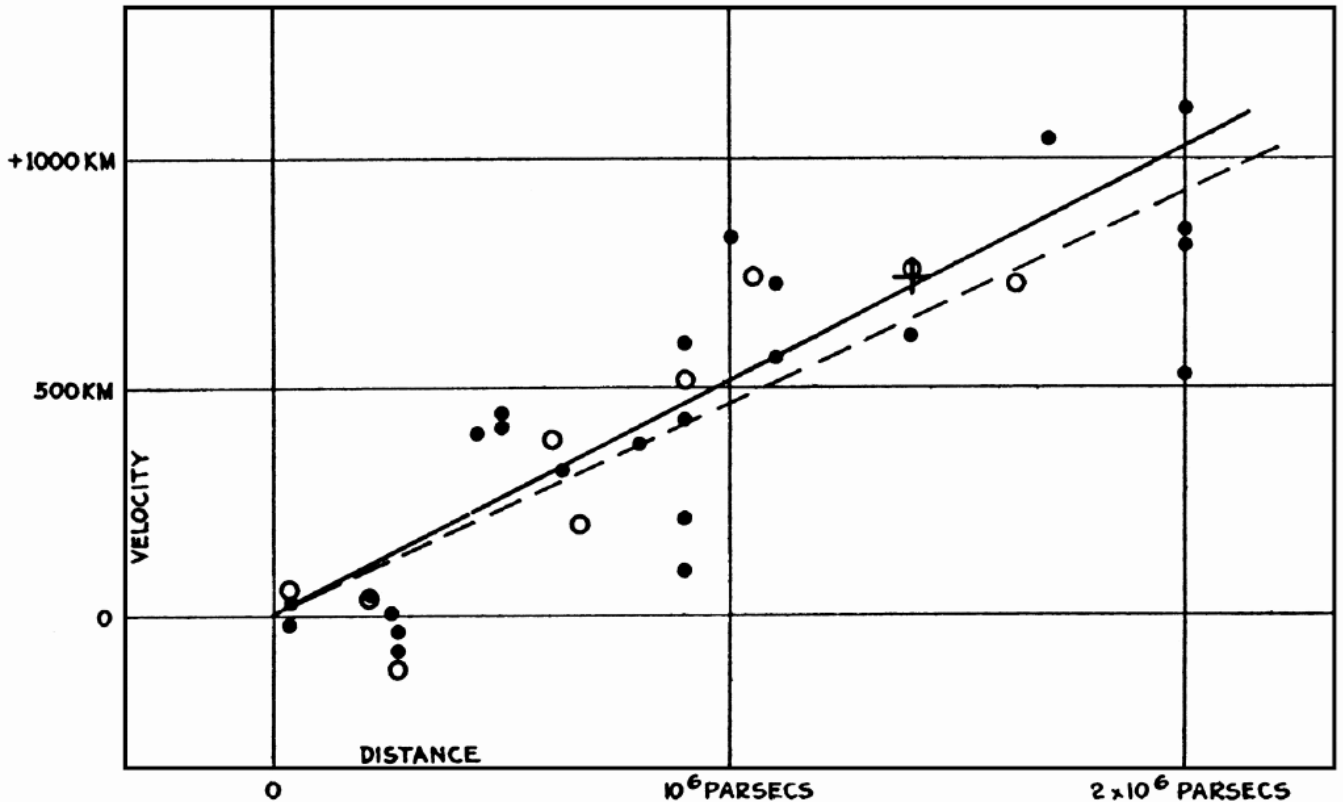


FIG. 22: The original Hubble diagram, showing that the Universe is uniformly expanding. Note that the vertical axis represents the Hubble velocity in [km/s], and not in [km], as it stands.

we easily infer,

$$\frac{dN}{dm} \propto 0.6 \times \ln(10) \times 10^{0.6m} \propto 10^{0.6m} \,. \tag{110}$$

In the 1920s is was known that dN/dm varies slower with apparent magnitude m than expected from (110). For bright galaxies, for which $m \in (15, 20)$, the law (110) is well satisfied, while for fainter (more distant) galaxies, it was observed that

$$\left(\frac{dN}{dm}\right)_{\rm obs} \propto 10^{0.45m} \,. \tag{111}$$

The disagreement is explained by the combined galactic evolution and the Universe expansion.

If this was the only evidence for Universe's expansion, it would not have been taken seriously, since the evidence is marred with evolution effects. In order to interpret properly the evidence from galactic redshift data shown in figure 22, a reliable information about the distance to galaxies is needed. An ingenious estimate to the Andromeda distance was proposed by Öpik. If spiral nebulae were galaxies like the Milky Way it is reasonable to assume that the mass-to-luminosity ratio of Andromeda is comparable

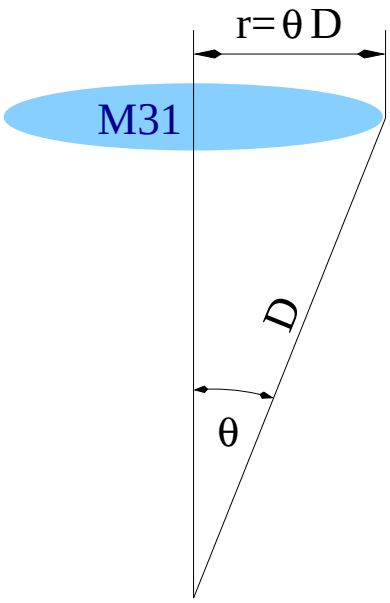


FIG. 23: The geometry used by Öpik to estimate the distance to the Andromeda galaxy, also known as M31.

to that of the Milky Way, which is about 3 times larger than the ratio for the Sun,

$$\frac{M_{\rm MW}}{\mathcal{L}_{\rm MW}} \simeq 3\,\frac{M_\odot}{\mathcal{L}_\odot} \tag{112}$$

The Milky Way estimate is made based on the mass estimate from (virialised) star velocities, and from star counts, and statistical estimate of star distances. Next, assuming that Andromeda is a virialised system of stars, the gravitational field (the force per unit mass) is on average balanced by the centrifugal force,

$$-g = \frac{G_N M}{(\theta D)^2} \sim \frac{v_c^2}{\theta D}\,, \tag{113}$$

where v_c is the average star speed at the edge of the disk. From this and $\mathcal{F} = \mathcal{L}/(4\pi D^2)$, it then immediately follows that the distance to Andromeda is,

$$D = \frac{v_c^2\theta}{4\pi G_N \mathcal{F}}\,\frac{\mathcal{F}}{M}\,. \tag{114}$$

Since v_c, θ and $\mathcal{F}$ are all measured quantities, and $\mathcal{F}/M$ can be estimated based on (112), D can be estimated,

$$D_{\rm Öpik} \simeq 450\ {\rm kpc}\,, \tag{115}$$

which is to be compared with the modern value, based on the *cepheid* variable star distance measurements,

$$D_{\rm M31} = 770 \pm 30\ {\rm kpc}\,. \tag{116}$$

Cepheid stars are currently the most widely used *standard candles* for the distance estimation on megaparsec scales. In passing, we mention that Hubble's estimate of the Andromeda distance, which was based on cepheid variable stars, was in fact poorer, $D_{\rm Hubble} \simeq 300$ kpc.

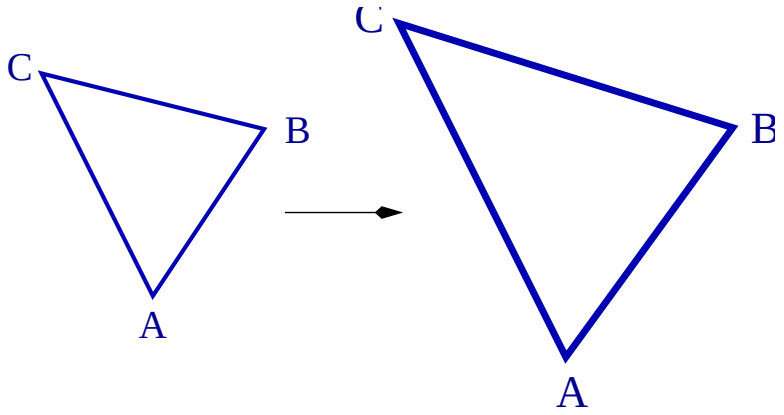


FIG. 24: An illustration of the Hubble Law. The same triangle at an earlier and a later instant in time. The points A, B and C can be thought of as being attached to some inertial frame of a local distribution of matter, *e.g.* the center of a cluster of galaxies. Due to the uniform expansion of space, at a later time, the distance between the points becomes larger.

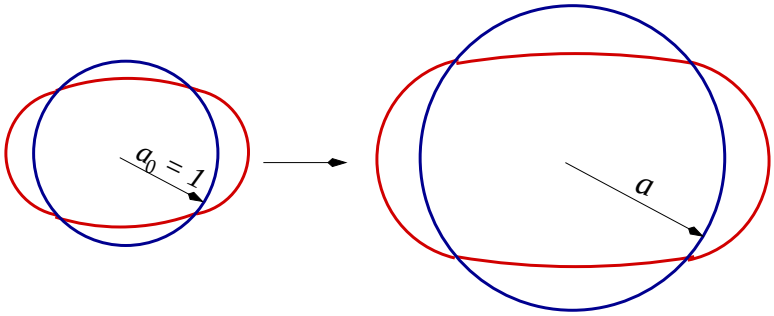


FIG. 25: An illustration for how the Hubble Law affects photons. In an imaginary situation, the photon traverses around a circle of radius $a_0 = 1$. After the Universe has expanded, the cirlce has expanded to $a > a_0$, but the number of nods of the photon must be conserved, implying that the photon wavelength must increase linearly with the size of the Universe.

The Öpik's method predates a widely used method for distance estimation based on the Tully-Fisher relation, which establishes a relationship between the velocity dispersion and the absolute luminosity of elliptical galaxies.

The Hubble Law states that the expansion of the Universe is uniform on large scales, implying that objects at a physical distance $r_{\rm ph}$ move away from us with an averaged speed,

$$v_H = H_0 r_{\rm phys} \,, \tag{117}$$

where H_0 is the Hubble constant. Figure 24 shows a triangle at an earlier and a later time. The points A, B and C can be thought to be attached to some definite galaxies. The Universe's expansion is uniform, in the sense that the distance between each pair of points increases uniformly with time.

An important question is how the expansion influences quantum fields. A simple example is the electromagnetic field, which is quantum mechanically a fluctuating field of massless photons. In order to see that the photon wavelength stretches with the expansion, in figure 25 we show an imagined photon, which wraps around a circle. Due to the Universe expansion the circle becomes bigger, and as a consequence of the conservation of the number of crests, the wavelength of the photon must increase. Hence we have,

$$\lambda_0 \to \lambda = a\lambda_0 \tag{118}$$

where $a_0 = 1$ and $a > 1$ represent the scale factors at the earlier and later time, respectively. Since $p = hc/\lambda$ and $E = pc$, we conclude that

$$p_0 \to p = \frac{p}{a}\,, \qquad E_0 \to E = \frac{E_0}{a}\,, \tag{119}$$

which is nothing but the gravitational redshift of light due to the Universe expansion. Another simple consequence of Universe's expansion is the temperature redshift, which can be argued as follows. Take a photon fluid, in which photons are distributed according to the Bose-Einstein distribution function (the Planck's black body spectrum), such that the photon occupancy of phase space is given by,

$$n_\gamma = \frac{1}{\mathrm{e}^{E/k_B T} - 1}\,, \tag{120}$$

where $k_B = 8.617342 \pm 0.000015 \times 10^{-5}$ eV/K denotes the Boltzmann constant. Since the number of photons per unit phase-space cell must be conserved, the energy redshift, $E_0 \to E = E_0/a$, implies the following temperature redshift,

$$T_0 \to T = \frac{T_0}{a}\,, \tag{121}$$

This is the redshift law that is observed by the microwave background photons, and which allows us to infer that the Early Universe was much hotter than it is today.

The measurement of the rate at which the Universe expands, which is encoded in the Hubble parameter H, has a long history, and only quite recently the measurements have become somewhat more precise. Until the 1990s, the uncertainly in the Hubble parameter was about a factor two, and for that reason, the Hubble parameter today was standardly represented as,

$$H_0 = 100h \text{ km/s/Mpc} \tag{122}$$

where h is a dimensionless number between 0.5 and 1. More recently (in 2000), based on relatively distant cepheid star measurements by the Hubble Space Telescope (HST), the HST Key Project came up with a more precise determination of the Hubble parameter,

$$H_0 = (72 \pm 3_{\text{stat}} \pm 7_{\text{syst}}) \text{ km/s/Mpc}\,, \tag{123}$$

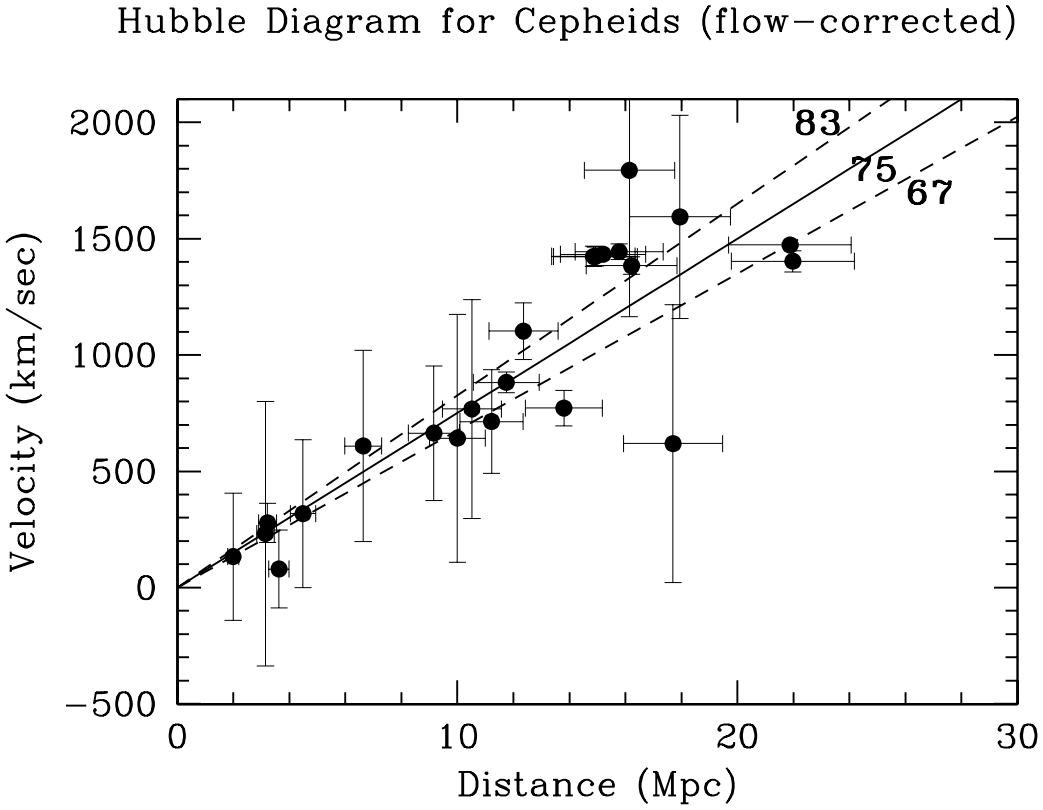


FIG. 26: The receding velocity *vs* distance of the galaxies measured by the Hubble Space Telescope Key Project (Freedman et al, 2000), based on which the Hubble parameter (123) was determined.

whereby the first error bars are statistical and the second are systematic. Systematic errors are dominated by the error in determination of the distance to standard candles, which are in this case cepheid stars. The HST Key Project data are shown in figure 26. Since the maximum distance at which a cepheid could be found by the Hubble Space Telescope is at about 20 Mpc, a significant scatter in H_0 can still be observed. In the figure, $H_0 = 75$ km/s/Mpc is shown, flanked by the $\pm 10\%$ lines, which indicate the level of the data scatter.

In figure 27 we show the Hubble diagram for a compilation of data, which include Type Ia supernovae (squares), Tully-Fisher clusters (solid circles), fundamental plane clusters (triangles), surface brightness fluctuation galaxies (diamonds), Type II supernovae (open squares). All of the second distance indicators are calibrated by cepheid variables, found by the Hubble Space Telescope, hence the consistency of the HST Key Project result in figure 26 with the data in figure 27 is not incidental, since the two data sets are not independent.

In figure 28 we show the uncertainly of various methods used to detemine H_0 in figure 27, and the gaussian envelope for all measurements taken together, which gives a good estimate of the total uncertainly in the measurement of the Hubble parameter.

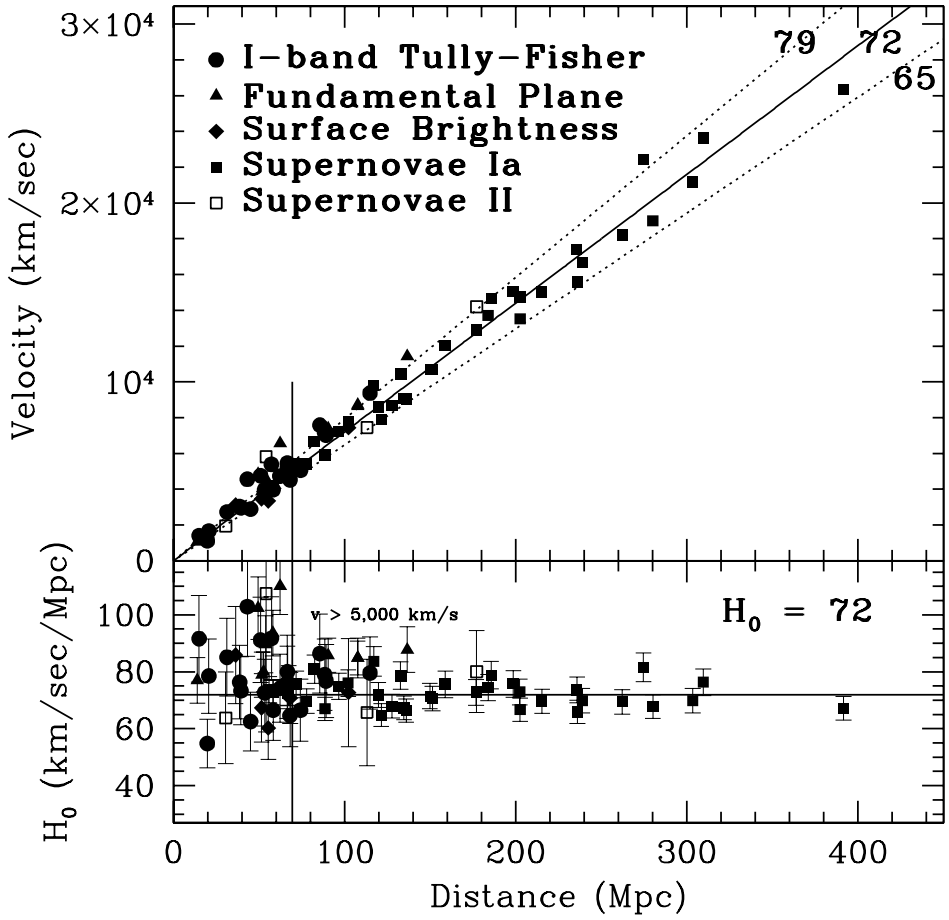


FIG. 27: The Hubble diagram of velocity *vs.* distance for second distance indicators, all calibrated by cepheid variables. The symbols are, Type Ia supernovae (squares), Tully-Fisher clusters (solid circles), fundamental plane clusters (triangles), surface brightness fluctuation galaxies (diamonds), Type II supernovae (open squares). A slope $H_0 = 72$ km/s/Mpc is shown, flanked by $\pm 10\%$ lines. Bottom panel shows the value of H_0 as a function of distance. As opposed to figure 26, the largest distance on this diagram is about 400 Mpc.

Remarkably, recent analyses of the WMAP measurements of temperature fluctuations of the cosmic microwave background (CMB) resulted in the cosmological parameters which are fully consistent with the HST Key Project result. The WMAP result for the Hubble parameter today is,

$$H_0 = 71 \pm 4 \text{ km/s/Mpc} \qquad (WMAP)\,. \tag{124}$$

A useful quantity, which gives a good idea about the age of the Universe, is the inverse Hubble parameter,

$$H_0^{-1} = 9.78 h^{-1} \times 10^9 \text{ year} = 13.7 \pm 0.5 \text{ Gy}\,, \tag{125}$$

which is to be compared with the WMAP estimate of the age of the Universe, $t_0 \simeq 13.7 \pm 0.2$ Gy. In a matter dominated universe, $H_0 = 2/(3t_0)$, such that the age is much lower, $t_0 = (2/3)H_0^{-1}$. The

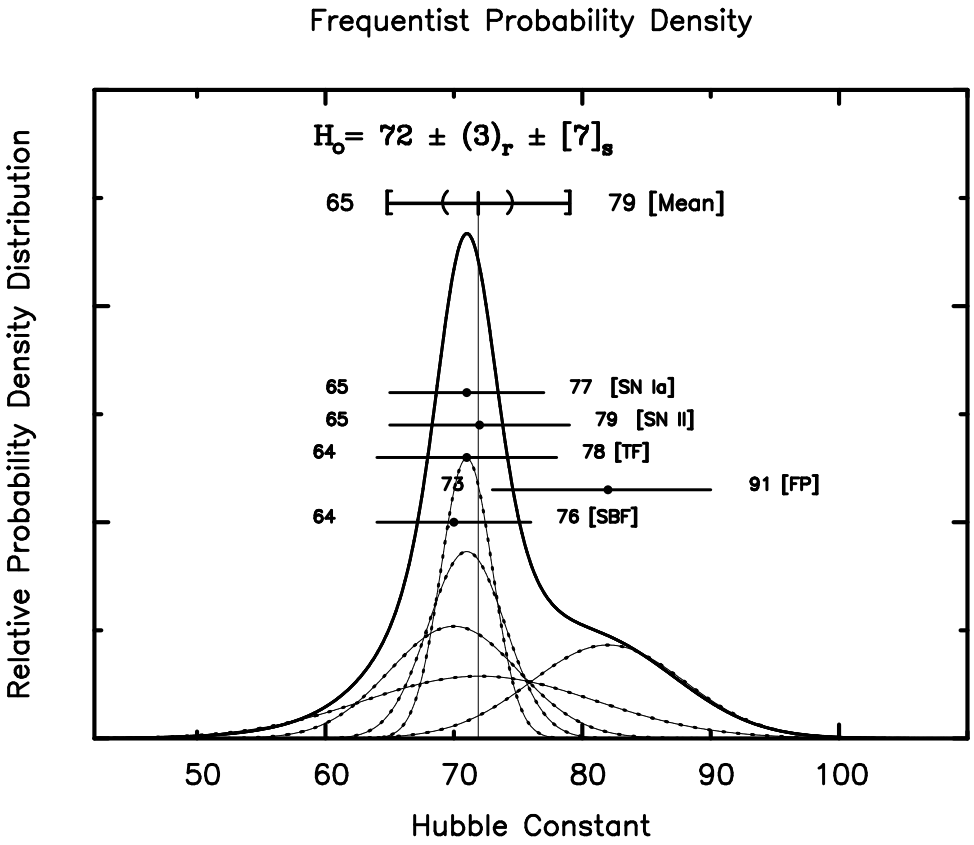


FIG. 28: The dispersion of the Hubble parameter H_0 in the HST Key Project. The various curves represent the uncertainties in H_0 for Type 1a supernovae, the Tully-Fisher relation, Type II supernovae, etc. All are calibrated by cepheid variables. The upper curve is obtained by summing the individual Gaussians.

explanation is to be found in a dark energy component of the Universe, which is believed to drive an accelerating expansion of the Universe today, and prolongs its age.

3. The Universe is accelerating

The discovery that the Universe is accelerating came as a surprise, and one still finds a few skeptics, who claim that it is sufficient to assume that distant supernovae Ia are about 40% dimmer (evolution effect), and the evidence for the Universe acceleration goes away. This contention is not unreasonable, since the physics of supernovae explosions is not well understood. Nevertheless, when taken together with all other measurements, which include the CMB measurements, Hubble parameter measurements, and the dynamical mass measurements, the evidence for accelerating Universe is quite convincing.

It is hence worth putting some effort into understanding the basis for the use of supernovae Ia as standard candles. In figure 29 we show the spread in the absolute luminosity of nearby supernovae Ia, before and after the spread is corrected for. Before the correction, the scatter is quite large, as can be

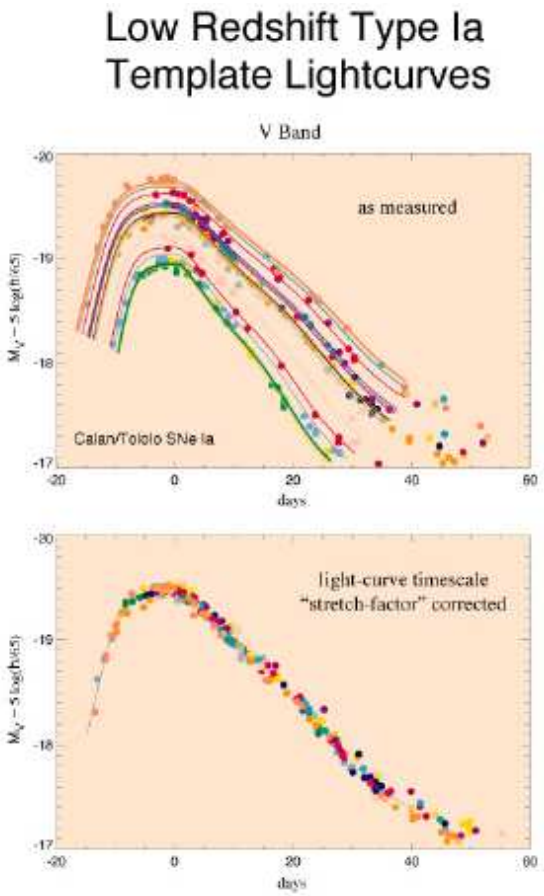


FIG. 29: The uncorrected and corrected light curves of nearby supernovae Ia. The correction is based on a phenomenological relation between the luminosity at the maximum and the total duration of the explosion, captured by the timescale of the light curve. A simple linear relationship between the absolute magnitude and and the timescale of the ligtcurve holds. The so corrected luminosities fall on top of each other (the remaining scatter is quite small, about $\Delta m \sim 0.2 - 0.3$), suggesting that the corrected absolute luminosities of SNIae can be used as standard candles.

seen in the top panel. There seems to be however a simple relation between the maximum luminosity and the timescale of the light curve. When a simple linear relation between the absolute magnitude and the timescale is assumed, and the luminosities are thus corrected for, the resulting luminosities fall within a narrow range, as shown in the bottom panel. Similar results have recently been obtained by the High-z Supernova Search team, as can be seen in figure figure 22. The corrected light curves have been plotted in three frequency bands (U, B and V) for both near and distant supernovae Ia (observed by the Hubble Space Telescope), and a concordance between the near and distant samples has been observed, which indicates that no dimming occurs for distant supernovae. The method is phenomenological, because there is no fundamental understanding of dependence of the absolute luminosity on shape of the light curves, and it is awaiting a confirmation from fundamental understanding of the physics of supernovae explosions.

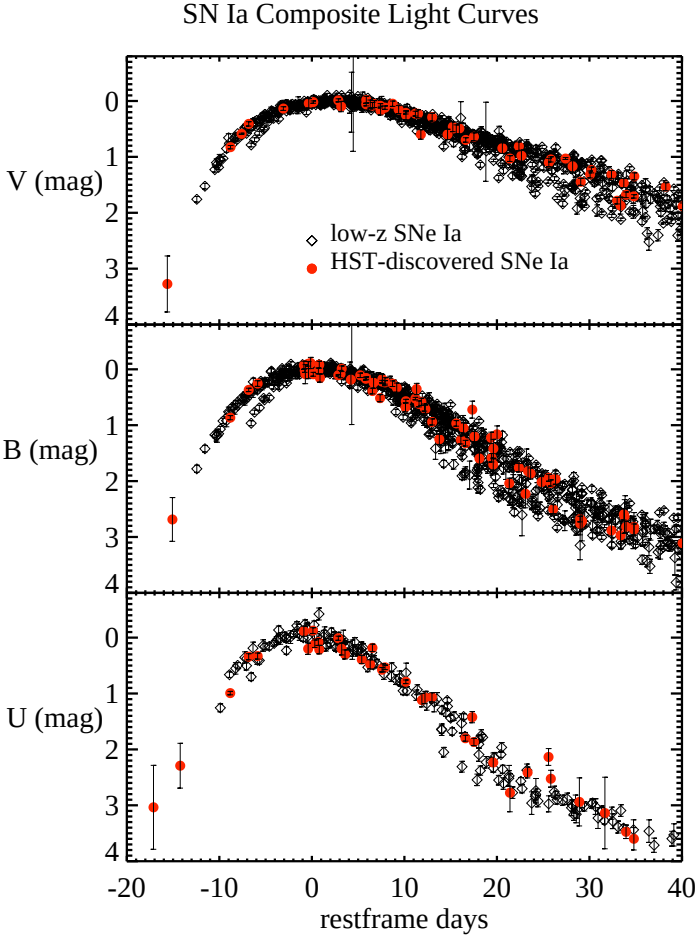


FIG. 30: A comparison of the corrected light curves for near ($z < 0.1$, open symbols) and distant supernovae Ia (observed by the Hubble Space telescope, filled symbols) in three frequency bands (B, U and V). No evolution (dimming) effects are observed for distant supernovae. From the High-z Supernova Search (A. Riess et al, 2004).

Based on the contention that distant supernovae Ia can be used as standard candles (not in any way different from nearby supernovae Ia), in 1998 the two teams, the Supernova Cosmology Project (Perlmutter et al, see figure 31) and the High-z Supernova Search (A. Riess et al) presented evidence that the Universe is accelerating.

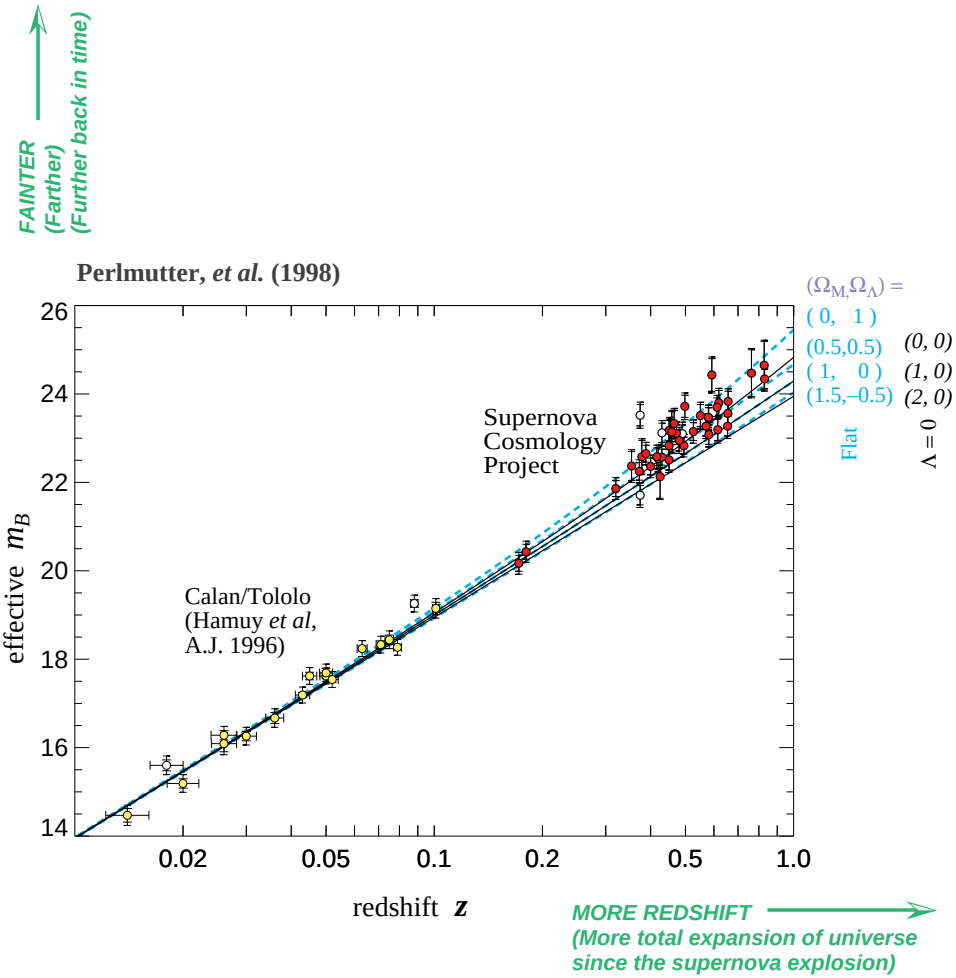
In order to get a better grasp on what it means, let us recall that in a Friedmann-Lemaître-Robertson-Walker (FLRW) homogeneous and expanding universe, with the line element,

$$ds^2 = c^2dt^2 - a^2\frac{dr^2}{1-kr^2} - a^2r^2(d\theta^2 + \sin^2(\theta)d\varphi^2) \tag{126}$$

the Einstein equations are,

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_N}{3c^2}\rho + \frac{\Lambda}{3} - \frac{kc^2}{a^2} \tag{127}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3c^2}(\rho + 3\mathcal{P}) + \frac{\Lambda}{3}\,. \tag{128}$$



In flat universe: $\Omega_M = 0.28$ [± 0.085 statistical] [± 0.05 systematic]

Prob. of fit to $\Lambda = 0$ universe: 1%

FIG. 31: An early evidence for accelerating Universe from Supernova Cosmology Project (Perlmutter et al., 1998), with about 40 supernovae Ia light curves. The diagram shows an effective apparent magnitude m_B of supernovae as a function of redshift z. Several curves of m_B as a function of z in different cosmologies are also shown.

The Universe is accelerating when the active gravitational mass is negative, $\rho + 3\mathcal{P} < 0$, and/or the cosmological term is positive, $\Lambda > 0$. While the first possibility presents an experimenter a moderate worry (nobody has ever observed such a matter in laboratory), the second option leaves an unconfortable sentiment by many theorists (a theoretical understanding of why the cosmological term is so close to zero is known the cosmological constant problem).

A convenient rewriting of equations (127–128) for today (when $a(t_0) = a_0 = 1$) is,

$$1 = \Omega_{\rm m} + \Omega_\Lambda + \Omega_k \tag{129}$$

FIG. 32: The Ω_Λ *vs.* $\Omega_{\rm m}$ plots. The left (right) panel shows the probability contours (68% and 90%) for 6 (40) distant supernovae Ia. The line $\Omega_{\rm m} + \Omega_\Lambda = 1$ is also shown, which divides an open from a closed universe. The age of the Universe is also indicated.

$$q_0 = \frac{1}{2}\Omega_{\rm m} - \Omega_\Lambda \,, \tag{130}$$

where we assumed that the matter energy density ρ is dominated by nonrelativistic particles, $\rho \simeq \rho_m$, $\mathcal{P} \simeq \mathcal{P}_m = 0$,

$$\Omega \simeq \Omega_{\rm m} = \frac{8\pi G_N}{3c^2}\,\frac{\rho_m}{H_0^2}\,, \qquad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}\,, \qquad \Omega_k = -\frac{c^2 k}{H_0^2}\,. \tag{131}$$

These equations are a convenient tool for understanding of the diagrams plotted based on distant supernovae measurements.

Assume that there is a standard candle (an object with a well understood absolute luminosity $\mathcal{L}$), which is so bright that it can be observed up to cosmological distances, that is up to a redshift of the order unity, $z \sim 1$. Assuming that supernovae Ia are standard candles, allows one to represent them on

a plot, which makes use of the nontrivial functional dependence of the luminosity distance,

$$d_L = \Big(\frac{\mathcal{L}}{4\pi\mathcal{F}}\Big)^{1/2} \tag{132}$$

on redshift z, where $\mathcal{F}$ is the measured flux. Indeed, when expanded in powers of z, at order z^2 d_L acquires a correction to the Hubble law,

$$d_L = H_0^{-1}\Big(z + \frac{1}{2}(1-q_0)z^2 + ..\Big)\,, \tag{133}$$

where

$$H_0 = \dot{a}(t_0), \qquad q_0 = -H_0^{-2}\,\ddot{a}(t_0) \qquad (a_0 = a(t_0) = 1) \tag{134}$$

and $z = a(t)^{-1} - 1$. While the linear term is sensitive to the Hubble parameter, the quadratic correction is sensitive to the Universe's acceleration, and it is quantified by the deceleration parameter, q_0.

In figure 31 we show a diagram of the effective (corrected) apparent magnitude m_B as a function of redshift z for about 40 distant supernovae Ia, which represent an early evidence for accelerating Universe from the Supernova Cosmology Project (Perlmutter et al, 1998). Several curves of m_B as a function of z in different cosmologies are also shown, with the density in nonrelativistic matter and cosmological term (relative to the critical energy density), $(\Omega_{\rm m}, \Omega_\Lambda) = (0,1), (0.5, 0.5), (1,0)$, and $(1.5, -0.5)$. All of these models correspond to universes with flat spatial sections. From figure 31 it is clear that the data support an accelerating universe with $q_0 \simeq -0.5$, such that a positive cosmological term is favoured. It is a curious fact that an empty (open) universe also fits the data reasonably well. (An empty universe is of course excluded by dynamical mass measurements, which yield $\Omega_{\rm m} \approx 0.3$.)

Based on figure 31, one can plot an $(\Omega_{\rm m}, \Omega_\Lambda)$ diagram, which is shown in figure 32, where $\Omega_{\rm m} = \rho_m/\rho_{\rm cr}$, and $\Omega_\Lambda = \Lambda/(3H_0^2)$, and $\rho_{\rm cr} = (3c^2H_0^2)/(8\pi G_N)$ denotes the critical energy density. According to the right panel, the Einstein-de Sitter Universe ($\Omega_{\rm m} = 1, \Omega_\Lambda = 0$) is excluded.

In a spatially flat universe ($\Omega_k = 0$), which is supported by the CMB measurements (WMAP), Eqs. (130), (52) and (55) then yield the following estimate for the deceleration parameter today (134),

$$q_0 = -0.60 \pm 0.05\,. \tag{135}$$

In figure 33 we show the joint confidence intervals for the *Gold Sample* of 157 supernovae Ia (solid contours) (A. Riess et al, the High-z Supernova Search, 2004), to be compared with the dotted contours (A. Riess et al, 1998). The confidence contours suggests that the Universe is accelerating ($q_0 < 0$) with more than 3σ confidence level. Assuming a universe with flat spatial sections, $\Omega_{\rm m} + \Omega_\Lambda = 1$, the best fit model gives, $q_0 \simeq -0.5$. While the data show a near degeneracy along the q_0 =constant lines, they favour slightly (at 1σ) a closed universe. This preference has by more recent measurements largely disappeared.

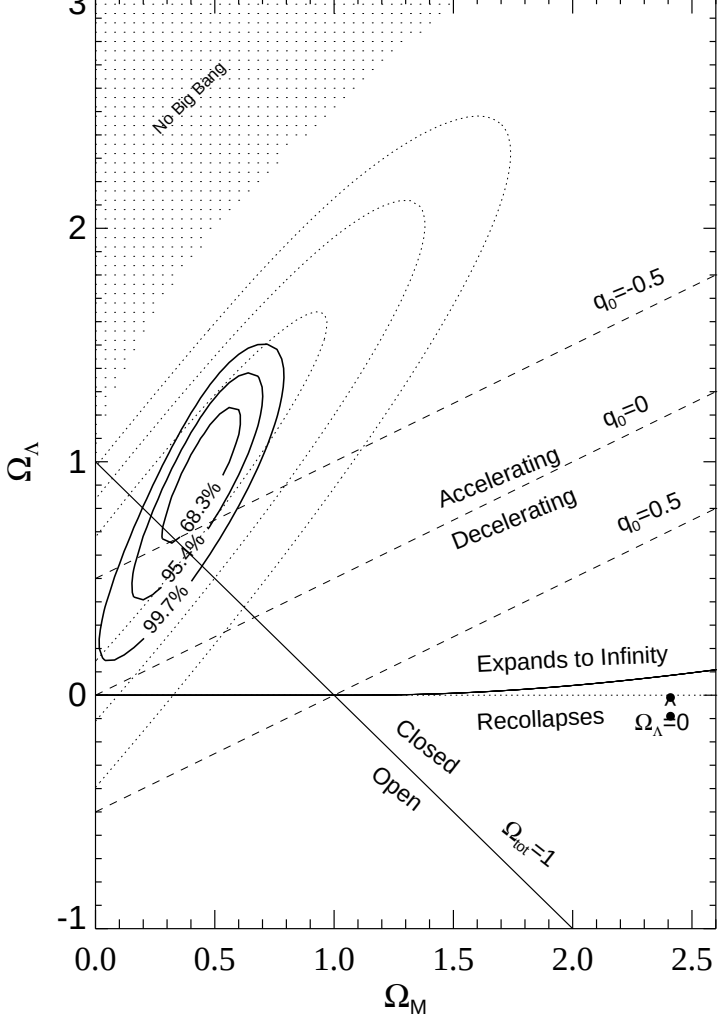


FIG. 33: Joint confidence intervals (1σ : 68%, 2σ : 95%, 3σ : 99.7%) in an $(\Omega_{\rm m}, \Omega_\Lambda)$ diagram for the *Gold Sample* of 157 supernovae Ia (solid contours) (A. Riess et al, the High-z Supernova Search, 2004). The dotted contours are from the earlier work of the Team (1998). The regions representing some specific cosmological scenarios are also shown.

In figure 34 we show the residual Hubble diagram (after the linear Hubble Law is subtracted) for the *Gold Sample* of supernovae Ia, in which cosmological models and models for astrophysical dimming are compared. While a simple model with astrophysical dimming (gray dust at large redshift) does poorly, a model with 'replenishing' gray dust does quite well in fitting the data, representing thus an alternative to acceleration. Among the cosmological models in the figure, an empty Universe ($\Omega_{\rm tot} = 0$) and the Einstein-de Sitter universe ($\Omega_{\rm m} = 1$, $\Omega_\Lambda = 0$) are both excluded by the *Gold Sample* of SNe Ia. Assuming a flat universe (which is supported by the CMB measurements), $\Omega_{\rm m} + \Omega_\Lambda = 1$ ($\Omega_k = 0$), the best fit is reached for the model with $\Omega_\Lambda \approx 0.74$ and $\Omega_{\rm m} \approx 0.25$.

In figure 35 we show a diagram of effective apparent magnitude m_B versus redshift, for a compilation of supernovae Ia measurements, which include near and distant supernovae Ia. The data support an accelerating universe. In figure 35 we illustrate how one may use the supernovae Ia measurements to

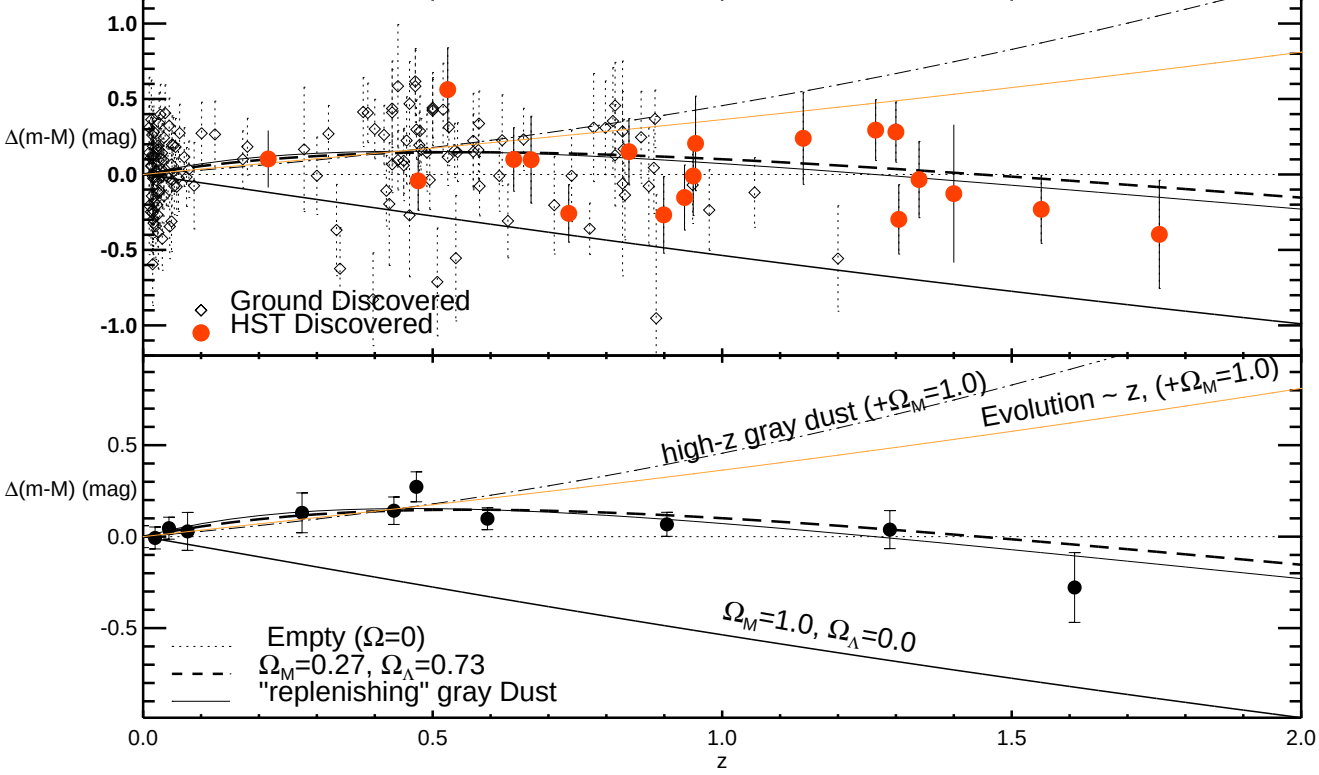


FIG. 34: Residual Hubble diagram (after the linear Hubble Law is subtracted) for the *Gold Sample* of supernovae Ia (Riess et al, 2004) comparing cosmological models and models for astrophysical dimming. Upper panel: ground based discovered SNe Ia are shown as diamonds, HST-discovered SNe Ia are shown as filled symbols. Bottom panel: Data and models are shown relative to an empty universe model ($\Omega = 0$), and for the Einstein-de Sitter Universe ($\Omega_{\rm m} = 1$, $\Omega_\Lambda = 0$).

infer about the age of the Universe, and even something about its future. These figures are taken from a popular article by Perlmutter (Physics Today 2003).

When one includes the constraints from the CMB measurements by the WMAP satellite, one gets constraints on the $\Omega_{\rm m}$ and Ω_Λ as shown in figure 37. When all data are taken together, a concordance is reached for $\Omega_\Lambda = 0.74 \pm 0.04$ and $\Omega_{\rm m} = 0.25 \pm 0.03$, such that the total energy density is consistent with a flat Universe, $\Omega_{\rm tot} = \Omega_\Lambda + \Omega_{\rm m} = 1$.

According to the CMB and distant supernovae measurements, the Universe is filled with a homogeneously distributed dark energy of unknown composition. In figure 38 we show constraints on the equation of state of the dark energy, $w = p/\rho$, as a function of matter density $\Omega_{\rm m}$ and the Hubble parameter h (in units of 100 kms^{-1}Mpc^{-1}). When the constraints from the CMB measurements are combined with the 2dF Galaxy Survey and distant supernovae, the 1σ and 2σ contours are shown in the upper right panel. When the supernova constraints are replaced by the HST Key Project constraints on the Hubble parameter, one gets the contours shown in the lower right panel. The contours favour the simplest explanation, the dark energy of the Universe is well represented by the cosmological term,

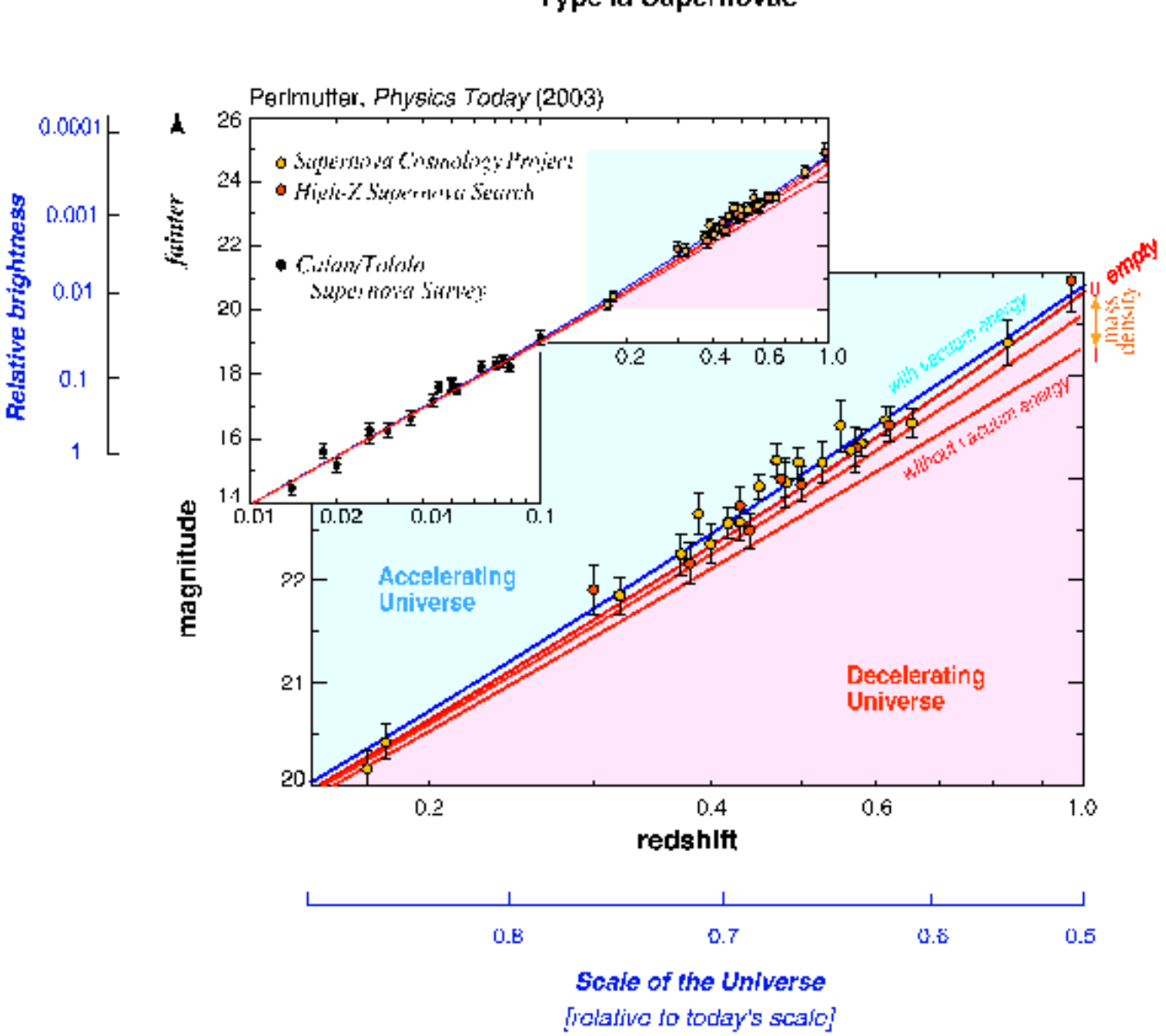


FIG. 35: A compilation of data, showing an effective apparent magnitude m_B versus redshift z, which include near SNe Ia data from the Calan/Tololo Supernova Survey (black circles) and distant supernovae from both the Supernova Cosmology Project (yelow circles) and the High-z Supernova Search (red circles) (Perlmutter, Physics Today 2003).

with an equation of state, $w_\Lambda = -1$.

An alternative explanation is a Q-matter (quintessence), according to which the Universe is accelerated by a weakly coupled homogeneous scalar field Q, with an appropriately tuned potential, $V = V(Q)$, such that at late times, the field dynamics is governed by a negative equation of state,

$$w_Q = \frac{p_Q}{\rho_Q} = \frac{\frac{1}{2}\dot{Q}^2 - V(Q)}{\frac{1}{2}\dot{Q}^2 + V(Q)} < -\frac{1}{3}\,. \tag{136}$$

According to the results shown in figure 38, these models are quite severely constrained to the current value, $w_Q \leq -0.8$ (today). The equation of state of the Q-matter is in general evolving with the Universe

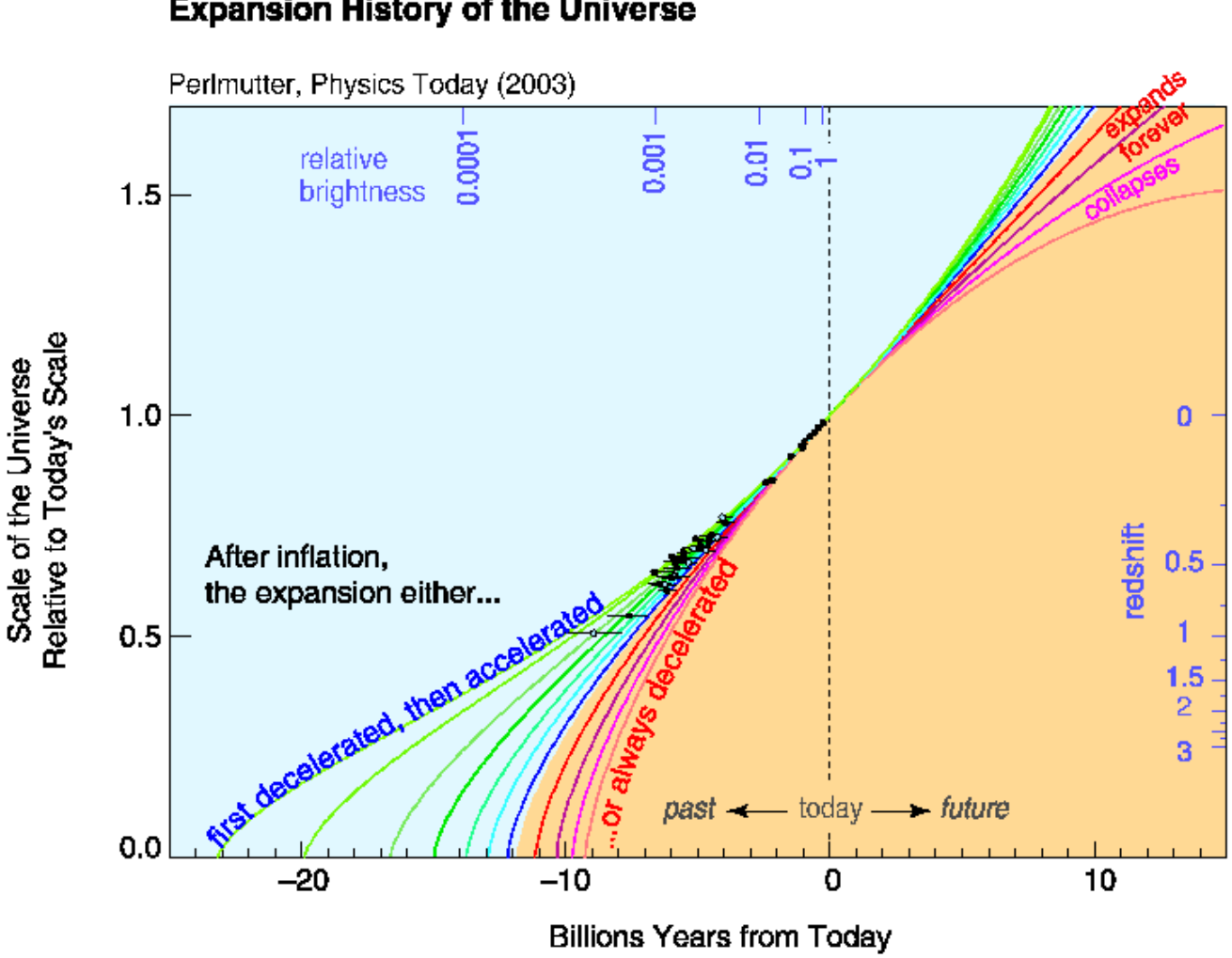


FIG. 36: An illustration how the distant supernovae data may be used to conclude about the history, age and future of the Universe (Perlmutter, Physics Today 2003).

expansion, $w_Q = w_Q(z)$. At this moment there is no experimental support for quintessence with an evolving equation of state. The fact that the data show a slight preference for $w_Q < -1$ has lead to a significant body of research on 'phantom fields' for which $w_{\rm phantom} < -1$. The more recent measurements by the Supernova Legacy Survey and by the WMAP 3 year data (2006) lend no significant support for phantom models.

C. A brief history of the Universe: the Big-Bang Model

The Universe underwent several distinct phases in its evolution, among which the following three epochs take a special role.

`Cosmic inflation` is a hypothetical epoch of an accelerated expansion of the early Universe, which is supported by the following indirect evidence: (a) the Universe has nearly flat spatial sections and (b) a nearly scale invariant spectrum of cosmological perturbations seed large scale structures

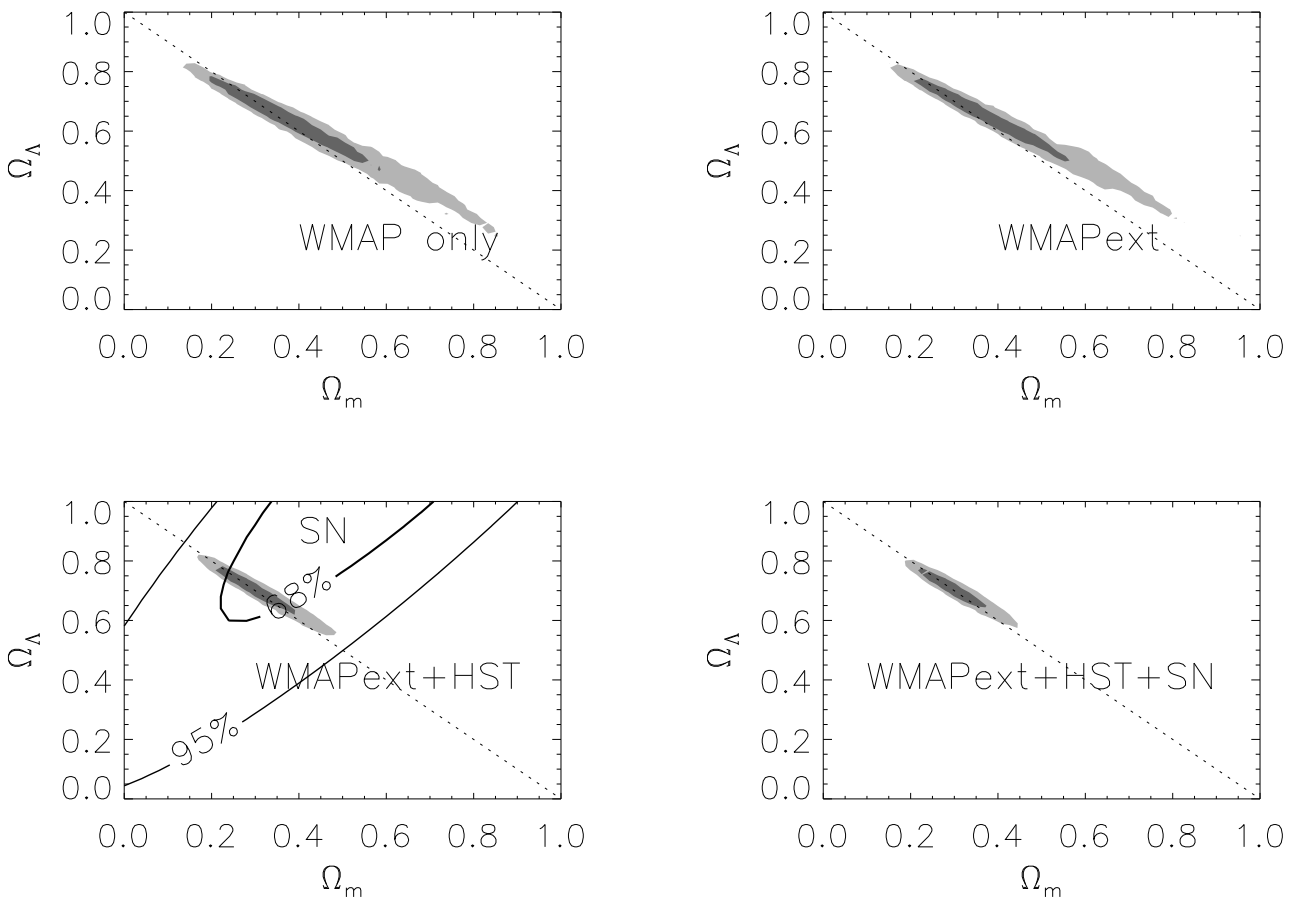


FIG. 37: Constraints on the geometry of the Universe in the $\Omega_{\rm m}$–Ω_Λ plane. The figure shows the two dimensional likelihood survfece for various combination of data: WMAP only (upper left panel), WMAP plus other CMB experiments (upper right panel), WMAP plus other CMB experiments, with the Hubble Space Telescope Key Project (HST) prior on the Hubble parameter, together with the contours from the distant supernovae Ia measurements (lower left panel), and WMAP plus other CMB experiments, with the HST prior, together with SNe Ia measurements (lower right panel) (Spergel et al, WMAP 2003).

of the Universe. During inflation energy density is dominated by a matter with negative pressure, $\mathcal{P} < -\rho/3$. In its simplest disguise energy density during inflation is approximately constant such that, according to the Friedmann equation (42), the Hubble parameter is also approximately constant and the Universe expands exponentially with cosmic time $a \propto {\rm e}^{H_I t}$, where H_I denotes the Hubble parameter during inflation.

`Radiation Era`, during which the energy density is dominated by relativisitic particles, with pressure, $\mathcal{P} = \rho/3$. Eq. (44) then implies that during radiation era $\rho \simeq \rho_{\rm rad} \propto 1/a^4$, such that the Friedmann equation (42) is solved by, $a \propto t^{1/2}$ and $H = 1/(2t)$.

`Matter Era`, during which the energy density is dominated by nonrelativisitic particles, whose pressure is negligible, $\mathcal{P} \approx 0$. From Eq. (44) we easily infer that during matter era $\rho \simeq \rho_{\rm m} \propto 1/a^3$,

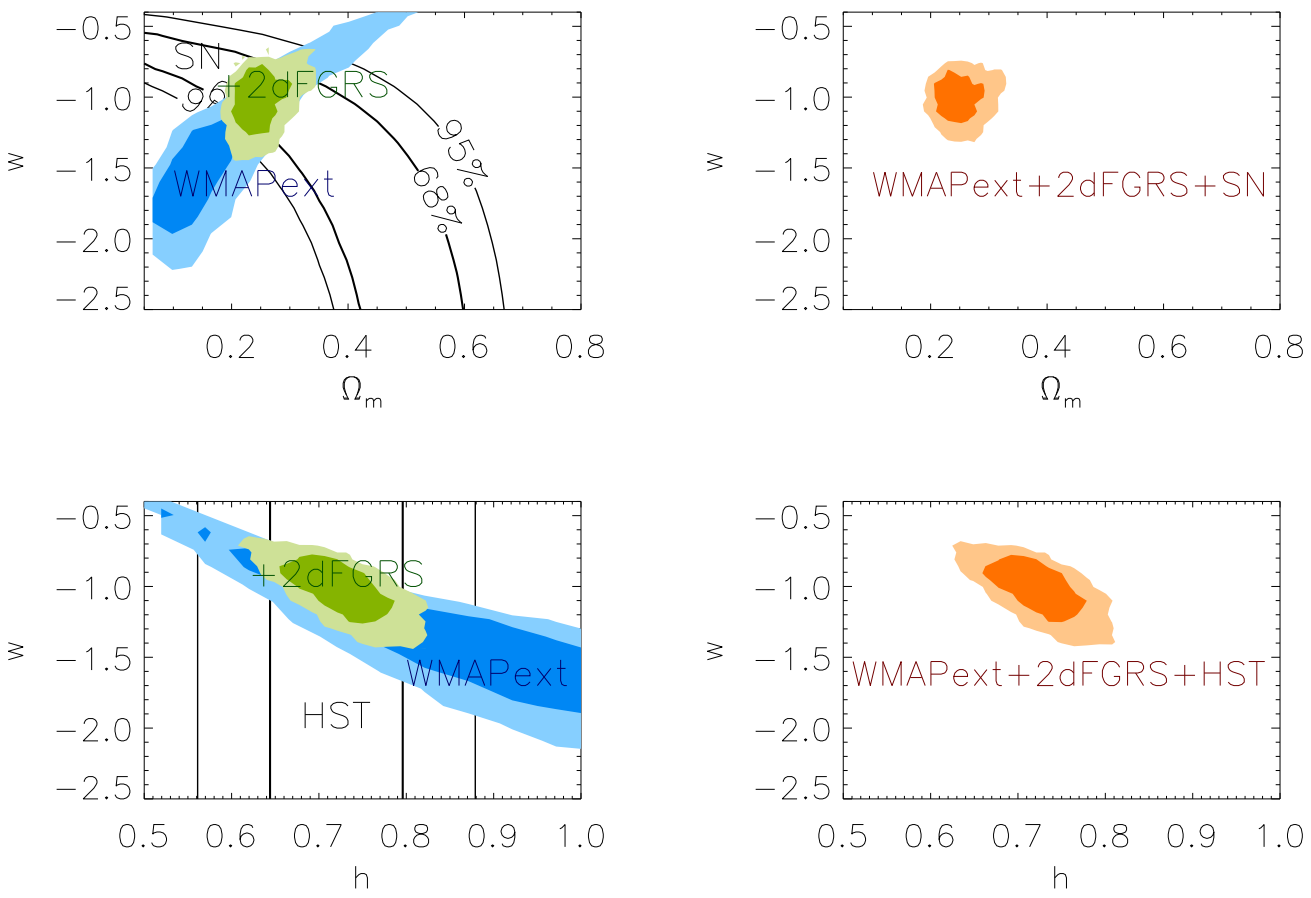


FIG. 38: Constraints on the dark energy equation of state, $w = p/\rho$ as a function of matter density Ω_m and the Hubble parameter h, in units of 100 kms^{-1}Mpc^{-1}. The upper left panel shows the 68% and 95% likelihood contours for the CMB experiments only (WMAPext) and for WMAPext plus the 2dF Galaxy Survey data, and the confidence contours from distant supernovae (Perlmuter et al. 1999). The contours in the upper right panel include the constraints from WMAPext, 2dF and supernovae Ia. In the lower panels the supernovae data are replaced with the HST Key Project constraints. From Spergel et al (WMAP 2003).

such that the Friedmann equation (42) is solved by $a \propto t^{2/3}$ and $H = 2/(3t)$.

Figure 39 shows a thermal history of the Universe. Along the vertical axis are the cosmic time, the scale factor and the (inverse) temperature (not to proportion). The epochs preceding the epoch of nucleosynthesis ($t \sim 1$ sec, $k_BT \sim 1$ MeV) are all hypothetical. For the later epochs there is a direct observational evidence. According to some ideas the Universe is begins with the 'Big Bang' at the Planck Epoch ($t \sim 10^{-43}$ sec, $k_BT \sim 10^{19}$ GeV). This epoch is highly hypothetical, if not speculative. A more modest beginning is by an epoch of cosmic inflation (not marked), which occurs at $t \sim 10^{-35}$ sec, $k_BT \sim 10^{16}$ GeV, which is also the scale of unification of electromagnetic, weak and strong forces. After inflation the Universe (re)heats, and its energy is dominated by relativistic particles. At the electroweak scale, $t \sim 10^{-8}$ sec, $k_BT \sim 100$ GeV, a thermally induced electroweak phase transition takes place, at which the gauge symmetry, $SU(3)_c \times SU(2)_L \times U(1)_Y$

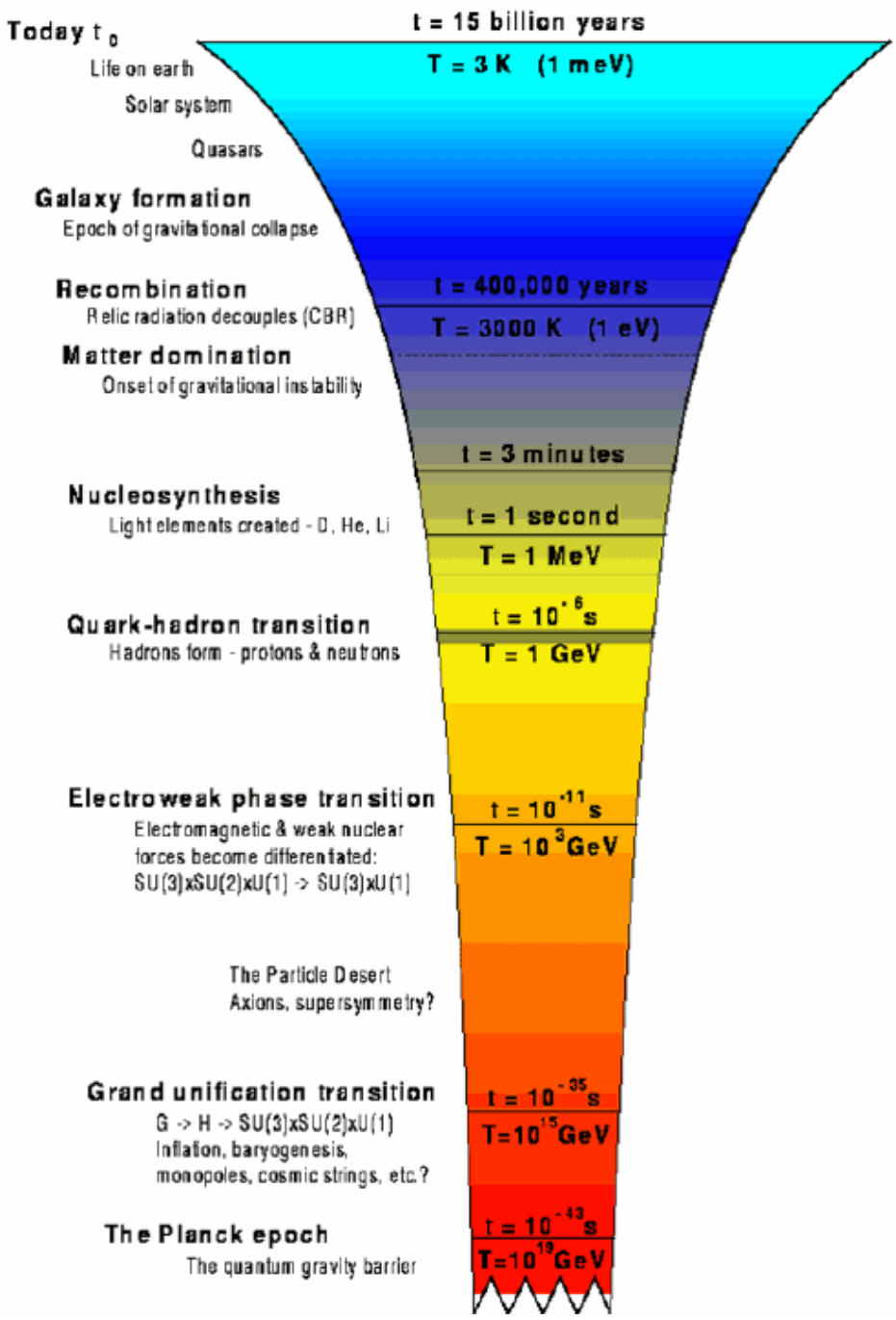


FIG. 39: A thermal history of the Universe.

is "broken" to $SU(3)_c \times U(1)_{EM}$, where $SU(3)_c$ denotes the symmetry of strong interactions, mediated by the 8 gluons, which live in the adjoint representation of $SU(3)_c$. $SU(2)_L$ is the symmetry of the W-gauge bosons, which mediate weak interactions, and $U(1)_Y$ is the symmetry of the hypercharge gauge field B. At the electroweak transition three out of the four gauge bosons of $SU(2)_L \times U(1)_Y$ symetry group acquire mass and the fourth, the photon field, remains massless. The remaining gauge symmetry is that of electromagnetism, denoted by $U(1)_{EM}$, which is still unbroken today. Since photons are massless and they interact weakly they can easily propagate to cosmological distances, such that the Universe appears transparent. Baryonic and dark matter

are probably created by a hypothetical mechanism somewhere between the scale of inflatin and the electroweak scale. $k_B T_{QCD} \simeq 160$ MeV is the scale of the QCD transition, which is a crossover (not marked). This is the scale at which the quantum chromodynamics (QCD) becomes strongly coupled, and quarks get confined by the gluon exchange into baryons and mesons.

At a scale, $k_B T \sim 1$ MeV, $t \sim 1$ sec, the protons and neutrons get bound into light elements, deuterium, helium and lithium are created (heavier elements are created in tiny unmeasurable quantities). This epoch is known as the Big Bang Nucleosynthesis (BBN). At about $k_B T \sim$ eV, $z_{\rm eq} \sim 3230$, the Universe becomes matter dominated. At $k_B T \sim 0.3$ eV ($T \sim 3000$ K, $z_{\rm rec} \simeq 1090$) the electrons and protons recombine into neutral hydrogen, and the photons start freely streaming. Finally, at $z \sim 10 - 30$ ($T \sim 100$ K) first stars, galaxies and clusters begin forming. Explosions of these early stars reionise the Universe. Finally, today $t_0 \simeq 13.7$ Gy, the structures are still in the process of formation. As of recently ($z \sim 0.4$), the Universe entered an accelerated expansion.

1. Cosmic inflation as paradigm

Cosmic inflation is an epoch of accelerated expansion of the Universe, which occured at an energy scale, $k_B T \sim 10^{16}$ GeV. During inflation the second FLRW equation (43) simplifies to,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3c^2}(\rho_I + 3\mathcal{P}_I) + \frac{\Lambda}{3} > 0\,. \tag{137}$$

An accelerated expansion can be realised either by a negative active gravitational mass, $\rho_I + 3\mathcal{P}_I < 0$, or by a positive cosmological term, $\Lambda > 0$.

In its simplest disguise, the equation of state of the inflationary matter (I-matter) is to a good approxmation that of the cosmological term,

$$\mathcal{P}_I \simeq -\rho_I\,, \tag{138}$$

such that $\rho_I + 3\mathcal{P}_I \simeq -2\rho_I$, and the Universe undergoes an approximately exponential expansion,

$$a(t) \simeq a_I \mathrm{e}^{H_I t}\,, \tag{139}$$

where a_I is a constant, and H_I denotes the Hubble parameter,

$$H_I \simeq \sqrt{\frac{8\pi G_N \rho_I}{3c^2}} \sim 10^{13}\ \mathrm{GeV}\hbar^{-1}\,. \tag{140}$$

We know from recent observations of cosmological supernovae Ia that the Universe is expanding at an accelerating pace at present, giving thus credence to an epoch of primordial inflation.

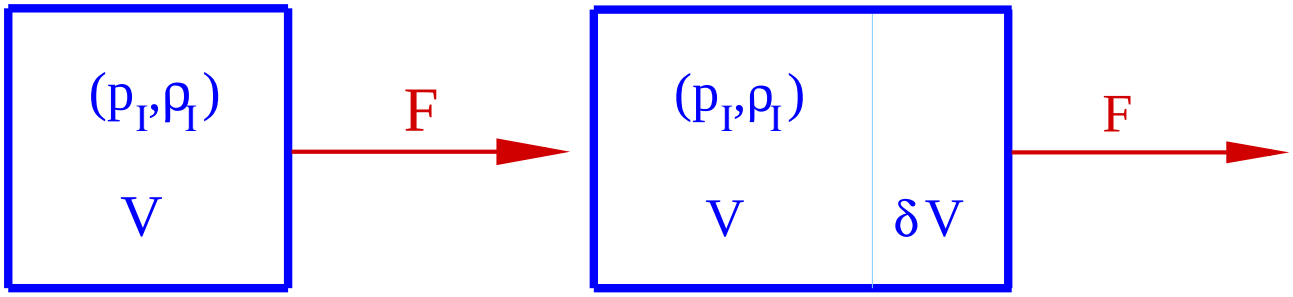


FIG. 40: The box of volume V filled with inflationary matter, $(\mathcal{P}_I, \rho_I)$, such that $\mathcal{P}_I \simeq -\rho_I$. An external force ought to do work $\delta W = -F\delta x = \mathcal{P}_I \delta V$ in order that the box expands by δV.

In order to get a feel for the properties of matter that can drive inflation, let us consider the following simple situation illustrated in figure 40. An external force $\vec{F}$ acts on a box of volume V filled with inflationary (I-)matter, $\mathcal{P}_I \simeq -\rho_I$, such that the box expands. The first law of termodynamics,

$$\delta W = \delta U + \mathcal{P}_I \delta V \,, \qquad \delta U = 0 \,, \tag{141}$$

then tells us that, if $\mathcal{P}_I < 0$, the work $\delta W < 0$ done by the piston by on the system is negative, meaning that the force actually does work on the system. This should not come as surprise, since the expansion enlarged the energy inside the box by $\delta E = \rho_I \delta V$. This energy increase is of course provided by the external force F, which does the work $F\delta x$. It turns out that is not hard to find matter with such properties: any rubber or metal object resists expansion thus effectively excerting negative pressure and negative work on the piston. Negative pressure ought to be so large in magnitude that it ought to exceed 1/3 of the total energy (which is at least the rest mass $\times c^2$). It is highly unlikely that matter with such properties can be produced in laboratory in a conceivable future.

While it is unclear how to build inflationary matter which could drive cosmic inflation in laboratory, from a theorist's perspective it is quite an "easy" task. Take for example a homogeneous scalar field, whose kinetic energy is small when compared to potential energy $\dot{I}^2/2 \ll V(I)$, such that the equation of state reads,

$$w_I = \frac{\mathcal{P}_I}{\rho_I} = \frac{\frac{1}{2}\dot{I}^2 - V(I)}{\frac{1}{2}\dot{I}^2 + V(I)} \simeq -1 + \frac{\dot{I}^2}{V(I)} \simeq -1 \,. \tag{142}$$

In this case, $\rho_I + 3\mathcal{P}_I \simeq -2V_I[1 - \dot{I}^2/V(I)] \simeq -2V_I$, which is the desired property of the I-matter.

While the above example illustrates very nicely the basic properties of matter which can drive cosmic inflation, there is nevertheless one bizarre fact. In the Universe there is no external force to pull the piston. The question is then how it is possible that, even in the absence of an external force, the Universe can still expand!

The answer is in the unusual property of the Einstein equation (137), according to which for a matter with negative active gravitational mass $\rho_I + 3\mathcal{P}_I < 0$, gravitation becomes effectively a repulsive force. The repulsive nature of gravitation during inflation then drives the accelerated expansion.

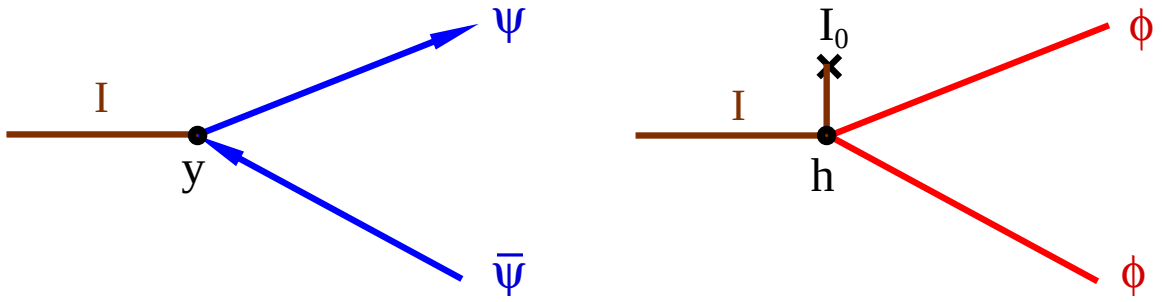


FIG. 41: Perturbative decay channels of scalar inflationary matter into fermions and scalars after inflation with the rates given in (146).

2. *Radiation era*

Preheating

At the end of inflation, the Universe is filled with matter whose energy density $\rho_I \sim 10^{16}\ \mathrm{GeV}^4(\hbar c)^{-3}$. This estimate is implied by the Hubble parameter during inflation (140),

$$H_I = \left(\frac{\hbar\rho_I}{3cM_{\rm Pl}^2}\right)^{1/2} \sim 10^{13}\ \mathrm{GeV}\hbar^{-1}\,, \tag{143}$$

where

$$M_{\rm Pl} = \left(\frac{\hbar c}{8\pi G_N}\right)^{1/2} \simeq 2.4\times 10^{18}\ \mathrm{GeV}/c^2 \tag{144}$$

denotes the reduced Planck mass, to be distinguished from the Planck mass, $m_{\rm Pl} = (\hbar c/G_N)^{1/2} \simeq 1.2\times 10^{19}\ \mathrm{GeV}/c^2$.

Assuming that the interaction lagrangian density takes the form,

$$\mathcal{L}_{\rm int} = -yI\bar\psi\psi - \frac{1}{2}hI^2\phi^2\,, \tag{145}$$

where $y \ll 1$ and $h \ll 1$ denote the (dimensionless) coupling constants, and $\psi[\mathtt{m}^{-3/2}]$ and $\phi[\mathtt{m}^{-1}]$ denote a light fermion field ($m_\psi c^2 \ll \hbar H_I$) and a light scalar field ($m_\psi c^2 \ll \hbar H_I$), respectively, to which the $I[\mathtt{m}^{-1}]$ field predominantly decays.

A perturbative estimate of the decay rate (calculated in the Born approximation) then results in the following estimate for the decay rates into fermions and scalars,

$$\begin{aligned}\Gamma_\psi &\sim \frac{|y|^2}{4\pi\hbar}E_I\\ \Gamma_\phi &\sim \frac{|h|^2}{4\pi\hbar}\frac{\hbar^2c^2|I_0|^2}{E_I}\,,\end{aligned} \tag{146}$$

where $E_I^2 = d^2V(I)/dI^2(\hbar c)^2$ defines the energy of the decaying I particles, and $I_0[\mathtt{m}^{-1}]$ denotes the amplitude of the homogeneous mode of the scalar field I. The corresponding tree-level perturbative decays are shown in figure 41. The scalar field I decays either into a fermion-antifermion pair, or into two scalar particles.

The I-matter decays when the decay rate, $\Gamma = \Gamma_\psi + \Gamma_\phi$ becomes comparable to the Hubble parameter,

$$\Gamma \sim H \,, \tag{147}$$

that is when the decay time $\tau = \Gamma^{-1}$ becomes of the order the expansion time, $t_H = 1/H$. Since the produced particles are assumed to be light ($m_\psi c^2, m_\phi c^2 \ll \hbar H$), they are relativistic, a typical energy of the produced particles is much larger than the corresponding rest energy, $E_\psi = (p^2c^2 + m_\psi^2 c^4)^{1/2} \gg m_\psi c^2$, $E_\phi = (p^2c^2 + m_\phi^2 c^4)^{1/2} \gg m_\phi c^2$, such that to a good approxmation, $E \simeq cp$, where $p = |\vec{p}|$ denotes the particle's momentum.

Assuming that the produced particles interact efficiently, they quickly thermalise, such that to a good approximation the bosons obey the Bose-Einstein distribution,

$$n_b \to n_{\rm BE} = \frac{1}{\exp\left(\frac{E-\mu}{k_B T}\right) - 1} \,, \tag{148}$$

while the fermions are Fermi-Dirac distributed,

$$n_f \to n_{\rm FD} = \frac{1}{\exp\left(\frac{E-\mu}{k_B T}\right) + 1} \,, \tag{149}$$

where for generality we allow for a nonvanishing chemical potential μ which parametrises departure from thermal equilibrium.

During this period of *preheating* or *reheating*, the entropy of the Universe increases dramatically, since the entropy of the I-matter is typically very small (I is to a nearly homogeneous field), while the entropy of a thermalised medium is maximised.

In the above we have assumed that conversion of the energy that drives inflation is perturbative. In the early 1990s it was realised that a nonperturvative conversion of the inflation energy ρ_I may be operative if inflation is driven by a scalar field, which oscillates after inflation. This mechanism is termed *parametric resonance*, and it may be operative both in the scalar and fermionic decay channels (Brandenberger, Traschen, 1992).

If ρ_I decays predominantly into very massive particles, they may be created nonrelativistic, in which case the Universe goes through an intermediate matter dominated post-inflationary epoch. While this is a theoretical possibility, there is no experimental evidence supporting such an epoch.

The energy density during radiation era can be calculated from a general expression for the stress-energy tensor, which, when expressed in terms of distribution functions, n_a, reads,

$$T^\mu_{\ \nu} = \sum_a g_a \int \frac{d^3p}{(2\pi\hbar)^3} \frac{c^2 p^\mu p_\nu}{E_a} n_a \tag{150}$$

where p^μ denotes the physical momentum, g_a is the number of degrees of freedom of species a, and $E_a = (p^2c^2 + m_a^2c^4)^{1/2}$ is the energy of species a. For an ideal fluid (39) and in the fluid rest frame, this reduces to,

$$T^\mu_\nu = \mathrm{diag}(\rho, -\mathcal{P}, -\mathcal{P}, -\mathcal{P})\,, \tag{151}$$

where ρ denotes the energy density and $\mathcal{P}$ the pressure of the fluid given by,

$$\rho \;=\; \sum_a g_a \int \frac{d^3p}{(2\pi\hbar)^3} E_a(p) n_a \tag{152}$$

$$\mathcal{P} \;=\; \sum_a g_a \int \frac{d^3p}{(2\pi\hbar)^3} \frac{c^2p^2}{3E_a(p)} n_a\,. \tag{153}$$

The integrals in (152–153) are easily evaluated in the relativistic limit for bosons and fermions, provided the distribution functions are given by (148) and (149), respectively, in which case, $E_a - \mu_a \simeq pc$. For example, for bosons, the integral in (152) can be easily reduced to the integral,

$$\int_0^\infty \frac{x^3 dx}{\mathrm{e}^x - 1} = \sum_{n=1}^\infty \int_0^\infty x^3 \mathrm{e}^{-nx} dx = 6\zeta(4)\,, \tag{154}$$

where $\zeta(4) = \pi^4/90$. $\zeta = \zeta(z)$ denotes the Riemann ζ-function defined by,

$$\zeta(z) = \sum_{n=1}^\infty \frac{1}{n^z}\,. \tag{155}$$

Similarly, for fermions the relevant integral can be reduced to the form,

$$\int_0^\infty \frac{x^3 dx}{\mathrm{e}^x + 1} = \sum_{n=1}^\infty (-1)^{n-1} \int_0^\infty x^3 \mathrm{e}^{-nx} dx = 6 \sum_{n=1}^\infty (-1)^{n-1} \frac{1}{n^4} = 6\Big(1 - \frac{1}{2^3}\Big)\zeta(4)\,. \tag{156}$$

With this, we can write for the energy density (152) and pressure (153) of a relativistic fluid,

$$\rho \;=\; g_* \frac{\pi^2}{30} \frac{(k_B T)^4}{(\hbar c)^3} \tag{157}$$

$$\mathcal{P} \;=\; \frac{\rho}{3}\,, \tag{158}$$

where we took account of $\zeta(4) = \pi^4/90$, and

$$g_* = \sum_{a=\text{bosons}} g_a + \frac{7}{8} \sum_{a=\text{fermions}} g_a \tag{159}$$

denotes the effective number of relativistic degrees of freedom. In a more general situation, the Universe may be filled with several noninteracting fluids at different temperatures, T_a. This can happen when fluids do not interact or when they interact very weakly. In this case Eqs. (157–158) are still valid, provided the definition for g_* is generalised to,

$$g_* = \sum_{a=\text{bosons}} g_a \Big(\frac{T_a}{T}\Big)^4 + \frac{7}{8} \sum_{a=\text{fermions}} g_a \Big(\frac{T_a}{T}\Big)^4\,. \tag{160}$$

Important to note is the equation of state (EOS) of relativistic fluid (157–158),

$$\mathcal{P} = w\rho\,, \qquad w = \frac{1}{3} \qquad (\texttt{relativistic fluid EOS})\,. \tag{161}$$

Another important quantity is entropy density, which in an expanding Universe can be written as,

$$s = \frac{\rho + \mathcal{P}}{T} = \frac{4}{3}\frac{\rho}{T}\,, \tag{162}$$

which is in relativistic fluid,

$$s = \frac{4}{3}\frac{\rho}{T} = k_B g_* \frac{2\pi^2}{45}\Big(\frac{k_B T}{\hbar c}\Big)^3\,. \tag{163}$$

When the Universe is filled with several noninteracting fluids at different temperatures, T_a, the entropy density grants the more general expression,

$$s = k_B g_{*S} \frac{2\pi^2}{45}\Big(\frac{k_B T}{\hbar c}\Big)^3\,. \tag{164}$$

where

$$g_{*S} = \sum_{a=\text{bosons}} g_a \Big(\frac{T_a}{T}\Big)^3 + \frac{7}{8}\sum_{a=\text{fermions}} g_a \Big(\frac{T_a}{T}\Big)^3\,. \tag{165}$$

While the energy density is not conserved in an expanding universe (recall that the stress-energy tensor is covariantly conserved, resulting in Eq. (44)), the entropy per comoving volume $S = a^3 s$ is conserved,

$$\frac{dS}{dt} = 0, \tag{166}$$

even when g_{*S} changes, which rends this quantity important for cosmology.

The entropy conservation law (166) can be obtained from the covariant energy density conservation (44), and from the second law of termodynamics for entropy,

$$dS(V,T) = \frac{\mathcal{P}dV + d(\rho V)}{T} = \frac{\mathcal{P}+\rho}{T}dV + \frac{V}{T}\frac{d\rho}{dT}dT \equiv \left(\frac{\partial S}{\partial V}\right)_T dV + \left(\frac{\partial S}{\partial T}\right)_V dT\,. \tag{167}$$

Now from $(\partial/\partial T)\partial S/\partial V = (\partial/\partial V)\partial S/\partial T$ it follows,

$$\frac{\partial}{\partial T}\Big(\frac{\mathcal{P}+\rho}{T}\Big) = \frac{\partial}{\partial V}\Big(\frac{V}{T}\frac{d\rho}{dT}\Big) \tag{168}$$

which then implies,

$$\frac{\partial \mathcal{P}}{\partial T} = \frac{\mathcal{P}+\rho}{T} = s\,. \tag{169}$$

Rewriting this as

$$\partial\mathcal{P}/\partial t = s dT/dt\,, \tag{170}$$

and adding it to Eq. (44) results in

$$\frac{\partial s}{\partial t} = -3\frac{\dot{a}}{a}s\,, \tag{171}$$

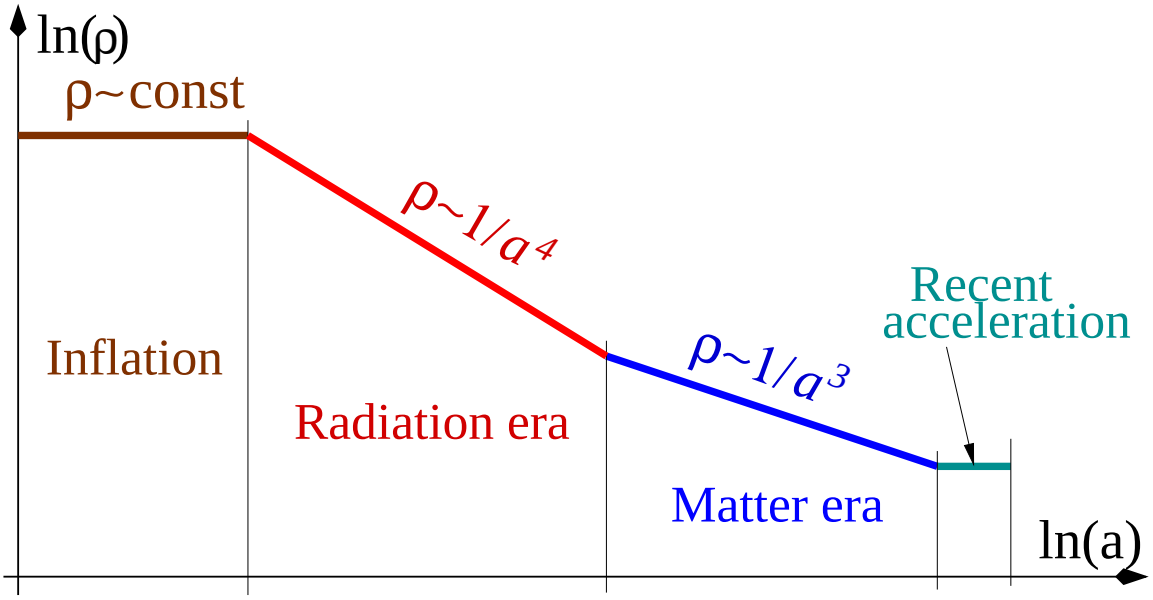


FIG. 42: The scaling of energy density during cosmic inflation, radiation era, matter era, and the recent epoch of acceleration. The figure is not to scale.

from which the entropy conservation law (166) follows.

For completeness, we also quote particle number density, which is defined as the *zeroth* component of vector current j^μ. When antiparticles are neglected one gets,

$$N = j^0 = \sum_a g_a \int \frac{d^3p}{(2\pi\hbar)^3} n_a \,. \tag{172}$$

In a relativisitic plasma in chemical equilibrium this evaluates to,

$$N = g_{*N} \frac{\zeta(3)}{\pi^2} \Big(\frac{k_B T}{\hbar c}\Big)^3 , \tag{173}$$

where $\zeta(3) \simeq 1.202$ and

$$g_{*N} = \sum_{a=\text{bosons}} g_a \Big(\frac{T_a}{T}\Big)^3 + \frac{3}{4} \sum_{a=\text{fermions}} g_a \Big(\frac{T_a}{T}\Big)^3 . \tag{174}$$

It is very easy to integrate the Friedmann equation (42) for radiation era,

$$\Big(\frac{\dot a}{a}\Big)^2 \simeq \frac{8\pi G_N}{3c^2} \rho_{\rm r} \,. \tag{175}$$

Note first that Eq. (44) implies,

$$\rho_{\rm r} + 4\frac{\dot a}{a}\rho_{\rm r} = 0, \tag{176}$$

such that

$$\rho_{\rm r} = \rho_0 \Big(\frac{a_0}{a}\Big)^4 . \tag{177}$$

In figure 42 we show the scaling of energy density during primordial inflation, radiation era, matter era, and the epoch of recent Universe's acceleration. Radiation era ends at matter radiation equality,

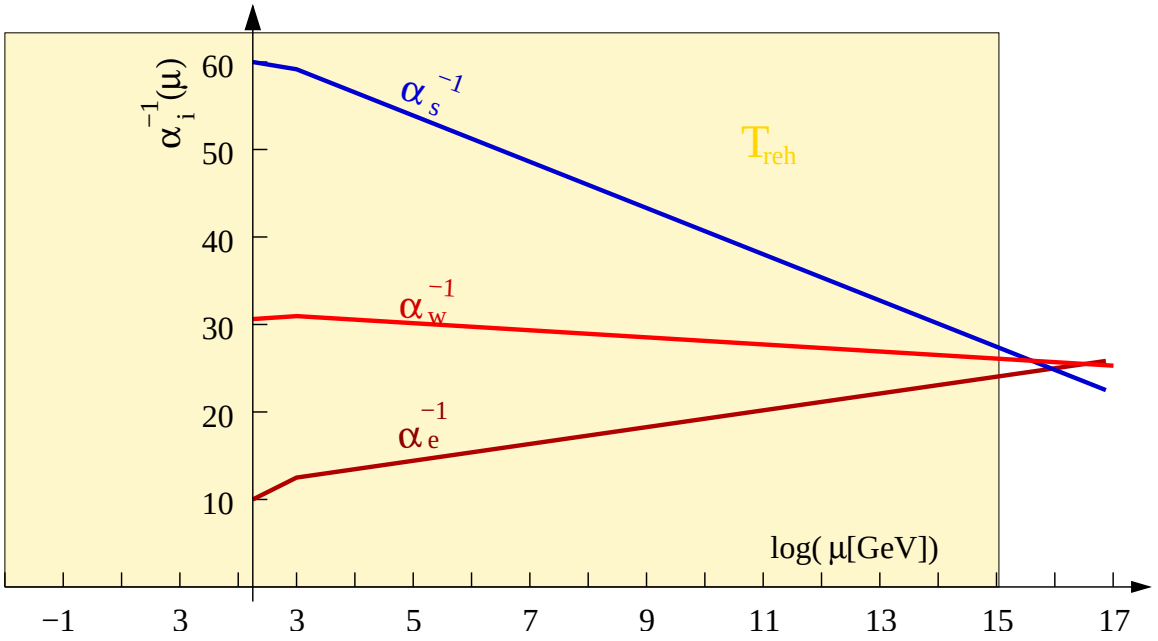


FIG. 43: The strong (α_s^{-1}), weak (α_w^{-1}) and electromagnetic (α_e^{-1}) coupling constants as a function of the logarithm of the energy scale. When supersymmetry is included, the couplings meet at the energy scale, $E = E_{\rm GUT} \sim 10^{16}$ GeV. The yellow shaded region signifies the allowed reheat temperature $T_{\rm reh}$.

at a redshift, $z_{\rm eq} \simeq 3230 \pm 210$ (WMAP constraint), and at $z \simeq 0.4$ the Universe begins to accelerate anew.

Taking account of (177), Eq. (175) can be integrated to give,

$$a = a_0 \Big(\frac{t}{t_0}\Big)^{1/2}, \qquad H = \frac{1}{2t}, \tag{178}$$

implying also that $\rho_{\rm r} \propto 1/t^2$. Note that the Hubble radius $R_H = cH^{-1} = 2ct$ grows linearly with cosmic time t during radiation era.

The scaling of the relativistic energy density (177), together with Eq. (157), imply that the temperature during radiation era scales inversely with the scale factor, $T \propto a^{-1}$. Stricly speaking, this scaling law is correct only when g_* is constant. When g_* is changing, the temperature scaling is more complicated and it can be obtained from the entropy conservation law (166), $a^3 g_{*S} T^3 = $ const., from which it follows,

$$T \propto a^{-1} g_{*S}^{-1/3}\,. \tag{179}$$

For example, when the temperature falls below the mass scale of some species, this species photodissociates into photons, damps the energy into the remaining relativistic particles, which heats up the fluid. The effective g_* is thus reduced. This heating is isentropic, and it is quantitatively expressed by (179), whereby $g_{*S} = g_{*S}(a)$ decreases with a.

Radiation era begins when most of the energy ρ_I is converted into relativistic particles with $E \simeq pc$.

This happens when (see Eqs. (42) and (144))

$$\Gamma \sim H\,, \tag{180}$$

from which one arrives at the following (perturbative) estimate of the reheat temperature,

$$k_B T_{\rm reh} \sim \Big(\frac{90}{g_*\pi^2}\Big)^{1/4}\sqrt{\hbar\Gamma M_{\rm Pl}c^2}\,, \qquad \text{when } \ \Gamma \ll H_I\,, \tag{181}$$

and

$$k_B T_{\rm reh} \sim \Big(\frac{30}{g_*\pi^2}\rho_I(\hbar c)^3\Big)^{\frac{1}{4}} = \Big(\frac{90}{g_*\pi^2}\Big)^{\frac{1}{4}}(\hbar H_I M_{\rm Pl}c^2)^{\frac{1}{2}}\,, \qquad \text{when } \ \Gamma \geq H_I\,. \tag{182}$$

Since we do not know how the inflationary energy decays, we do not know the reheat temperature $T_{\rm reh}$ of the Universe. This uncertainty in $T_{\rm reh}$ is illustrated by the yellow shaded region in figure 43. Whatever $T_{\rm reh}$ is, most likely $k_B T_{\rm reh}$ is below the Grand Unified Scale, $E_{\rm GUT} \sim 10^{16}$ GeV, which is the scale at which the strong, weak and electromagnetic couplings unify, as shown in figure 43. This can be seen by observing that the maximum reheat temperature is reached when preheating is instantaneous, $\Gamma \gg H_I$. Indeed, in this case, the reheat temperature can be estimated from (182), $T_{\rm reh} \simeq 2\times 10^{15}(100/g_*)^{1/4}$ GeV $\sim 10^{15}$ GeV, which is significantly below the grand unified scale. A consequence of this is that most of the grand unified particles with a mass comparable to the grand unified scale, $E_{\rm GUT}$, are produced at the end of inflation only in a very small amount. Strictly speaking, the lowest reheat temperature is restricted to be above the primordial nucleosynthesis scale, $T_{\rm reh} \geq E_{\rm BBN} \sim 0.1$ MeV, at which protons and neutrons combine into deuterium and helium.

Recall that the reheat temperature, at which the energy density in inflationary matter falls below the energy density in relativistic particles, is not necessarily the hottest temperature. Hence it is worth giving a lower bound on the highest temperature, $T_{\rm max}$. Even when the decay rate of the I-matter is very slow, the quantum fields couple gravitationally to the background space-time, leading to particle production during inflation. The process is operative as long as the mass of particles is below the Hubble parameter during inflation H_I, and the field coupling to gravitation is not conformal. One can show that in this case the energy density in particles at the end of inflation is of the order, $\rho \sim [H_I/(2\pi)]^4$. After inflation these particles thermalise, leading to a lower limit on the maximum temperature, $k_B T_{\rm max} \geq \hbar H_I/(2\pi)$. Since in most inflationary models, $H_I \sim 10^{13}$ GeV$/\hbar$, we arrive at a reasonable lower bound, $k_B T_{\rm max} \geq 10^{12}$ GeV. The dynamics of the Universe is governed by the I-matter until temperature drops below the reheat temperature (182). While one can conceive of inflationary models which occur at a scale significantly below $H_I \sim 10^{13}$ GeV, $k_B T_{\rm max}$ must be above the electroweak scale, $E_{\rm ew} \sim 100$ GeV, if one requires a dynamical mechanism for production of the (baryonic) matter-antimatter asymmetry of the Universe (BAU). This is so because below the electroweak scale the baryon

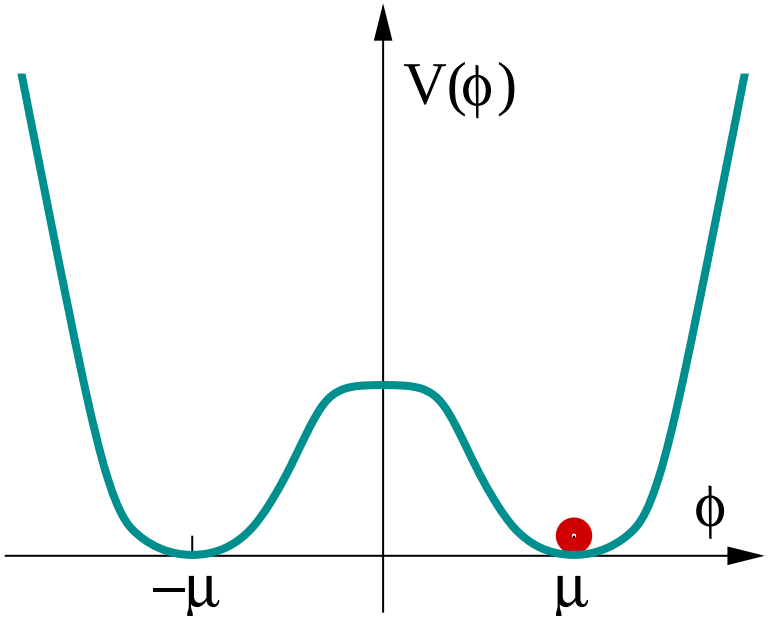


FIG. 44: Toy model potential for a real scalar field with two degenerate local minima used to illustrate a radiatively induced phase transition.

number violating processes of the Minimal Standard Model are switched off, making a dynamical BAU impossible.

Usually one assumes that both the maximum and reheat temperature are much higher than the mass scale of any of the Minimal Standard Model particles (quarks, leptons, gauge bosons, Higgs particles), such that the effective number of relativistic degrees of freedom above the electroweak scale is at least equal to the number of degrees of freedom in the Minimal Standard Model,

$$g_*|_{\rm SM} \simeq 107\,. \tag{183}$$

If supersymmetry is realised at the TeV scale, then g_* above the TeV scale is about $2g_*|_{\rm SM}$. Since the thermal occupancy of bosons and fermions (148–149) is different, supersymmetry is not manifest during the thermal history of the Universe.

Electroweak phase transition.

Radiative (quantum loop) finite temperature corrections can induce phase transitions (Kirzhnits, Linde, 1972). To illustrate this, consider the following tree-level potential for a real scalar field,

$$V(\phi) = \lambda(\phi^2 - \mu^2)^2\,, \tag{184}$$

such that the potential has two degenerate local minima at $\phi = \pm\mu$, and one local maximum, $\phi = 0$, as shown in figure 44.

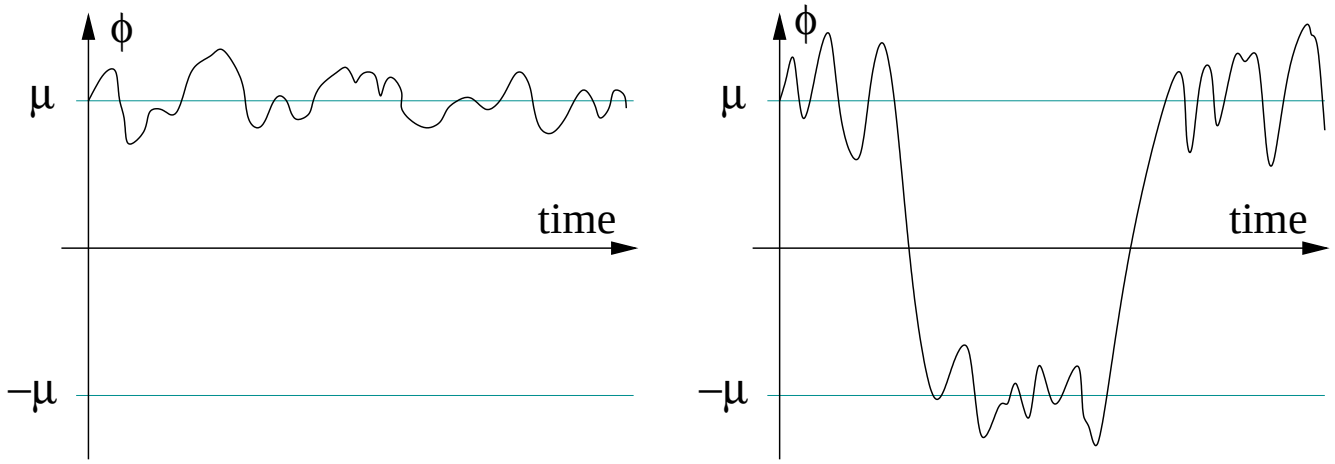


FIG. 45: The time evolution of the field ϕ at a finite temperature T with the potential (184). The field fluctuates around the minimum at $\phi = \mu$. As the temperature increases, so does the amplitude of fluctuations. On the left panel temperature is small such that the amplitude of fluctuations is small, and the field is confined to fluctuate around the minimum $\phi \sim \mu$. On the right panel temperature is comparable to the critical temperature T_c, and the amplitude of fluctuations is large. From time to time the field fluctuates over the potential barrier into the other minimum. When $T \gg T_c$ transitions between the two minima become rapid.

Even at zero temperature, the field fluctuates around one of the minima (say $\phi = \mu$) and occasionally tunnels to the other minimum. When the potential barrier between the two minima is large, this quantum tunneling process is exponentially suppressed. Let us decompose the field into a slowly changing mean field, ϕ_0, and the fast changing fluctuations, φ, as follows, $\phi = \phi_0 + \varphi$. As the temperature increases, so does the amplitude of fluctuations, φ. The fluctuations couple to ϕ_0 through the interaction terms, $\delta V = 6\lambda\phi_0^2\varphi^2 + 4\lambda\phi_0\varphi^3$, which 'kick' ϕ_0 more and more violently as the temperature increases, as illustrated in the left panel in figure 45. One can think of φ as particles which hit ϕ_0, such that ϕ_0 begins fluctuating in time. The kicks become stronger as the temperature increases. At a certain critical temperature T_c, fluctuations of the field become so large, that the kicks induce transitions between the two minima, and the field begins to fluctuate around both minima, as shown in the right panel of figure 45. As the temperature increases further, the transitions between the two minima become more and more rapid. When averaged over a large time interval Δt, the field expectation value vanishes, $\lim_{\Delta t\to\infty}\langle\phi\rangle_{\Delta t} = 0$.

The physical picture is as follows. At low temperatures, $T \ll T_c$, the field expectation value remains confined to one of the two minima. Even though the Lagrangian is symmetric under $\phi \to -\phi$, the symmetry is (spontaneously) broken by a definite choice of the vacuum, $\phi = +\mu$ or $\phi = -\mu$. At temperatures above the critical temperature, $T \geq T_c$, the symmetry is restored, and the field expectation value vanishes, $\langle\phi\rangle = \phi_0 = 0$.

A quantitative description of such phase transitions is provided by the average action (average poten-

tial). If one is interested in a phenomenon that occurs at a scale ℓ which corresponds to a momentum, $p = h/\ell$ (here h denotes the Planck constant) and an energy $E = \sqrt{p^2c^2 + m^2c^4}$, one constructs an average action S_p by averaging (smoothing) the fields on scales up to the scale ℓ, or equivalently by integrating out the modes whose momentum is larger than p. The mathematical formalism, based on which one can construct $S_p = S_p[\phi]$ is known as the renormalisation group (RG) procedure (Wilson), and it is studied in many field theory textbooks. Alternatively, one can construct a thermal effective action, $S_T = S_T[\phi]$, which is obtained by a standard procedure of Legendre transform of the tree-level action. The thermal effective action corresponds to the effective action obtained by calculating thermal loop corrections to the tree-level action, whereby one uses thermal propagators. While S_p can be used to study the phenomena which correspond to momenta smaller than p, the effective action $S_T = S_T[\phi]$ is suitable for studying the phenomena whose momenta are smaller than about $\lambda^n k_B T$, where λ denotes the relevant coupling constant, and $n = 1, 2, ..$ is an integer. For example, when all contributions with momenta $p \gg \lambda k_B T$ to the effective action S_T are taken account of by integrating the appropriate loop contributions, than S_T is primarily suitable for studying phenomena with $p \leq \lambda k_B T$. A systematic techniques for resumming the relevant finite temperature loop diagrams to get thermal effective actions for interacting (non-Abelian) gauge fields have been developed by Braaten and Pisarski and by Bödeker in the 1990s, and we shall not study them further here (see *e.g.* Michel Le Bellac, "Thermal Field Theory," Cambridge University Press, 1996).

In order to illustrate a calculation of an effective action, let us consider the example of a real scalar field with the following tree-level action,

$$\begin{aligned} S_0[\phi] &= \int d^4x \mathcal{L}_0 \\ \mathcal{L}_0 &= \frac{1}{2}\eta^{\mu\nu}(\partial_\mu\phi)(\partial_\nu\phi) - V(\phi)\,, \qquad \eta^{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)\,, \end{aligned} \tag{185}$$

where $\mathcal{L}_0$ denotes the tree-level Lagrangian, and the potential reads,

$$V(\phi) = \bar{V}_0 + \frac{1}{2}\frac{m_0^2c^2}{\hbar^2}\phi^2 + \frac{\lambda}{4!}\phi^4\,. \tag{186}$$

When $m_0^2 \geq 0$, the field expectation value, $\langle\phi\rangle = 0$, such that the field is in its symmetric minimum, $\langle\phi\rangle = 0$. If, on the other hand,

$$m_0^2 = \hbar^2 c^{-2}\frac{d^2V}{d\phi^2}(\phi = 0) < 0\,, \tag{187}$$

the Z_2-symmetry, $\phi \to -\phi$, is spontaneously broken by the vacuum, such that the scalar vacuum corresponds to one of the two minima, $\phi = \pm\mu$, $\mu^2 = -6m_0^2c^2/(\lambda\hbar^2)$. In this case it is convenient to consider the action for a shifted scalar field,

$$\phi = \phi_0 + \varphi\,. \tag{188}$$

Note that the shift corresponding to the true minimum of the potential is given by,

$$\phi_0^2 = \mu^2 \equiv -\frac{c^2}{\hbar^2}\frac{6m_0^2}{\lambda} > 0\,. \tag{189}$$

The quadratic part of the shifted action reads,

$$\begin{aligned}\int d^4x\mathcal{L}_0 \;&=\; \int d^4x \int d^4y \frac{1}{2}\varphi(x) i\mathcal{D}^{-1}(x-y;\phi_0)\varphi(y) + O(\varphi^3) \\ i\mathcal{D}^{-1}(x-y;\phi_0) \;&=\; \Big(-\Box_x - \frac{c^2}{\hbar^2}m^2\Big)\delta^4(x-y)\,, \qquad m^2 = m^2(\phi_0) = m_0^2 + \frac{1}{2}\lambda\hbar^2c^{-2}\phi_0^2\,,\end{aligned} \tag{190}$$

where $\Box_x = \partial_x^2 = c^{-2}\partial_t^2 - \nabla_{\vec{x}}^2$. At the true minimum, $m^2 = -2m_0^2 > 0$. The thermal scalar (Feynman) propagator is defined as,

$$D_T(x-y) = \frac{\mathrm{Tr}[\mathrm{e}^{-\beta\hat{H}}\mathcal{T}(\phi(x)\phi(y))]}{\mathrm{Tr}[\mathrm{e}^{-\beta\hat{H}}]}\,, \tag{191}$$

where $\mathcal{T}$ denotes the time ordering operation,

$$\hat{H} = \hbar c\int d^3x\Big[\frac{1}{2c^2}(\dot{\phi})^2 + \frac{1}{2}(\nabla\phi)^2 + V(\phi)\Big] \tag{192}$$

is the Hamiltonian operator, $\dot{\phi} = \partial_t\phi$ and $\beta = 1/(k_BT)$. The propagator (191) of the shifted theory (190) obeys the equation of motion (*cf.* $i\mathcal{D}^{-1}(x-y;\phi_0)$ in Eq. (190)),

$$(-\Box_x - \hbar^{-2}m^2c^2)D_T(x-y) = i\delta^4(x-y) \tag{193}$$

with the boundary condition,

$$D_T(z^0 = 0, \vec{z}) = D_T(z^0 = -i\hbar c\beta, \vec{z}) \qquad (z = x-y)\,, \tag{194}$$

such that in the imaginary time formalism (see *e.g.* Le Bellac, "Thermal Field Theory," Cambridge University Press, 1996),

$$iD_T(x-y) = \frac{k_BT}{\hbar c}\sum_{n=-\infty}^{\infty}\int\frac{d^3p}{(2\pi\hbar)^3}\mathrm{e}^{-i\hbar^{-1}[2\pi nk_BT(t_x-t_y)-(\vec{x}-\vec{y})\cdot\vec{p}]}iD_T(p)\,, \tag{195}$$

with

$$iD_T(p) = \frac{c^2\hbar^2}{-\omega_n^2 + c^2\vec{p}^2 + m^2c^4}\,, \tag{196}$$

where $\omega_n = 2\pi i n k_BT$ $(n = 0, \pm1, \pm2, ..)$ denote the imaginary Matsubara frequencies.

This propagator can be used to calculate a thermal effective action at any loop order. In order to illustrate how to calculate the one-loop thermal contribution to the effective action whose Feynman graph representation is shown in figure 46, we consider a scalar field theory whose classical (tree level) action reads,

$$S_0[\phi] = \int d^4x\sqrt{-g}\left(\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right)\,. \tag{197}$$

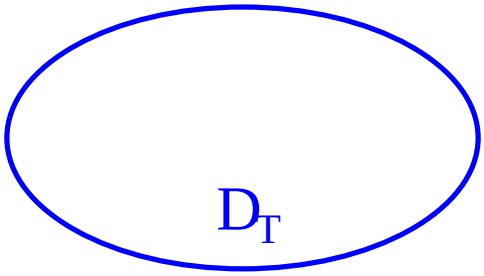


FIG. 46: The one-loop Feynman diagram contributing at order $\hbar$ to the effective action of a real scalar theory.

When ϕ is expanded into an averaged (ϕ_0) and fluctuating part φ (*cf.* Eq. (198)),

$$\phi = \phi_0 + \varphi \tag{198}$$

the action (197) splits into three parts,

$$S_0[\phi] = S_0[\phi_0] + S_\varphi[\phi_0, \varphi] + S_{\rm int}[\phi_0, \varphi] \tag{199}$$

where $S_0[\phi_0] = S_0[\phi = \phi_0]$ and

$$S_\varphi[\phi_0, \varphi] \;=\; \int d^4x\sqrt{-g}\left(\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\varphi - \frac{1}{2}c^2\hbar^{-2}m_\phi^2(\phi_0)\varphi^2\right) \tag{200}$$

$$S_{\rm int}[\phi_0, \varphi] \;=\; \int d^4x\sqrt{-g}\left(-\frac{1}{6}\partial_\phi^3 V(\phi_0)\varphi^3 - \frac{1}{4!}\partial_\phi^4 V(\phi_0)\varphi^4 - \dots\right)\,, \tag{201}$$

where $m = \hbar^2 c^{-2}\partial_\phi^2 V_{\phi=\phi_0}$ defines the scalar field mass.

The effective action is obtained from the tree level action by integrating the fluctuating field φ,

$$\mathrm{e}^{i\Gamma[\phi_0]} \;=\; \int \mathcal{D}\varphi \exp\left(iS_0[\phi_0] + S_\varphi[\phi_0, \varphi] + S_{\rm int}[\phi_0, \varphi]\right)\,. \tag{202}$$

To calculate the one-loop contribution to the effective action, note first that $S_{\rm int}$ in Eq. (202) can be neglected, simply because the interactions (201) do not contribute to the bubble diagram in figure 46. The integral $\int \mathcal{D}\varphi$ denotes the Feynman path integral over all field configurations. Since S_φ is quadratic in the fluctuating field, the resulting integral can be reduced to a product of gaussian integrals. Upon evaluation one obtains the standard result,

$$\mathrm{e}^{i\Gamma[\phi_0]} = \mathrm{e}^{iS_0[\phi_0]}\left[\det\left(\sqrt{-g}(-\Box - c^2\hbar^{-2}m^2(\phi_0)\right)\right]^{-1/2} \times \text{higher loops}\,, \tag{203}$$

where $\Box = (-g)^{-1/2}\partial_\mu(-g)^{1/2}g^{\mu\nu}\partial_\nu$ denotes the d'Alambertian operator.

The determinant in Eq. (203) can be exponentiated resulting in the following contribution to the effective action,

$$\delta\Gamma_1[\phi_0] = \frac{i}{2}\mathrm{Tr}\ln\left[\sqrt{-g}\left(-\Box - c^2\hbar^{-2}m^2\right)\right]\,, \tag{204}$$

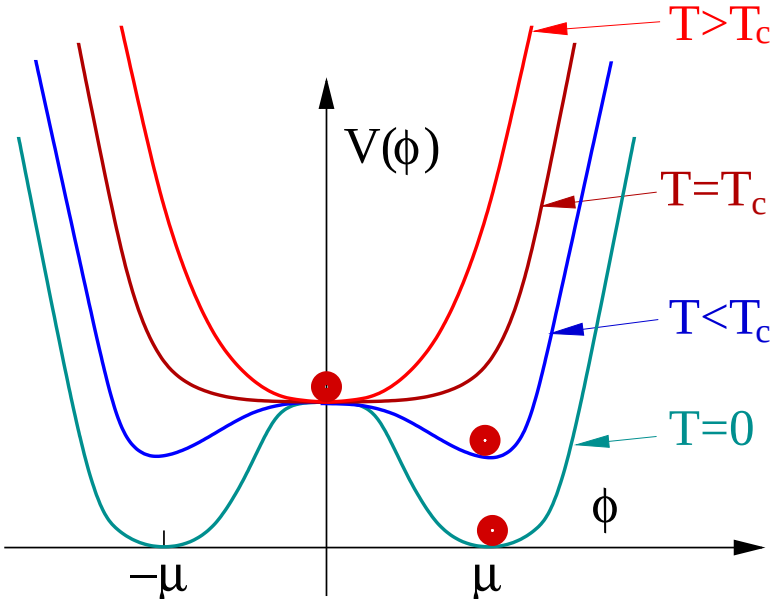


FIG. 47: Toy model potential for a real scalar field with two degenerate local minima used to illustrate a radiatively induced phase transition. Crossover or a second order transition.

where here Tr stands for the integral $\int d^4x$. This can be evaluated by observing that

$$c^{-2}\hbar^2\frac{\partial\delta\Gamma_1[\phi_0]}{\partial m^2} = -\frac{1}{2}\langle x|\frac{i}{\sqrt{-g}\,(-\Box - c^2\hbar^{-2}m^2)}|x\rangle = -\frac{1}{2}i\Delta(x;x)\,. \tag{205}$$

where $i\Delta(x;x)$ denotes the scalar field propagator evaluated at the coincidence $x'^\mu = x^\mu$.

We shall now evaluate (205) by making use of the thermal propagator (191) and in Minkowski space. Even though we want to describe a rapidly expanding Universe, this simplification is justified since we are interested in describing the situations where the processes which lead to thermalisation are much faster than the expansion rate of the Universe. At the electroweak scale this is justified since the expansion rate is $\hbar H_{\rm EW} \sim (k_B T_{\rm EW})^2/(M_P c^2) \sim 10^{-5}$ eV, which is indeed much smaller than the equilibration rate which is of the order $\Gamma_{\rm EW} \sim 10 - 100$ GeV. The one-loop contribution to the effective potential for a real scalar field (up to an unphysical constant) reads,

$$V_{T1}(\phi) = \frac{1}{2}\frac{k_B T}{\hbar c}\sum_{n=-\infty}^{\infty}\int\frac{d^3p}{(2\pi\hbar)^3}\ln\left[i\tilde{\mathcal{D}}^{-1}(k;\phi_0)\right] \tag{206}$$

where

$$i\tilde{\mathcal{D}}^{-1} = \omega_n^2 - \vec{p}^{\,2}c^2 - m^2c^4 \tag{207}$$

denotes the inverse propagator (190) in momentum space.

The integral in (206) can be evaluated to yield (see *Problem 2.11*),

$$V_{T1} = \Delta V_{0,\rm vac} + \Delta V_{T1}\,, \tag{208}$$

where

$$\Delta V_{0,\rm vac} = \int\frac{d^3p}{(2\pi\hbar)^3}\frac{E}{2c\hbar}\,, \qquad E = (p^2c^2 + m^2c^4)^{1/2} \tag{209}$$

denotes the contribution from zero temperature vacuum fluctuations, which is usually combined with $\bar{V}_0$ in Eq. (186) to a finite constant potential term, $V_0 = \bar{V}_0 + \Delta V_{0,\rm vac}$. The latter term in (208) corresponds to the contribution from the purely thermal fluctuations, which is finite,

$$\Delta V_{T1} \;=\; \frac{(k_BT)^4}{2\pi^2(\hbar c)^4}\int_0^\infty x^2dx\ln\Big[1-\exp\Big(-(x^2+m^2c^4/(k_BT)^2)^{1/2}\Big)\Big]\,. \tag{210}$$

In the high temperature limit, $k_BT \gg mc^2$, this can be expanded in a Taylor series in $mc^2/(k_BT) \ll 1$ to yield,

$$\begin{aligned}\Delta V_{T1}(\hbar c)^4 \;&=\; -\frac{\pi^2}{90}(k_BT)^4 + \frac{1}{24}m^2c^4(k_BT)^2 - \frac{1}{12\pi}m^3c^6k_BT - \frac{1}{64\pi^2}m^4c^8\Big[\ln\Big(\frac{m^2c^4}{(k_BT)^2}\Big) - c_0\Big] \\ &+\; O(m^6c^{12}/(k_BT)^2)\,,\end{aligned} \tag{211}$$

where $m^2 = m_0^2 + \lambda\hbar^2c^{-2}\phi_0^2/2$, $c_0 = (3/2) + 2\ln(4\pi) - 2\gamma_E \simeq 5.4076$, $\gamma_E \equiv -\psi(1) = 0.577215..$ denotes the Euler constant, and $\psi(z) = d[\ln\Gamma(z)]/dz$, $\Gamma(z) = (z-1)!$ denotes the (factorial) Γ-function. Combining all terms together, we get for the 1-loop thermal effective potential,

$$\begin{aligned}V + V_{T1} \;&=\; V_0(T) + \frac{1}{2}\frac{m_0^2c^4 + \frac{\lambda}{24}(k_BT)^2}{(\hbar c)^2}\phi_0^2 - \frac{1}{12\pi}\Big(m_0^2c^4 + \frac{1}{2}\lambda\hbar^2c^2\phi_0^2\Big)^{3/2}\frac{k_BT}{(\hbar c)^4} \\ &+\; \frac{\lambda}{4!}\phi_0^4 - \frac{1}{64\pi^2}\frac{m^4(\phi_0)c^8}{(\hbar c)^4}\Big[\ln\Big(\frac{m^2(\phi_0)c^4}{(k_BT)^2}\Big) - c_0\Big] + O(m^6(\phi_0)c^{12}/(k_BT)^2)\,,\end{aligned} \tag{212}$$

where

$$V_0(T) = V_0 - \frac{\pi^2}{90}\frac{(k_BT)^4}{(\hbar c)^4} + \frac{1}{24}m_0^2c^4\frac{(k_BT)^2}{(\hbar c)^4}\,. \tag{213}$$

Strictly speaking this effective potential is applicable only when, $m^2 \geq 0$. When $m^2 < 0$, V_{T1} becomes complex-valued, which illustrates the difficulties one encounters when trying to study finite temperature induced phase transitions by the means of perturbative techniques. Nevertheless, some generic properties of the transition can be obtained by analysing the effective potential (212).

The question we are interested in is the dynamics of the scalar field governed by the action (185) in a finite temperature bath, whose temperature slowly decreases, as it is the case in an expanding universe. Snapshots of the effective potential at different temperatures are shown in figures 47 and 48. At very high temperatures, $(k_BT)^2 \gg -(24/\lambda)m_0^2c^4$, the effective mass squared around the origin $\phi_0 = 0$ is positive, and the symmetry is restored. As the temperature drops below the critical temperature a second (deeper) minimum develops, and the field acquires an expectation value. As the temperature drops further the effective potential approaches the zero temperature potential, with $\phi_0 = \pm\mu$ defined in Eq. (189).

A thermally induced transition can be either a crossover or a first order transition, a second order transition being the limiting case. Let us first consider a crossover transition, as shown in figure 47.

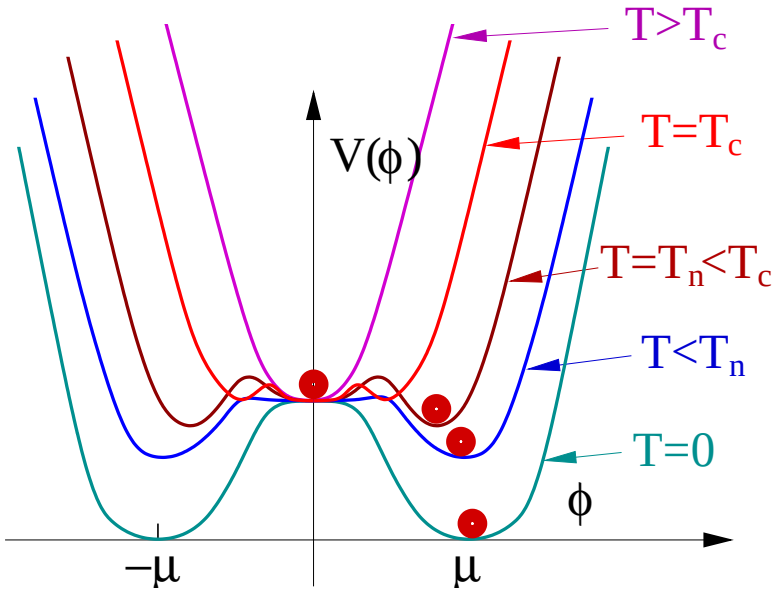


FIG. 48: Toy model potential for a real scalar field with two degenerate local minima used to illustrate a radiatively induced phase transition. First order transition.

(As regards the effective potential, a second order transition looks alike.) At high temperatures, the term in (211), which is proportional to m^2, dominates over the tree level mass term $\propto m_0^2$ in Eq. (186), such that the symmetry is restored. As the temperature drops to a critical value, T_c, at which the effective mass of scalar excitations vanishes, the field develops large scale fluctuations, still retaining zero expectation value. Below the critical temperature, $T < T_c$, the field begins "rolling down the hill", and develops an expectation value, resulting thus in a spontaneous breakdown of the Z_2-symmetry. As the temperature tends to zero, the field expectation value approaches one of the two minima, $\phi_0 = \pm\mu$.

In figure 48 we show snapshots of a first order transition. The main difference with respect to figure 47 are the secondary minima of the effective potential, which occur slightly above the critical temperature, T_c. The secondary minima are primarily generated by the negative cubic term in (212),

$$-B_T k_B T \phi = -\frac{1}{12\pi}\frac{m^3 c^6}{(\hbar c)^4} k_B T \approx -\frac{1}{12\pi}\left(\frac{\lambda}{2}\right)^{3/2}\frac{k_B T}{\hbar c}\phi_0^3\,, \tag{214}$$

which is hence responsible for thermally induced first order phase transitions. When the one-loop calculation is repeated for fermions, no cubic term of the form (214) is found, implying that, when treated in the high-temperature approximation, fermions cannot induce a first order transition. The cubic term namely arises due to the infrared singularity in the Bose-Einstein distribution, $n_{\rm BE} = 1/(\mathrm{e}^{\beta E} - 1) \to k_B T/E$ when $E \to 0$, which is absent in the Fermi-Dirac distribution. At the critical temperature, $T = T_c$, all minima are degenerate, but the field remains trapped at the origin, $\phi_0 = 0$, by the barrier generated by the cubic term (214). As the temperature drops, the secondary minima become deeper than the minimum at $\phi = 0$, and the Universe *supercools*. As the temperature drops further, the potential barrier between the minima decreases, reaching eventually a nucleation temperature, T_n, at

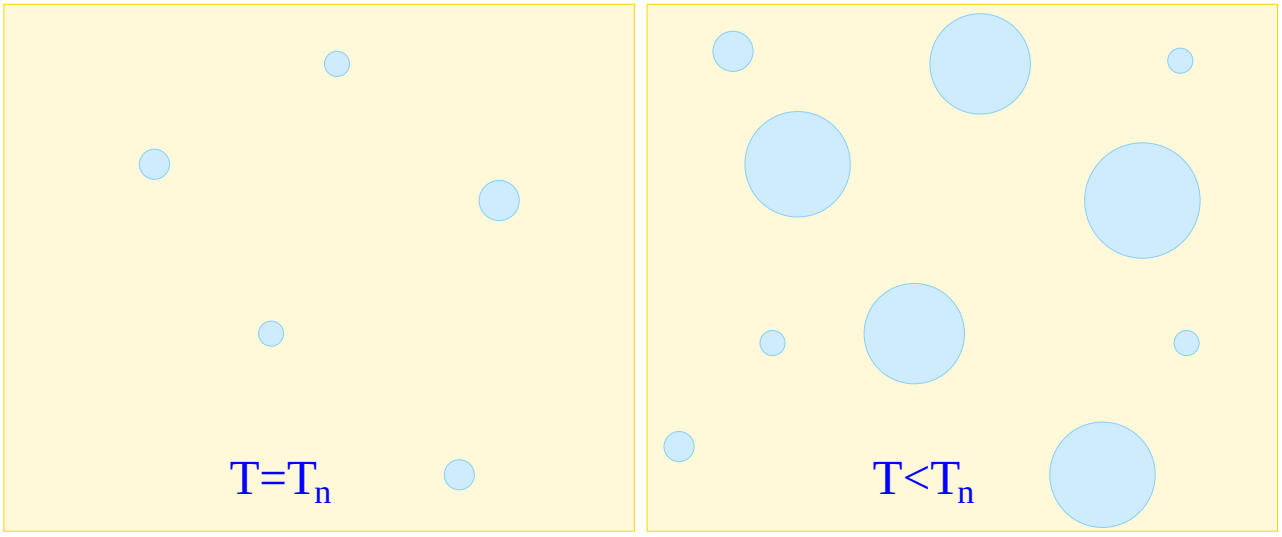


FIG. 49: Phase coexistence and bubble nucleation at a first order phase transition. In the left panel we show a snapshot of a supercooled plasma, whose temperature is close to the nucleation temperature, $T \simeq T_n$, such that bubbles of the true vacuum phase have just began nucleating. In the right panel we show a snapshot of the same phase transition at a later time, at which the temperature has further decreased, $T < T_n$. The bubbles, which were already formed in the left panel, have grown, and new ones have formed.

which the probability (per unit time) for the field to tunnel through the barrier to the true minimum becomes of the order the expansion rate of the Universe,

$$\frac{dP_{\rm tunnel}}{dt} \sim H \sim 10^{-16} k_B T_n/\hbar\,. \tag{215}$$

As a result of the tunneling, bubbles of the broken phase form at random locations. Since the free energy inside the bubbles is lower than in the surrounding plasma, the bubbles expand. The expansion is fuelled by latent heat, defined as the difference betweed the free energy of the two phases, which is released as the bubbles expand. This process is illustrated in figure 49. In the left panel we show an early stage of bubble nucleation at $T \simeq T_n$. In the right panel we show a snapshot of the transition at a later time, $T < T_n$. Note that bubbles nucleated in the left panel have grown in size, and in addition new ones have formed. The growth rate of bubbles is limited by the friction generated by the interactions between the Higgs field bubble interface and the excess of particles in the surrounding plasma (Moore, Prokopec, 1995).

Let us now consider the electroweak phase transition in the Minimal Standard Model (MSM). The relevant Lagrangian of the Higgs sector is of the form,

$$\mathcal{L}_{\rm Higgs} = \eta^{\mu\nu}(D_\mu H)^\dagger (D_\nu H) - V(H^\dagger H)\,, \tag{216}$$

with the Higgs potential,

$$V = \hbar^{-2}c^2 m_H^2 H^\dagger H + \lambda_H (H^\dagger H)^2\,, \tag{217}$$

where we approximated the background metric by the flat Minkowski metric. This is appropriate as long as we are considering phenomena that occur at much shorter time scales than the Hubble time, which is at the electroweak transition,

$$t_H = H^{-1} \sim 10^{16} \frac{\hbar}{k_B T} \,. \tag{218}$$

The Higgs field in (216) is a complex scalar doublet charged under $SU(2)_L$,

$$H = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \phi_0 + \varphi_3 + i\varphi_4 \end{pmatrix} , \tag{219}$$

where H^+ and H^0 denote an electrically charged and neutral complex scalar Higgs field, respectively, φ_1, φ_2, φ_3 and φ_4 are real scalar fields, whose expectation value vanishes, and ϕ_0 is the Higgs vacuum expectation value, which is at zero temperature, $\phi_0 = v_0$, $(\hbar c)^2 v_0^2 = -2m_H^2 c^4/\lambda_H \simeq (246 \text{ GeV})^2$.

The Higgs field couples minimally to the weak and hypercharge gauge bosons,

$$D_\mu H = \Big(\partial_\mu + ig\frac{\tau^a}{2} W_\mu^a + ig'\frac{Y}{2} B_\mu\Big) H \,, \tag{220}$$

where τ^a $(a = 1, 2, 3)$ denote the Pauli matrices, W_μ^a $(a = 1, 2, 3)$ are the $SU(2)_L$ gauge boson fields, B_μ is the hypercharge field, g is the $SU(2)_L$ coupling and $g'Y/2$ $(Y = 1)$ is the $U(1)_Y$ hypercharge coupling of the Higgs field.

The electroweak transition takes place at about $k_B T \simeq 100$ GeV. Above the transition, $k_B T > k_B T_c \sim 100$ GeV, the Higgs expectation value vanishes, such that the electroweak symmetry

$$G_{\rm SM} = SU(3)_c \times SU(2)_L \times U(1)_Y \tag{221}$$

is restored, and all Minimal Standard Model (MSM) particles are massless. Below the transition temperature the Higgs field acquires expectation value, such that the symmetry of the MSM is "broken"

$$G_{\rm SM} \to SU(3)_c \times U(1)_{EM} \,, \tag{222}$$

where $U(1)_{EM}$ is the Abelian symmetry of electromagnetism. Note that the vacuum manifold of the Higgs sector is equivalent to a three-sphere, $\mathcal{M} = G_{\rm SM}/[SU(3)_c \times U(1)_{EM}] \sim S^3$, such that, after the symmetry breaking, three out of four scalar Higgs fields in (219), say φ_1, φ_2 and φ_3, the want-to-be Goldstone bosons of the symmetry breaking, and thus correspond to massless excitations. This is however not the case, since the symmetry is local (gauged). Consequently, the fields φ_1, φ_2 and φ_3 are to the pseudo-Goldstone bosons, which are massive. Physically, they correspond to the massive longitudinal excitations of the $W^\pm$ and Z gauge bosons.

Stricly speaking, this picture of the transition is naïve, in the sense that thermal effects 'dress' particles in the plasma (including gauge bosons), giving them a thermal mass (see Michel Le Bellac, "Thermal Field Theory," Cambridge University Press, 1996).

As a consequence of the electroweak symmetry breaking, the gauge bosons, quarks and leptons of the Minimal Standard Model acquire a mass by the means of the Brout-Englert-Higgs (BEH) mechanism. The transition can be either a first order phase transition, as illustrated in figure 48, or a crossover, as illustrated in figure 49, the second order transition being the limiting case between the two.

According to the BEH mechanism, the two charged W-bosons,

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2) \tag{223}$$

and the Z-boson,

$$Z_\mu \;=\; \cos(\theta_W) W_\mu^3 - \sin(\theta_W) B_\mu \tag{224}$$

aquire a mass, while the photon field,

$$A_\mu \;=\; \sin(\theta_W) W_\mu^3 + \cos(\theta_W) B_\mu \,, \tag{225}$$

remains massless, and likewise gluons remain massless. Here θ_W denotes the Weinberg angle, which is defined as,

$$\sin(\theta_W) = \frac{g'}{\sqrt{g^2 + g'^2}} = 0.23117 \pm 0.00033 \,. \tag{226}$$

The (zero temperature) masses of the W- and Z- bosons are,

$$\begin{aligned} m_W \;&=\; \frac{1}{2} g v_0 \hbar/c = 80.412 \pm 0.043 \text{ GeV}/c^2 \\ m_Z \;&=\; \frac{1}{2}(g^2 + g'^2)^{1/2} v_0 \hbar/c = \frac{m_W}{\cos(\theta_W)} = 91.1887 \pm 0.0044 \text{ GeV}/c^2 \,, \end{aligned} \tag{227}$$

where $v_0 \simeq 246$ GeV$/(\hbar c)$. The masses of the quarks and leptons of the MSM are generated by the Yukawa terms, which have the generic form,

$$\mathcal{L}_{\text{Yukawa}} = -y H \bar{\psi}_R \psi_L + \text{h.c.} \tag{228}$$

such that the masses are,

$$m_f = y \frac{v_0}{\sqrt{2}} \hbar/c \,. \tag{229}$$

The heaviest fermion in the MSM is the top quark, with mass (`http://www-cdf.fnal.gov/physics/new/top/top.html`),

$$m_t = 174.5 \pm 2.3 \text{ GeV}/c^2 \,, \qquad (\texttt{CDF + D0 FERMILAB collaboration Mar 2006}) \tag{230}$$

which implies, $y_t \simeq 1$ (note that coincidentally, $246/\sqrt{2} = 173.95$). All other fermions are much lighter, and the corresponding Yukawa couplings are much smaller than one. Finally, the (zero temperature) Higgs mass is bounded from below,

$$m_H = \sqrt{2\lambda_H}\, v_0 \hbar/c \geq 114 \text{ GeV}/c^2 \qquad (\text{LEP bound}) . \tag{231}$$

During the last year of the LEP accelerator operation (located at the CERN in Geneva, Schwitzerland), an evidence emerged that the MSM Higgs mass may be, $m_H \simeq 114$ GeV/c^2. However, the evidence has too low a significance to be conclusive. The Higgs particle is expected to be discovered at the Large Hadron Collider (LHC) (proton-proton collider), currently under constribuction at CERN, and to become operational in 2007.

The precise nature of the electroweak transition depends on how many particles that couple (strongly) to the Higgs field are (at least moderately) relativistic, and that is not known with certainty. Apart from the uncertainty due to the unknown Higgs mass, the transition in the Minimal Standard Model (MSM) is however well understood. The transition in the Minimal Supersymmetric Standard Model, and in some other (supersymmetric) extensions of the MSM is moderately well understood.

The basic picture of the transition is as follows. In the MSM the transition is first order when the Higgs mass is smaller than, $m_H = 72 \pm 2$ GeV/c^2, and a crossover for larger values of the Higgs mass (Kajantie, Laine, Rummukainen, Shaposhnikov, 1996). At $m_H \approx 72$ GeV/c^2 there is a second order transition, which belongs to the universality class of the three-dimensional Ising model (Rummukainen, Tsypin, Kajantie, Laine, Shaposhnikov, 1998).

The nature of the phase transition in the Minimal Standard Model was studied by considering a two-loop effective thermal potential in the early 1990s (Arnold, Espinosa, 1992), and nonperturbatively by the technique of dimensional reduction in the mid 1990s (Kajantie, Laine, Rummukainen, Shaposhnikov, 1996). The perturbative techniques are useful only when the transition is strongly first order or moderately strongly first order. When the transiton is weakly first order or a crossover, perturbative techniques fail, because the relevant fields, which govern the transition, develop long range correlations, which cannot be described by the perturbatively constructed effective potentials. Dimensional reduction is a technique based on the integration of the Matsubara modes, which correspond to the thermal excitations with a nonvanishing energy in equilibrium systems. Mathematically speaking, Matsubara modes correspond to the $n \neq 0$ modes of the propagator (196), with the energies, $E_n = -i\hbar\omega_n$. The technique of dimensional reduction yields accurate results only when there are no bosonic excitations in the theory with a mass of the order the first Matsubara energy, $-i\hbar\omega_1/c^2 = 2\pi k_B T/c^2$.

The result of dimensional reduction of the (reduced) Minimal Standard Model is the tree-dimensional

effective theory,

$$S_T^{(3)}[\phi] = \int d^3x \Big(\frac{1}{2} \mathrm{Tr}[F_{ij}F_{ij}] + c^2\hbar^{-2} m_3^2 H^\dagger H + \lambda_3 (H^\dagger H)^2 \Big) \tag{232}$$

where $F_{ij}^a = \partial_i A_j^a - \partial_j A_i^a + g_3 \epsilon^{abc} A_i^b A_j^c$ $(a,b,c = 1,2,3;\, i,j, = 1,2,3)$, $F_{ij} = T^a F_{ij}^a$, $D_i H = (\partial_i - i g_3 T^a A_i^a) H$, $T^a = \tau^a/2$, τ^a are the Pauli matrices, and ϵ^{abc} is the antisymmetric symbol in 3 dimensions. The theory (232) can be solved numerically by lattice gauge techniques.

The meaning of the parameters in the action (232) is as follows,

$$\begin{aligned} g_3^2 &= g^2(\mu) k_B T + \texttt{loop corrections} \\ \lambda_3 &= \lambda(\mu) k_B T + \texttt{loop corrections} \\ m_3^2 &= -\nu^2(\mu) + \Big(\frac{1}{2}\lambda_3 + \frac{3}{16} g_3^2 + \frac{1}{16} {g'}_3^2 + \frac{1}{4} y_t^2 \Big) k_B T c^{-4} + \texttt{higher order corrections}\,. \end{aligned} \tag{233}$$

where $-\nu^2$ denotes the (negative) tree-level Higgs mass squared (calculated at the origin), $-\nu^2 = (\hbar/c)^2 d^2V/dH dH^\dagger|_{H=0}$. The dependence on the parameter μ in Eq. (233) indicates that the couplings and masses correspond to those of the renormalised theory at an energy scale given by the temperature, $\mu \sim k_B T$. The parameter $g = g(\mu)$ is the $SU(2)_L$ gauge coupling constant, and $\nu = \nu(\mu)$, $g' = g'(\mu)$, $y_t = y_t(\mu)$ are to leading order related to the physical parameters as follows,

$$\begin{aligned} \nu^2 &= \frac{m_H^2}{2} \\ g' &= g \tan(\theta_W) \\ y_t &= \frac{g}{\sqrt{2}} \frac{m_t}{m_W} \\ \lambda_H &= \frac{g^2}{8} \frac{m_H^2}{m_W^2}\,, \end{aligned} \tag{234}$$

where $m_H \geq 114$ GeV/c^2 denotes the Higgs mass, θ_W denotes the Weinberg angle, $m_W = 80.412 \pm 0.043$ GeV/c^2 is the W-boson mass, $m_Z = 91.1887 \pm 0.0044$ GeV/c^2 is the Z-boson mass, $m_t = 178 \pm 4$ GeV/c^2 is the top quark mass. Here we neglected the corrections induced by the renormalisation of the bare Higgs parameters (Kajantie, Laine, Rummukainen, Shaposhnikov, 1995).

The result of the analysis based on dimensional reduction is shown in figure 50, which represents the phase diagram of a simplified Minimal Standard Model, obtained by removing the hypercharge $U(1)_Y$ field, B_μ. The vertical and horizontal axes are xy_c and x respectively, where x and y are the following dimensionless quantities,

$$x = \frac{\lambda_3}{g_3^2}\,, \qquad y = \frac{m_3 (g_3)^2}{g_3^4}\,, \tag{235}$$

and y_c is the critical value for y. One can namely show that adding the B_μ field influences the electroweak transition only very weakly. The diagram shows that the transition is first order for $x < x_c = 0.0983 \pm 0.0015$, or equivalently when $m_H < 72 \pm 2$ GeV/c^2. For Higgs mass above 72 ± 2 GeV,

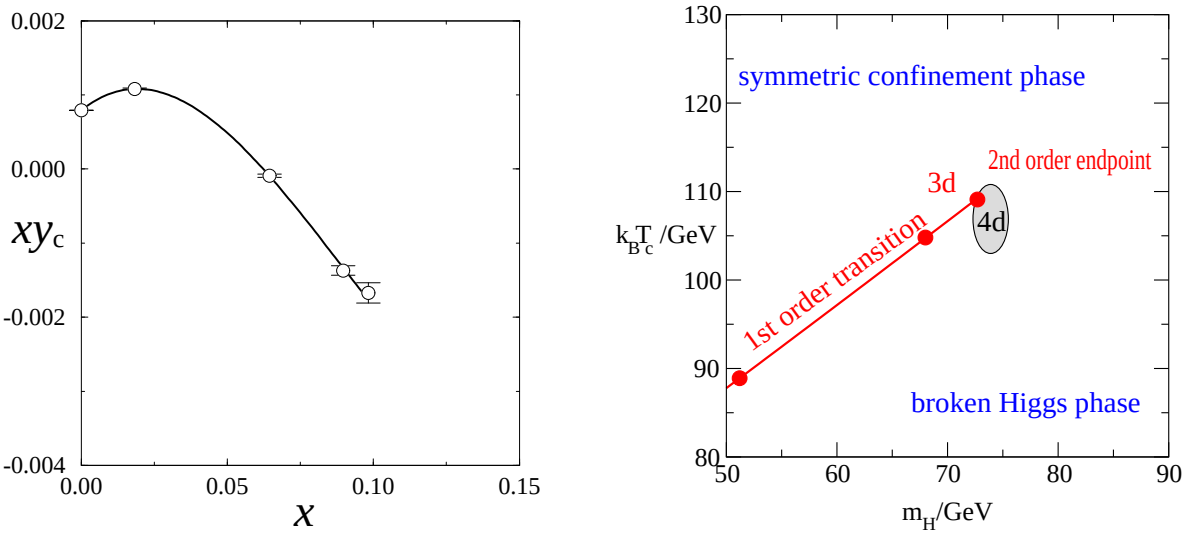


FIG. 50: *Left panel.* The phase diagram of the Minimal Standard Model, with the Abelian hypercharge symmetry $U(1)_Y$ neglected. The phase diagram is based on a dimensionally reduced $SU(2)_L$ theory with quarks and leptons. The horizontal axis represents the (dimensionless) Higgs mass, $x = \lambda_3/g_3^2$, and on the vertical axis a product of x and a critical value of the y parameter, $y = m_3(g_3)^2/g_3^4$, see Eq. (233). Here g_3 denotes the effective gauge coupling in the 3-dimensional effective dimensionally reduced theory (232), $m_3 = m_3(\mu)$ is the corresponding Higgs mass (calculated at a scale $\mu = g_3$) and λ_3 the corresponding quartic coupling (from Rummukainen, Tsypin, Kajantie, Laine, Shaposhnikov, 1998). The first order transition ends at the critical point, $x_c = 0.0983 \pm 0.0015$ ($m_H = 72 \pm 2$ GeV/c^2), which corresponds to a second order transition end-point, which belongs to the universality class of the three-dimensional Ising model.

Right panel. The same phase transition diagram but now with temperature on the vertical axis and the physical higgs mass $m_H c^2$ on the horizontal axis. The first order transition line (red solid line) ends at the second order endpoint: $m_H \simeq 72 \pm 2$ GeV/c^2, $k_B T_E \simeq 110$ GeV. For higher Higgs masses the transition is crossover. The three- and the four-dimensional numerical simulations agree remarkably well as regards the value of m_H but the agreement is less impressive as regards the critical temperature at the endpoint T_E.

the transition is a crossover. The first order transition ends at the critical point x_c, which corresponds to a second order transition end-point, belonging to the universality class of the three-dimensional Ising model. The critical temperature corresponding to the *endpoint* is $T_E \simeq 110$ GeV (see the right panel in figure 50). The nature of the phase transition in the Minimal Supersymmetric Standard Model is also well understood (Espinosa, Quiros, Zwirner, 1993; Laine, Rummukainen, 1998).

Understanding the nature of the electroweak transition is of relevance for cosmology for the following reason. It is quite plausible that during the electroweak transition a small matter-antimatter asymmetry

was generated, which would explain the excess of the matter over the antimatter of the Universe, manifest today as the baryonic matter (protons and neutrons) of the Universe. The mechanism can work only if the electroweak transition is strongly first order. Moreover, trere are indications that an excess of nonbaryonic matter over the corresponding antimatter was generated at or around the electroweak transition. If true, this would provide an elegant particle physics explanation for the origin of the dark matter of the Universe.

Even though the amount of dark and baryonic matter was initially tiny, a large mass of baryons ($m_p \sim m_n \sim 1$ GeV$/c^2$) and dark matter particles ($m \sim 10^2$ GeV$/c^2$) means that, as soon as the temperature falls below the relevant mass threshold, $k_B T < mc^2$, the energy density begins scaling as nonrelativistic matter. From that moment on, the relative fraction of the energy density in the baryonic and dark matter increases with the scale factor (and hence with time), $\rho_{\rm b}/\rho_{\rm r} \propto a$, $\rho_{\rm dm}/\rho_{\rm r} \propto a$, resulting eventually in matter domination after the matter-radiation equality at a redshift, $z_{\rm eq} = 3230 \pm 210$.

QCD phase transition

From numerical simulations (lattice gauge QCD) we know that the quantum-cromodynamics (QCD) transition in the early Universe is a crossover. At the transition the strong gauge coupling of the $SU(3)_c$ becomes large, quarks and gluons get confined into mesons (pions, rho-mesons, *etc.*) and baryons (protons and neutrons, *etc.*). The remaining symmetry gets 'broken' to the $U(1)_{EM}$ of electromagnetism,

$$SU(3)_c \times U(1)_{EM} \to U(1)_{EM} \, . \tag{236}$$

No observable remnants of the QCD transition in the early Universe are known, and hence the question of understanding the nature of the transition in the early Universe is primarily of academic interest. There is a large experimental effort undergoing at the Brookhaven National Laboratory (BNL) on Long Island (RHIC accelerator), with the purpose to create and study the state quark-gluon plasma and the QCD phase transition, and such to recreate the conditions existing in the early Universe. The experimenters smash gold nuclei on gold target, hoping that the gluon exchange interactions become so strong, that the resulting nuclei would heat up to a temperature above the QCD phase transition. The effort is to continue at the LHC at CERN.

A sketch of the QCD phase diagram in the (T, μ_B)-plane (temperature, baryon chemical potential) is shown in figure 51 for two light and one moderately heavy quark flavour ($m_u \sim m_d \ll m_s \ll T_c$). At high temperatures, $T > T_c$, the fundamental excitations are quarks and gluons. While the nonvanishing masses m_u and m_d break chiral symmetry, due to the smallness of the light quark masses, m_u and m_d, the symmetry is only approximate. In the limit when $m_u, m_d \to 0$ at low μ_B and below T_c the chiral

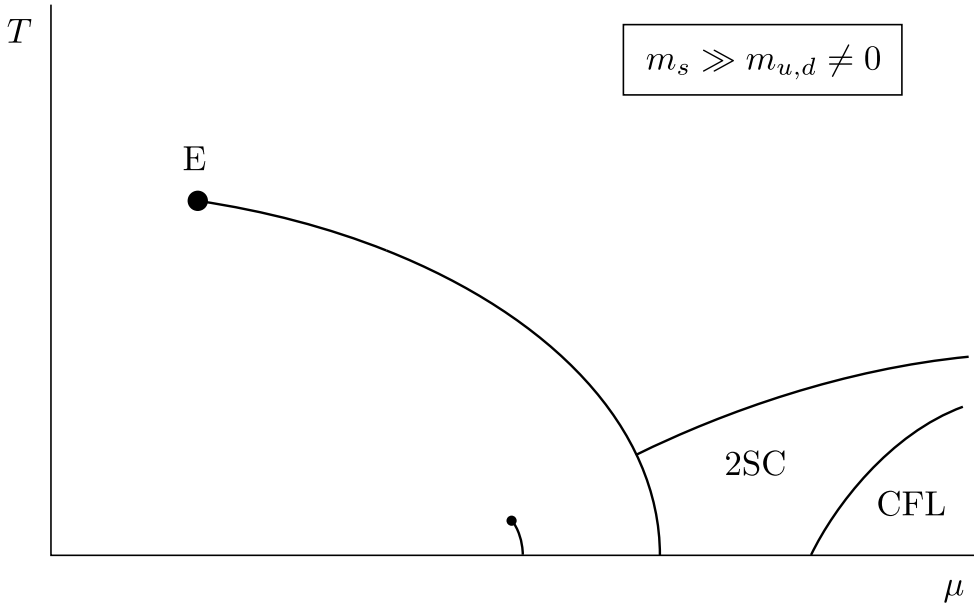


FIG. 51: The physical QCD phase space diagram (temperature *vs.* chemical potential) with three massive quarks, $m_u \sim m_d \ll m_s < T_c$ (not to scale). Other quarks (c, b, t) are too heavy to be thermally excited, and hence they are of no relevance for the transition. The relevant phases are: quark-gluon plasma (QGP) (high temperatures), hadronic phase (low temperature and low chemical potentials), nuclear matter, and the superconducting 2SC and CFL phases. Different phases are separated by the first order transition lines. For low chemical potentials, the transition between the QGP and hadronic phases is a crossover. The transition becomes second order at the end-point E, and first order above E.

condensate forms,

$$\langle \bar{\psi}^i_{L\alpha} \psi^{\alpha j}_R \rangle = M^{ij} = \sigma \delta^{ij} + \vec{\pi} \cdot (\vec{\tau})^{ij} \,, \tag{237}$$

such that the (approximate) chiral symmetry gets broken by the condensate,

$$SU(2)_L \times SU(2)_R \to SU(2)_{L+R} \,, \tag{238}$$

where $\vec{\tau} = (\tau^a)$ $(a = 1, 2, 3)$ are the Pauli matrices. The order parameter is $\phi = (\sigma, \vec{\pi})$, such that the chiral symmetry breaking (238) can be also viewed as, $O(4) \to O(3)$, with the vacuum manifold, $\mathcal{M} = O(4)/O(3) \sim S^3$, given by $\phi^2 = \sigma^2 + \sum_{a=1}^{3} (\pi^a)^2 =$ const. The chiral symmetry breaking is strictly speaking first order transition only in the unphysical limit of two massless quark flavours, $m_u = m_d = 0$ $(m_s \to \infty)$.

Apart from the chiral condensate (237), below the critical temperature quarks and gluons get confined into mesons and baryons (confinement). This low temperature phase is known as the hadronic phase, and we live in it.

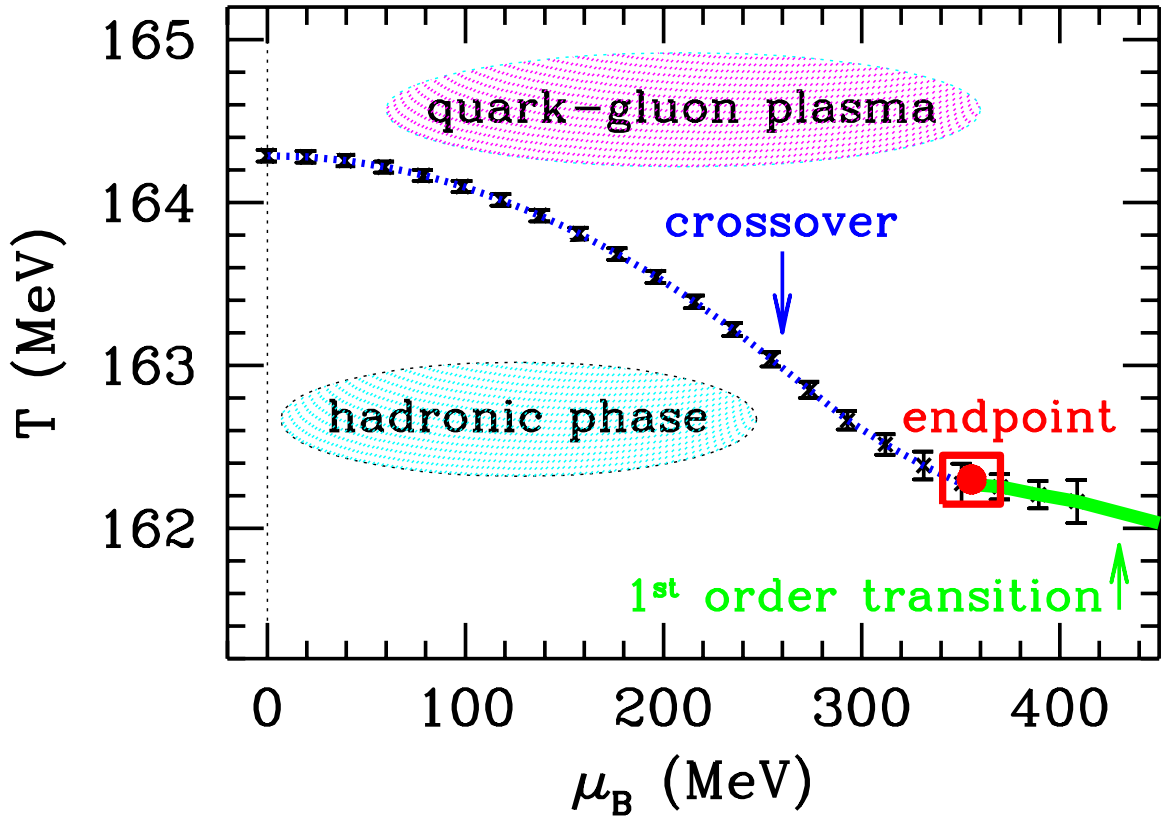


FIG. 52: The QCD phase space diagram for three massive quarks with the physical quark masses, $m_u = m_d = 0.0092\ k_B T_c/c^2$, and $m_s = 0.25\ k_B T_c/c^2$. The end point E lies at $k_B T_E = 162 \pm 2$ MeV, $\mu_E = 360 \pm 40$ MeV. The transition between the QGP and hadronic phases is a crossover for $\mu_B < \mu_E$ and first order for $\mu_B > \mu_E$ (Fodor, Katz, 2004).

The temperature of the transition is known to a relatively high precision,

$$T_c = T_{\rm QCD} = 164 \pm 2 \text{ MeV}\,. \tag{239}$$

The QGP-hadronic transition is a crossover up to the end point E, which is located at the temperature and baryonic chemical potential, as can be seen in figure 52,

$$\begin{aligned} \mu_E &= 380 \pm 40 \text{ MeV} \\ k_B T_E &= 162 \pm 2 \text{ MeV}\,. \end{aligned} \tag{240}$$

Above the end point the transition becomes first order. The transition line hits $T = 0$ at around $\mu_B \sim 1$ GeV (the precise value of the chemical potential is not known). At even higher chemical potentials and low temperatures, quark pairs develop a condensate, and QCD becomes *superconducting*. This was first proposed in the 1990s by Rajagopal and Wilczek by considering the Nambu-Jona-Lasinio (NJL) toy model of QCD, and demostrated by D. T. Son in 1998 by solving the one-loop dynamical 'gap' equation for QCD. The NJL model is a low-energy approximation of QCD, in which the (dynamical)

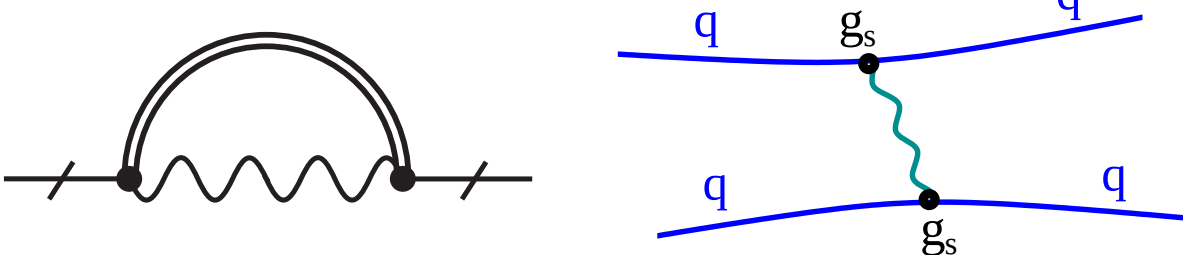


FIG. 53: The one-loop diagram used to calculate the quark self-energy (left panel). The diagram contains the relevant gluon exchange, which leads to an attractive interaction between quarks. Cutting the inner quark line in the diagram in the left panel (which amounts to forcing the loop quark propagator on-shell), results in the tree-level gluon exchange, shown in the right panel. When one sums the one-gluon exchange processes, one arrives at the color-flavor quark locking and the condensate as described in Eq. (241). The quark condensate is analogous to the Cooper electron pairs of conventional BCS superconductivity.

one-gluon exchange between two quarks is replaced by a contact (local) four-fermion interaction. Strictly speaking, the one-gluon exchange approximation applies in the limit, $\mu_B \to \infty$. Even though the gap equation is not (fully) gauge invariant, it is believed to yield trustable results.

As it can be seen from figure 51, there are two superconducting phases, which are acronymed as the 2SC and CFL phase, respectively. Recall that the conventional Bardeen-Schrieffer-Cooper (BCS) superconductivity is realised through the formation of Cooper pairs. Cooper pairs are the pairs of electrons with an opposite spin and momentum, bound (paired) by the exchange of phonons, which are vibrating deformations of charged atoms of the crystal lattice. Since a Cooper pair has a lower energy than the two freely moving electrons, an energy gap forms, such that superconducting (nondissipative) currents are established at low temperatures, below the energy gap, $k_B T < \Delta$. Analogous to the BCS ground state, the ground state of the color-flavor locking phase (CFL) of QCD corresponds to the condensate of quark pairs,

$$\langle \psi_{iL}^{a\alpha}(\vec{p},p^0)\psi_{jL}^{b\beta}(-\vec{p},p^0)\epsilon_{ab}\rangle = -\langle \psi_{iR}^{a\alpha}(\vec{p},p^0)\psi_{jR}^{b\beta}(-\vec{p},p^0)\epsilon_{ab}\rangle = \frac{\Delta_{\rm CFL}(\vec{p}^2,p^0)}{\hbar c}\epsilon^{\alpha\beta A}\epsilon_{ijA}\,, \tag{241}$$

where $\Delta_{\rm CFL}$ [GeV] is the energy gap, $\psi_{iL}^{a\alpha}$ and $\psi_{iR}^{a\alpha}$ denote the left- and right-handed two-spinors of quarks, respectively, ϵ_{ab}, $\epsilon_{\alpha\beta A}$, ϵ_{ijA} and the two- and three-indexed totally antisymmetric symbols, respectively, *e.g.* $\epsilon_{12}=1$, $\epsilon_{21}=-1$, $\epsilon_{11}=0$ and $\epsilon_{22}=0$, α,β denote the color indices, i,j denote the flavor indices, a,b are the spinor indices, and A is the index which links (locks) color and flavor. Note that the ground state is a scalar. Indeed, when the parity operation is applied on the left-handed condensate in (241), one gets a *minus* times the right-handed condensate, such that parity symmetry is respected by the vacuum state in (241).

Since the vacuum locks the $SU(3)_L$ to $SU(3)_c$, and likewise the $SU(3)_R$ is locked to $SU(3)_c$, in the

approximation of QCD plus three massless quarks, the vacuum breaks the symmetry as follows,

$$SU(3)_c \times SU(3)_R \times SU(3)_L \times U(1)_B \;\rightarrow\; G_{\rm CFL} = SU(3)_{c+L+R} \times Z_2\,, \tag{242}$$

such that the chiral symmetry is broken. Note that the the global baryon number symmetry, $U(1)_B$, is broken by the condensate to the discrete group, $Z_2 = \{1, -1\}$, the two elements corresponding to a positive and negative μ_B. Since the $SU(3)_c$ is fully broken by the CFL quark condensate (241), all of the eight gluons develop a mass,

$$m_G^2 = N_f \frac{g_s^2 \mu_B^2}{6\pi^2 c^4} + \Big(N_c + \frac{N_f}{2}\Big)\frac{g_s^2 (k_B T)^2}{9c^4}\,, \tag{243}$$

where $N_c = 3$ and $N_f = 3$ denote the number of colors and flavors, respectively. Consequently, all gluons exhibit the QCD analog of the Meisner effect known to occur for electromagnetic fields in the BCS superconductivity. Furthermore, all quarks that interact with the condensate also acquire a mass.

That a condensate of the form (241) indeed develops can be shown by considering the one-gluon exchange shown in the right panel in figure 53, which is obtained by 'cutting' the corresponding quark self-energy, shown in the left panel in figure 53. Recall that cutting a propagator corresponds to enforcing the loop quark propagator on-shell. The (one-loop) thermal quark self-energy can be thought of as a thermal nonlocal correction to the quark mass term. This gluon exchange is analogous to the one-photon exchange, which occurs as the radiative (loop) correction to propagation of charged particles. The gluon exchange generates an attractive force, resulting in an energy gap, which is at leading order in the coupling constant (Son, 1998; Pisarski, Rischke, 1999; Schäfer, Wilczek, 1999),

$$\Delta_{CFL} \sim b_{\rm CFL}\mu_B g_s^{-5} \exp\Big(-\frac{3\pi^3}{\sqrt{2}\,g_s}\Big)\,, \tag{244}$$

where $g_s = g_s(\mu_B)$ denotes the strong ($SU(3)_c$) coupling constant, μ_B is the chemical potential for baryons, and $b_{\rm CFL} \sim 512\pi^4 2^{-1/3}(2/3)^{5/2} \sim 10^4$ is a numerical constant.

The $2SC$ phase is a superconducting phase with two flavors of light quarks. (In this approximation the strange quark is too heavy to participate in the locking.) Cooper quark pairs form in the attractive $\bar{3}$-channel, and the relevant condensate is of the form,

$$\langle \epsilon_{\alpha\beta 3}\epsilon^{ij}\psi_i^\alpha(\vec{p}, p^0)\psi_j^\beta(-\vec{p}, p^0)\rangle = \frac{\Delta_{\rm 2SC}(\vec{p}^{\,2}, p^0)}{\hbar c}\,, \tag{245}$$

where $\Delta_{\rm 2SC}$ [GeV] is the 2SC energy gap. and the meaning of other symbols is identical as in (241). Note that the condensate is colored, *e.g.* blue, which is caused by the lack of a third light quark flavor to participate in pairing. The blue color of the condensate is compensated by the anti-blue color of the plasma, such that the whole system is color-neutral. The $SU(3)_c$ is only partially broken by the condensate (245),

$$SU(3)_c \;\rightarrow\; SU(2)_c \tag{246}$$

such that *five* out of eight gluons acquire a mass.

The gap for the 2SC phase is,

$$\Delta_{2SC} \sim b_{\rm 2SC}\mu_B g_s^{-5}\exp\Big(-\frac{3\pi^3}{\sqrt{2}\,g_s}\Big)\,, \tag{247}$$

where $b_{\rm 2SC} \simeq 512\pi^4 \simeq 3.5 b_{\rm CFL}$, such that the 2SC gap is bigger than the gap of the CFL phase (244), which is also indicated in figure 51.

Since the current baryon-to-photon ratio of the Universe is very small, $n_B/n_\gamma = 6.1 \pm 0.3 \times 10^{-10}$, we have good reasons to believe that the QCD transition in the early Universe occured at a small baryonic chemical potential, and hence it was a crossover. Low temperature and large density conditions are realised in the centers of neutron stars, such that they represent a candidate where the large μ_B and low temperature superconducting region of the QCD phase space diagram in figure 51 may be realised. So far there are no observations, which would indicate that a superconducting QCD state is realised in the interior of neutron stars.

Neutrino decoupling

The Universe is filled with cosmic background neutrinos of a similar number density as the cosmic microwave radiation, $n_\gamma = 410.4 \pm 0.9$ cm^{-3}, and yet the cosmic background neutrinos have never been observed, because they interact very weakly with ordinary matter.

At high temperatures, $k_B T \gg 1$ MeV, neutrinos are kept in equilibrium with the rest of the plasma (which at that epoch consisted of photons, electrons and positrons) by the weak interactions, examples being,

$$\begin{aligned} \bar\nu + \nu &\leftrightarrow e + e^+ \\ \bar\nu + e &\leftrightarrow \nu + e^+\,, \end{aligned} \tag{248}$$

etc.

On the other hand, the photo-dissociation of electron-positron pairs,

$$e^+ + e \leftrightarrow \gamma + \gamma\,, \tag{249}$$

is biased towards the *right* below the rest mass of the electron, $m_e = 0.511$ MeV$/c^2$, such that at temperatures much below $m_e c^2/k_B$, the number density of electrons and positrons is exponentially (Maxwell) suppressed, $n_e \sim n_{e^+} \propto {\rm e}^{-m_e c^2/(k_B T)}$.

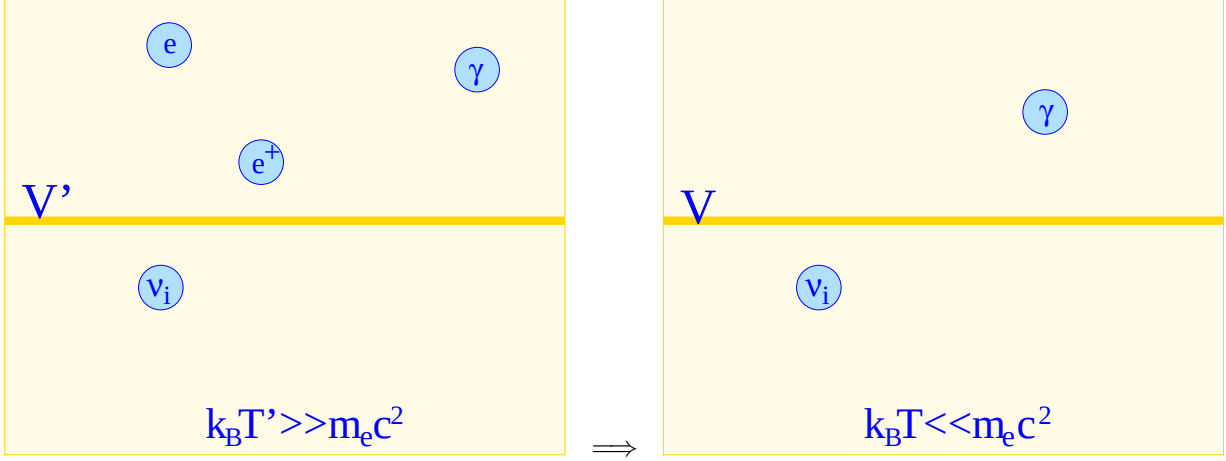


FIG. 54: The neutrino decoupling. In the left panel we show the Universe's contents above the temperature of the electron-positron photo-dissociation, $k_B T' \gg m_e c^2$, but much below any other particle mass scale of the MSM (except, of course, neutrinos). At these temperatures the Universe is to a good approximation a relativistic gas of photons, electrons, positrons and neutrinos. All species are at the same temperature, $T' = T'_\nu = T'_\gamma$. At a temperature below the electron rest mass, $T \ll m_e c^2/k_B$, most of the electrons and positrons have desintegrated, and the plasma consists mainly of photons and neutrinos. The resulting temperature of photons is higher than the temperature of neutrinos.

Since all neutrino species are light, $m_{\nu_i} \ll k_B T/c^2$, the neutrino energy and particle densities can be estimated by making use of the relativistic approximations (157) and (173), as follows,

$$\begin{aligned} \rho_\nu &= \frac{7}{4} N_\nu \frac{\pi^2}{30} \frac{(k_B T)^4}{(\hbar c)^3} \\ n_\nu &= \frac{3}{2} N_\nu \frac{\zeta(3)}{\pi^2} \Big(\frac{k_B T}{\hbar c}\Big)^3 , \end{aligned} \tag{250}$$

where we used, $g_{*N,\nu} = 2 \times (3/4) \times N_\nu$, $g_{*,\nu} = 2 \times (7/8) \times N_\nu$. Here $N_\nu \simeq N_f = 3$ denotes the number of (light) neutrino species, and $\zeta(3) \simeq 1.202$.

We shall now show that, assuming the neutrinos decouple at a temperature, $T' \gg T_{\nu,\rm dec} \gg m_e c^2/k_B$, the neutrino temperature drops below the photon temperature when $T_\gamma \ll m_e c^2/k_B$. The Universe content is shown in figure 54. In the left panel we show the Universe at a temperature, $T' \gg m_e c^2/k_B$. The Universe consists mainly of relativistic photons, electrons, positrons and neutrinos, all species being at the same temperature, $T' = T'_\nu = T'_\gamma$. As the temperature decreases neutrinos decouple. At temperatures below the electron rest mass $T \ll m_e c^2/k_B$ electrons and positrons desintegrate, thereby heating the photon fluid, but not the (decoupled) neutrino fluid.

To calculate the neutrino temperature we use the entropy conservation law (166)

$$\frac{d(Vs)}{dt} = 0 \,, \tag{251}$$

where $s = (\rho + \mathcal{P})/T$ denotes the entropy density (162–163), and $V \propto a^3$ the comoving volume. This then implies,

$$S = \frac{4}{3T'}\rho' V' = \frac{4}{3T}\rho V\,. \tag{252}$$

For photons, electrons and positrons we then have (157),

$$\begin{aligned} \rho' &= 2\Big(1 + 2\times\frac{7}{8}\Big)\frac{\pi^2}{30}\frac{(k_B T'_\gamma)^4}{(\hbar c)^3} \qquad (k_B T'_\gamma \gg m_e c^2) \\ \rho &= 2\frac{\pi^2}{30}\frac{(k_B T_\gamma)^4}{(\hbar c)^3} \qquad (k_B T_\gamma \ll m_e c^2)\,, \end{aligned} \tag{253}$$

and the energy density of neutrinos is given by Eq. (250). From this and the entropy conservation law (252), we then conclude,

$$\begin{aligned} \frac{11}{4}V' {T'_\gamma}^3 &= V T_\gamma^3 \qquad \text{(photons)} \\ V' {T'_\nu}^3 &= V T_\nu^3 \qquad \text{(neutrinos)}\,. \end{aligned} \tag{254}$$

Taking account of $T'_\gamma = T'_\nu$, this then implies,

$$T_\nu = \Big(\frac{4}{11}\Big)^{\frac{1}{3}} T_\gamma\,, \tag{255}$$

which is also valid today.

Since the photon temperature today is,

$$T_\gamma = T_0 = 2.725 \pm 0.001 \text{ K}\,, \tag{256}$$

the current temperature of the cosmic neutrino background is,

$$T_\nu \approx 1.95 \text{ K}\,. \tag{257}$$

This is a prediction of the Big Bang theory, which has not yet been tested. Due to the weakness of their interactions, most likely the cosmic background neutrinos will be observed only in a distant future.

From Eq. (257) and from the photon density, $n_\gamma \simeq 410 \text{ cm}^{-3}$, it is then quite easy to estimate the number density of neutrinos today (per relativistic species),

$$\frac{n_\nu}{N_\nu} = \frac{3}{4}\Big(\frac{T_\nu}{T_\gamma}\Big)^3 n_\gamma \simeq 112 \text{ cm}^{-3}\,. \tag{258}$$

Since we do not know the masses of neutrinos, we do not know the number of light (relativistic) species today.

In the derivation of the neutrino temperature (257) we have assumed that neutrinos decouple at a temperature, which is much higher than the temperature of the electron-positron photo-desintegration,

$T_{\nu,\rm dec} \gg m_e c^2/k_B$. In order to check the validity of this assumption, we note that neutrinos fall out of equilibrium when the time scale for weak interactions (248), becomes of the order the expansion time of the Universe,

$$t_{\nu,\rm dec} = t_H \equiv \frac{1}{H}\,. \tag{259}$$

The cross-section for weak interactions at low energies (with respect to the mass scale of the W- and Z-bosons, $m_W \sim m_Z \sim 10^2$ GeV$/c^2$) is given by

$$\sigma_\nu[\mathtt{m}^2] \sim G_F^2 (k_B T)^2\,, \tag{260}$$

where G_F denotes the Fermi constant,

$$G_F = 1.17 \times 10^{-5}\ ({\rm GeV})^{-2} (\hbar c)\,. \tag{261}$$

Recall now that

$$t_{\nu,\rm dec} \equiv \frac{1}{\Gamma_\nu} \simeq \frac{1}{n_\nu \sigma_\nu c}\,, \qquad H^2 = \frac{\hbar \rho}{3 c M_{\rm Pl}^2}\,, \tag{262}$$

which implies ($g_* = 43/4$)

$$t_{\nu,\rm dec} \sim \sqrt{\frac{360}{43\pi^2}} M_{\rm Pl} c^2 \frac{\hbar}{(k_B T)^2}\,, \tag{263}$$

and hence

$$k_B T_{\nu,\rm dec} \simeq 1.5\ {\rm MeV}/k_B\,, \tag{264}$$

such that the condition, $T_{\nu,\rm dec} \gg m_e c^2/k_B$ is moderately well satisfied. A more accurate analysis, based on a solution of the appropriate kinetic equations, shows that the approximation is actually better than the above simple analysis would suggest.

An important question is whether the primordial neutrinos can be the dark matter of the Universe. The short answer is no. The actual energy density in neutrinos is constrained by the WMAP one year data (Spergel et al, 2003) combined with the 2dF galaxy catalogue,

$$\Omega_\nu h^2 = \frac{\sum_{i=1}^3 m_{\nu_i} c^2}{94{\rm eV}} \leq 0.008\,, \tag{265}$$

where $h = 0.73 \pm 0.03$ is the Hubble parameter today (in units of 100 km/s/Mpc), such that the sum of neutrino masses is constrained as,

$$\sum_i m_{\nu_i} c^2 \leq 0.7\ {\rm eV}\,. \tag{266}$$

(Some recent studies (published in 2007) suggest an even tighter cosmological constraint on the neutrino masses: $\sum_i m_{\nu_i} c^2 \leq 0.2$ eV.) The constraint comes primarily from the observation that neutrinos do not cluster on small scales. Instead, they 'freely stream,' and prevent an early structure formation, as well as formation of structures on small scales, contrary to what has been observed.

The next important event in the Universe is the (primordial) Big Bang Nucleosynthesis (BBN), which takes place at about

$$k_B T_{bbn} \sim 0.1 \text{ MeV} \,, \tag{267}$$

which corresponds to a cosmic time, $t_{bbn} \sim 10^3$ sec. During the BBN primordial deuterium (D), helium (^3He and ^4He), and a small quantity of lithium (^7Li) are created by the nuclear interactions by which protons and neutrons get combined into light nuclei.

In order to perform a quantitative analysis of the BBN, we need to solve the kinetic Boltzmann equations for various nuclear interactions in an expanding Universe. We shall now devote some effort to present a derivation of the relevant Boltzmann kinetic equation, which has a broad range of applications.

Let us, for simplicity, begin with the Liouville equation, which is of the form

$$\frac{d}{dt} f \equiv (\partial_t + \vec{v} \cdot \partial_{\vec{x}} + \vec{F} \cdot \partial_{\vec{p}}) f = 0 \,, \tag{268}$$

where $f = f(\vec{x}, \vec{p}, t)$ denotes the distribution function, which measures the average density of particles on phase space, $\{\vec{x}, \vec{p}\}$, $\partial_t \equiv \partial/\partial t$, *etc.*, and hence the nature of (268) is statistical. The Liouville equation (268) applies to situations when interactions between particles can be neglected. Roughly speaking, $f = 1$, when there is one particle per unit volume of phase space, $\Delta x \Delta p_x \Delta y \Delta p_y \Delta z \Delta p_z = \hbar^3$. The velocity, $\vec{v}$ and the force, $\vec{F}$, in Eq. (268) are given by the Hamilton equations,

$$\begin{aligned} \vec{v} &= \dot{\vec{x}} = \partial_{\vec{p}} H \\ \vec{F} &= \dot{\vec{p}} = -\partial_{\vec{x}} H \,, \end{aligned} \tag{269}$$

where $H = H(\vec{x}, \vec{p})$ is the (classical) Hamiltonian.

The Liouville equation expresses the conservation of phase space densities, such that, as a system evolves in time, the regions of phase space where f does not vanish deform, but the total phase-space volume weighed by f (which represents the total number of particles in the system) is conserved by the evolution,

$$\int d^3x \, d^3p f = \text{const.} \tag{270}$$

While a general solution to the Liouville equation is hard to find, when there are conserved quantities in the system one can easily construct important classes of solutions. Perhaps the simplest case is when energy is conserved, implying that individual particles move along trajectories of a constant energy,

$$E = H(\vec{x}, \vec{p}) \tag{271}$$

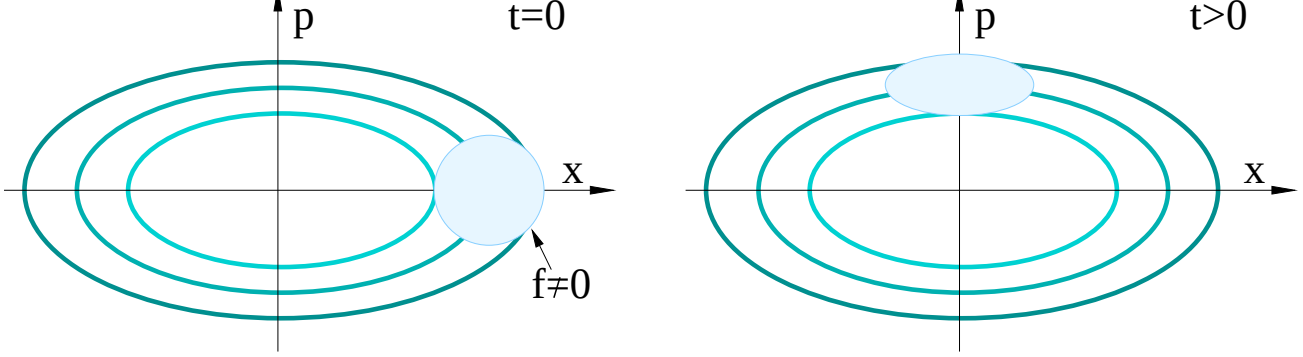


FIG. 55: The evolution of the distribution function f in phase space $\{x, p\}$ obeying the Liouville equation (268). The dynamics is dictated by the Hamiltonian of a one-dimensional harmonic oscillator (274), such that the individual particle trajectories are the ellipses of constant energy. An initial distribution function, $f(t=0)=1$ on the circle shown in the left panel evolves into an ellipse at $f(t>0)=1$ in the right panel. Since f is conserved by the evolution, the surface area of the ellipse in the right panel is equal to the surface area of the circle in the left panel, as dictated by the Liouville equation.

Since this Hamiltonian does not explicitly depend on time, the Liouville equation is solved by,

$$f = f(E)\,, \tag{272}$$

where $f = f(E)$ is any function of energy. Indeed, upon plugging (272) in Eq. (268), and making use of the Hamilton equations (269), results in,

$$\partial_t f(E) + \Big(\partial_{\vec{p}} H \cdot \partial_{\vec{x}} H - \partial_{\vec{x}} H \cdot \partial_{\vec{p}} H\Big) \frac{d}{dE} f(E) = \partial_t f(E) = 0\,, \tag{273}$$

which is indeed satisfied provided f does not explicitly depend on time. A simple way to understand this is to recall that f describes a collective motion of many individual particles, each of them moves along a trajectory given by (271), and where $\vec{x} = \vec{x}(t)$ and $\vec{p} = \vec{p}(t)$ describe classical trajectories, *i.e.* they are solutions of the Hamilton equations (269).

Perhaps the simplest example of this situation is a one-dimensional harmonic oscillator with the Hamiltonian,

$$H = \frac{p^2}{2m} + \frac{1}{2} m\omega^2 x^2 \tag{274}$$

such that

$$v \equiv \dot{x} = \frac{p}{m}\,, \qquad F = \dot{p} = -m\omega^2 x\,, \tag{275}$$

which are solved by ($x_0 = x(0)$ and $\dot{x}_0 = v(0)$ denote the initial position and velocity, respectively),

$$\begin{aligned} x &= x_0 \cos(\omega t) + \frac{\dot{x}_0}{\omega} \sin(\omega t) \\ p &= m\dot{x}_0 \cos(\omega t) - m x_0 \omega \sin(\omega t)\,, \end{aligned} \tag{276}$$

such that the trajectories of particles (of constant energy E) are ellipses in phase space shown in figure 55. In the left panel of figure 55 we show $f = f_0$ at an initial time $t = 0$ (circle), and in the right panel we show f at a later time, $t > 0$ (ellipse). Note that indeed one can think of the evolution of f as the evolution of a set of individual particles, each moving along a trajectory of constant energy. In this simple case the shape of the region in which f does not vanish deforms from a circle in the left panel into an ellipse in the right panel of figure 55, such that the surface area is preserved.

The Liouville equation (268) is not suitable for description of the early Universe dynamics, since it does not take account of particle interactions. In order to take account of interactions, we ought to make a suitable generalisation of the Liouville equation (268). This is the Boltzmann equation,

$$\frac{d}{dt}f = \mathrm{Coll}[f]\,, \tag{277}$$

where $\mathrm{Coll} = \mathrm{Coll}[f]$ denotes the collision functional (integral). This equation can in fact be derived from an effective action of a quantum field $\hat{\phi}$, which has a general form, $S_{\rm eff} = S_{\rm eff}[\phi, G]$, where ϕ and G denote the one point and two-point function (propagator). A variation of the effective action with respect to G results in a Schwinger-Dyson equation, to which the Boltzmann equation (277) is an approximate equation for the dynamics of the distribution function f, which is an on-shell projection of the two point (Whitman) function, $f = \int [dk_0/2\pi\hbar]G$.

When the distribution function f is integrated over the momenta, one obtains a particle number density,

$$n_a(\vec{x}, t) = \int \frac{d^3p}{(2\pi\hbar)^3} f_a(\vec{x}, \vec{p}_a, t) \tag{278}$$

where the index a refers to a species a. Since in an expanding universe the physical momentum scales inversely with the scale factor, $\vec{p} \propto 1/a$, integrating the left hand side of the Boltzmann equation (277) for $f \to f_1(\vec{x}, \vec{p}_1, t)$ results in the following continuity equation,

$$\begin{aligned}
\frac{1}{a^3}\frac{d}{dt}(a^3 n_1) + \nabla\cdot(\vec{u}_1 n_1) \;&=\; \int \frac{d^3p_1}{(2\pi\hbar)^3}\mathrm{Coll}[f_a] \\
&=\; -\prod_{a=1}^{4}\int \frac{d^3p_a}{(2\pi\hbar)^3 2E_a}(2\pi)^4\delta^3(\vec{p}_1+\vec{p}_2-\vec{p}_3-\vec{p}_4)\delta(E_1+E_2-E_3-E_4)|\mathcal{M}|^2 \\
&\qquad \times(\hbar^8 c^{12})\{f_1 f_2(1\pm f_3)(1\pm f_4) - f_3 f_4(1\pm f_1)(1\pm f_2)\}\,,
\end{aligned} \tag{279}$$

where $f_a \equiv f_a(\vec{x}, \vec{p}_a, t)$ are the distribution functions for species $a = 1, 2, 3, 4$, $E_a = (p_a^2c^2 + m_a^2c^4)^{1/2}$ is the corresponding energy, $\mathcal{M}$ is the scattering amplitude, $\vec{u}_a = \vec{u}_a(\vec{x}, t)$ denotes the fluid velocity of a species a, defined as,

$$\vec{u}_a = \frac{1}{n_a}\int \frac{d^3p_a}{(2\pi\hbar)^3}\frac{\vec{p}_a}{m_a} f_a \qquad (a = 1, 2, 3, 4)\,. \tag{280}$$

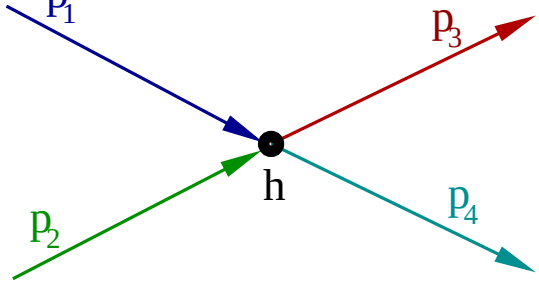


FIG. 56: An illustration of a two-to-two scattering process, whose full kinetic description is provided by the Boltzmann equation (279). Two incoming particles (1 and 2) interact and scatter into two outgoing particles (3 and 4). In this simple case, the interaction strength is given by the scattering amplitude, $\mathcal{M} = h$.

For the moment we assume that $\vec{u}_a = 0$, which is a reasonable assumption in a homogeneous expanding Universe, in which $a = a(t)$. In writing (279) we assumed a special form for the collision term, such that it describes two incoming particles, which interact, resulting in two outgoing particles,

$$1 + 2 \leftrightarrow 3 + 4\,. \tag{281}$$

The strength of the interaction is determined by the scattering amplitude, $\mathcal{M}$. Even though more complicated forms of the collision integral are possible, it is often the case that collisions are dominated by the two-to-two scatterings as in (279) and (281). The delta-functions in the collision integral (279) indicate that the energy and momentum are conserved by the interactions, in the sense that the sum of the incoming 4-momenta equals to the sum of the outgoing 4-momenta.

In the simplest case, the interaction is described by a local coupling constant, in which case the scattering amplitude is proportional to the coupling constant, as illustrated in figure 56. For example, the local quartic interaction for a real scalar field has the form, $\mathcal{L}_{\rm int} = -h\phi^4/4!$, which describes a two-to-two scattering, $\phi_1 + \phi_2 \leftrightarrow \phi_3 + \phi_4$, and whose corresponding scattering amplitude is simply,

$$\mathcal{M} = h\,. \tag{282}$$

More generally, one is interested in calculating the scattering amplitude of two-to-two scattering processes induced by one gauge boson exchange. In this case the calculation of the scattering amplitude is still quite simple, and we shall return to this question later.

The collision integral of the Boltzmann equation (279) describes the rate of change of the particle density due to the interactions. This rate depends not only on the interaction strength, but also on the actual occupancy of the incoming and outgoing states. More concretely, the scattering rate of the incoming bosons (fermions), $1 + 2 \to 3 + 4$, is statistically enhanced (suppressed) with respect to the vacuum interaction rate by (see Eq. (279)),

$$f_1 f_2 (1 \pm f_3)(1 \pm f_4) \tag{283}$$

where the upper (lower) sign refers to bosons (fermions). The scattering is enhanced if the outgoing states are highly populated bosons, $f_{3,4} \gg 1$, which is an expression of a stimulated bose emission. A dramatic example of this effect is the Bose-Einstein condensation, according to which a macroscopically large number of particles 'condenses' into a single quantum (ground) state. The Bose-Einstein condensation is currently being investigated in atomic systems. Since for fermions, $0 \leq (1-f_3)(1-f_4) \leq 1$, the scattering into occupied fermionc states is suppressed, which is an expression of the Pauli blocking phenomenon, well known from atomic physics (electron clouds keep atoms apart).

The scattering (decay) process, $1+2 \to 3+4$, is counteracted by the creation rate (note the minus sign in (279)), $3+4 \to 1+2$, whose rate is statistically enhanced (suppressed) as,

$$-f_3 f_4 (1 \pm f_1)(1 \pm f_2)\,. \tag{284}$$

The minus sign in front indicates that particles of species 1 are created by that process. In chemical equilibrium these two processes are balanced, such that,

$$\mathrm{Coll}[f] = 0\,, \tag{285}$$

implying that in chemical equilibrium, a special case of which is thermal equilibrium, the Liouville equation (268) is satisfied. Note that the collision integral (285) indeed vanishes for thermal equilibrium distribution functions,

$$f_{a,\mathrm{eq}} = \frac{1}{\mathrm{e}^{E_a/(k_B T)} \pm 1}\,, \tag{286}$$

which is a simple consequence of the energy conservation, $E_1 + E_2 = E_3 + E_4$, imposed by the energy δ-function in (279).

For now, we are primarily interested in simplifying the collision integral in (279). A careful look at (279), reveals that the continuity equation for n_a does not close (n_1 is given in terms of f_a, $a = 1,2,3,4$). A closure can be nevertheless achieved, if the fluid velocity is so small that it can be neglected, $\vec{u}_a \simeq 0$, and when the distribution functions can be approximated by the local thermal equilibrium form,

$$f_a = \frac{1}{\mathrm{e}^{\beta(E_a - \mu_a)} \pm 1}\,, \qquad E_a = \sqrt{m_a^2 c^4 + \vec{p}_a^{\,2} c^2} \qquad (a = 1,2,3,4)\,, \tag{287}$$

where $\beta = 1/(k_B T)$, and μ_a is the chemical potential of species a. This approximation is justified when the particle number changing interactions are weak in comparison to the particle number conserving interactions, an example of which is the 2-to-2 particle scattering (281), which clearly conserves particle number. An additional assumption is needed however. Namely, the system ought to be either near thermal equilibrium, or nonrelativistic. Here we analyse the simpler case, and assume that the system is nonrelativistic, in which case (287) simplifies to,

$$f_a \approx \exp\Big(-\frac{E_a - \mu_a}{k_B T}\Big)\,, \qquad E_a \simeq m_a c^2 + \frac{p_a^2}{2m_a} \gg k_B T \qquad (a = 1,2,3,4)\,, \tag{288}$$

for both bosons and fermions. In this nonrelativistic limit we have

$$\{f_1 f_2(1 \pm f_3)(1 \pm f_4) - f_3 f_4(1 \pm f_1)(1 \pm f_2)\} \simeq \mathrm{e}^{-(E_1+E_2)/(k_B T)}\Big(\mathrm{e}^{(\mu_1+\mu_2)/(k_B T)} - \mathrm{e}^{(\mu_3+\mu_4)/(k_B T)}\Big)\,, \quad (289)$$

such that we can reduce (279) to,

$$\frac{1}{a^3}\frac{d}{dt}\Big(a^3 n_1\Big) = -c n_{10} n_{20}\langle\sigma v\rangle\Big(\frac{n_1 n_2}{n_{10} n_{20}} - \frac{n_3 n_4}{n_{30} n_{40}}\Big) \quad (290)$$

where

$$\begin{aligned} n_{a0} &= g_a \int \frac{d^3 p_a}{(2\pi\hbar)^3}\mathrm{e}^{-E_a/(k_B T)} \\ &= g_a\Big(\frac{m_a k_B T}{2\pi\hbar^2}\Big)^{\frac{3}{2}}\mathrm{e}^{-m_a c^2/(k_B T)}\,, \qquad (m_a c^2 \gg k_B T)\,, \\ n_a &= n_{a0}\mathrm{e}^{\mu_a/(k_B T)}\,, \end{aligned} \quad (291)$$

and the thermally averaged cross section,

$$\langle\sigma v\rangle = \frac{\hbar^8 c^{12}}{n_{10} n_{20}}\prod_{a=1}^{4}\int \frac{d^3 p_a}{(2\pi\hbar)^3 2E_a}|\mathcal{M}|^2\delta^3(\vec{p}_1 + \vec{p}_2 - \vec{p}_3 - \vec{p}_4)\delta(E_1 + E_2 - E_3 - E_4)\mathrm{e}^{-(E_1+E_2)/(k_B T)}\,. \quad (292)$$

Analogous equations hold for other species, n_a ($a = 2, 3, 4$). Equation (290) is the desired continuity equation, which exhibits closure for particle densities, and we shall use it extensively in studying the kinetics of species in the expanding Universe. Since (290) is derived from the Boltzmann equation (279), it is often referred to as the Boltzmann equation.

In the limit of a large cross section (292), the system is driven towards a chemical equilibrium,

$$\frac{n_1 n_2}{n_{10} n_{20}} = \frac{n_3 n_4}{n_{30} n_{40}}\,, \quad (293)$$

which then implies,

$$\mu_1 + \mu_2 = \mu_3 + \mu_4\,. \quad (294)$$

While the chemical equilibrium alone does not guarantee that the individual chemical potentials are zero, it is often the case that many interactions are present in the plasma, such that the only self-consistent solution to the system of kinetic equations is the thermal equilibrium with all chemical potentials vanishing. The important exception are the chemical potentials associated with conserved quantities. An important example of a conserved quantity is the baryon number density $n_{\rm b}$. The quantity $a^3 n_{\rm b}$ is conserved below the electroweak phase transition, below which the Minimal Standard Model baryon violating processes are exponentially suppressed, and hence effectively inoperative.

Big Bang Nucleosynthesis

Big Bang Nucleosynthesis (BBN) refers to the creation of deuterium (D), helium (^3He and ^4He), and small quantities of lithium (^7Li) out of protons and neutrons, which were in turn formed in the early Universe at the QCD transition. The temperature at which this primordial nucleosynthesis takes place is,

$$k_B T_{bbn} \sim 0.1 \text{ MeV} \,. \tag{295}$$

which corresponds to a cosmic time of about, $t_{bbn} \sim 10^2$ sec.

Recall that at the temperature (295) neutrinos are already decoupled, $\Gamma(\nu e \leftrightarrow \nu e) \ll H$, and the number of electrons and positrons is Maxwell suppressed, $n_e \sim n_{e^+} \propto \mathrm{e}^{-m_e c^2/(k_B T)}$. The baryon-to-photon ratio has reached its current value,

$$\eta_b = \frac{n_b - n_{\bar{b}}}{n_\gamma} = 6.1 \pm 0.3 \times 10^{-10} \Big(\frac{\Omega_b h^2}{0.022} \Big) \,, \tag{296}$$

with $h = 0.73 \pm 0.03$. The asymmetry between baryons and antibaryons was created at the electroweak scale or at an even higher scale.

The weak interactions,

$$\begin{aligned} n + \nu &\leftrightarrow e + p \\ n + e^+ &\leftrightarrow \bar{\nu} + p \end{aligned} \tag{297}$$

are not any more in equilibrium. Due to the mass difference between the neutron and proton $Q = (m_n - m_p)c^2 = 1.293$ MeV, which is principally of electromagnetic origin, at $T \sim T_{bbn}$ both of these interactions are shifted towards the right hand side. On the other hand, around the same time neutrons begin decaying through the interaction,

$$n \rightarrow \bar{\nu} + p + e \,, \tag{298}$$

with an anomalously long lifetime (Serebrov et al., 2005),

$$\tau_n \simeq 880 \text{ s} \,. \tag{299}$$

The principal nuclear interactions leading to production of deuterium and helium are,

$$\begin{aligned} p + n &\leftrightarrow D + \gamma \\ D + D &\leftrightarrow {}^3He + n \\ {}^3He + D &\leftrightarrow {}^4He + p \,. \end{aligned} \tag{300}$$

Apart from small quantities of lithium, ^7Li, no heavy elements are produced in the early Universe. This can be understood from the binding energies per nucleon of the light elements shown in figure 57. No

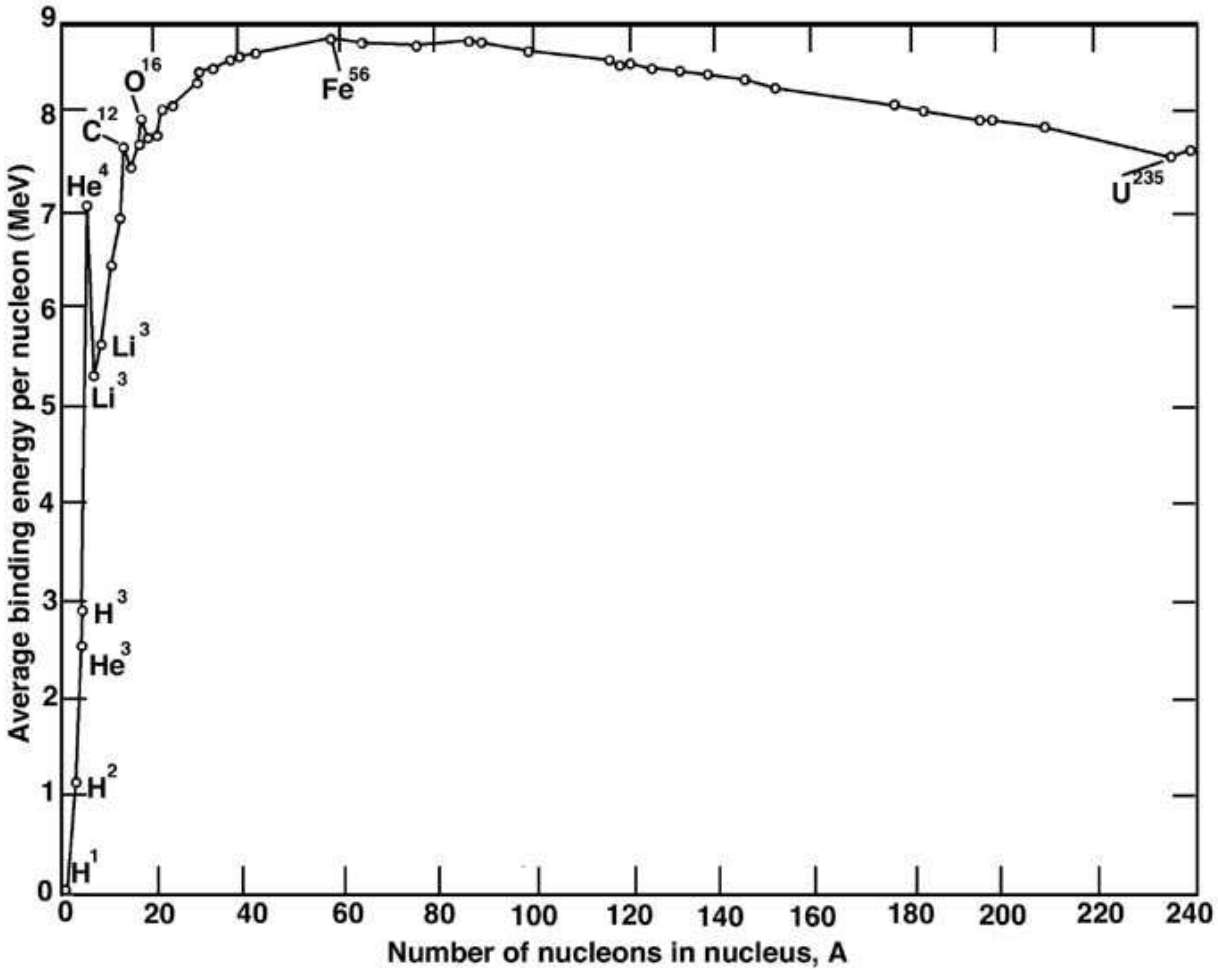


FIG. 57: The average binding energy per nucleon. Note that the binding energy of ^4He is anomalously large, such that it becomes larger again only for ^{12}C. Iron-56 ($^{56}_{22}Fe$) is the most stable isotope, with binding energy of almost 9 MeV per nucleon.

stable element with five nucleaons ($A = 5$) exists. Moreover, the binding energies per nucleon of lithium ($_3Li$), beryllium ($_4Be$) and boron ($_5B$) are all smaller than that of helium, making it energetically disfavourable to fuse helium atoms to produce Li, Be or B. The first element in the periodic table that is more stable than ^4He is carbon ^{12}C, followed by nitrogen ($^{14}_7N$), oxygen ($^{16}_8O$), *etc.* But in order to produce carbon, one ought to fuse three ^4He atoms, which is a very unlikely event in the primordial Universe. This process happens in stars though, producing thus heavier elements, which is of a crucial importance for the genesis of life.

Deuterium

Deuterium ($D = {}^2H$) is produced by the fusion of protons and neutrons,

$$p + n \;\leftrightarrow\; D + \gamma \tag{301}$$

When the process is in equilibrium, we may use the equilibrium form of the Boltzmann equation (290) to obtain,

$$\frac{n_D}{n_p n_n} = \frac{n_{D0}}{n_{p0} n_{n0}}\,, \tag{302}$$

where we took account of the fact that the photon density must be equal to that in equilibrium, $n_\gamma = n_{\gamma 0}$. This is so because the photon is its own antiparticle, and because the chemical potential of particles and antiparticles should add up to zero, implying $\mu_\gamma = 0$, and thus $n_\gamma = n_{\gamma 0}$. In nonrelativistic limit the energy of a particle species a can be expanded in a Taylor series as, $E_a \simeq m_a c^2 + p^2/2m_a$, and hence the equilibrium number densities are of the form (291), such that Eq. (302) can be rewritten as,

$$\frac{n_D}{n_p n_n} = \frac{3}{4}\Big(\frac{2\pi\hbar^2 m_D}{m_n m_p k_B T}\Big)^{3/2} \mathrm{e}^{B_D/(k_B T)}\,, \tag{303}$$

where

$$B_D = (m_n + m_p - m_D)c^2 = 2.22 \text{ MeV} \tag{304}$$

is the binding energy of deuterium. The factor 3/4 in Eq. (303) is the number of spin states of deuterium ($g_D = 3$) divided by the number of the spin states of the proton ($g_p = 2$) and by that of the neutron ($g_n = 2$). Now taking $m_D \simeq 2m_n \simeq 2m_p$ in the prefactor, and writing

$$n_p + n_n = n_b = \eta_b n_\gamma\,, \qquad n_p = X_p n_b\,, \qquad n_n = (1 - X_p) n_b \tag{305}$$

we get,

$$\frac{n_D}{n_b} = X_p(1 - X_p)\eta_b n_\gamma \frac{3}{4}\Big(\frac{4\pi\hbar^2}{m_p k_B T}\Big)^{3/2} \mathrm{e}^{B_D/(k_B T)}\,. \tag{306}$$

Upon taking for the density of photons,

$$n_\gamma = 2 \times \frac{\zeta(3)}{\pi^2}\Big(\frac{k_B T}{\hbar c}\Big)^3\,, \tag{307}$$

Eq. (306) can be recast as,

$$X_D \equiv \frac{n_D}{n_b} = \Big\{\frac{12\zeta(3) X_p(1 - X_p)}{\sqrt{\pi}}\Big\}\eta_b \Big(\frac{k_B T}{m_p c^2}\Big)^{3/2} \mathrm{e}^{B_D/(k_B T)}\,. \tag{308}$$

Noting that $X_p \in [0, 1]$, and $\max[X_p(1 - X_p)] = 1/4$ at $X_p = 1/2$, the prefactor in curly brackets is at most 2, it is reasonable to take it to be of the order one. From Eq. (308) it follows that it is primarily the smallness of η_b which inhibits production of deuterium until a fairly low temperature of about $k_B T \simeq 0.07$ MeV.

Neutrons

At high temperatures, $k_BT \gg 1$ MeV, the weak interactions (297) are efficient, implying that the equilibrium densities of neutrons and protons are almost equal, $n_{n0} \simeq n_{p0}$. When the temperature drops below about $k_BT \sim 1$ MeV, the interactions (297) are not any more so efficient. To study how the density of neutrons, protons and electrons change with temperature in more detail, consider Eq. (290) for the scattering processes (297),

$$\frac{1}{a^3}\frac{d}{dt}\Big(a^3 n_n\Big) = -\lambda_{np}\Big(n_n - n_p\frac{n_{n0}}{n_{p0}}\Big)\,, \tag{309}$$

where

$$\lambda_{np} = n_{l0}\langle\sigma v\rangle \tag{310}$$

denotes the scattering rate of neutrons into protons, and n_{l0} is the equilibrium density in leptons (neutrinos or electrons). Since in equilibrium,

$$\frac{n_{n0}}{n_{p0}} = \mathrm{e}^{-Q/(k_BT)}\,, \qquad Q = (m_n - m_p)c^2 = 1.293\ \text{MeV}\,, \tag{311}$$

and the number of baryons per comoving volume must be conserved,

$$\frac{d}{dt}(n_b a^3) = 0\,, \qquad n_b = n_n + n_p \tag{312}$$

we can recast Eq. (309) as,

$$\frac{dX_n}{dt} = -\lambda_{np}\Big(X_n - (1-X_n)\mathrm{e}^{-Q/(k_BT)}\Big)\,, \tag{313}$$

where we defined

$$X_n = \frac{n_n}{n_b}\,. \tag{314}$$

It is now useful to define a new (dimensionless) time variable,

$$\tau = \frac{Q}{k_BT}\,. \tag{315}$$

From the Friedmann equation,

$$\frac{1}{a}\frac{da}{dt} = -\frac{1}{T}\frac{dT}{dt} = H \tag{316}$$

it then follows that

$$\frac{d\tau}{dt} = -\tau H(\tau) \tag{317}$$

and Eq. (313) becomes,

$$\frac{dX_n}{d\tau} = -\frac{\tau\lambda_{np}(\tau)}{H(1)}\Big(X_n - (1-X_n)\mathrm{e}^{-\tau}\Big)\,, \tag{318}$$

with

$$H(1) = \Big(g_*\frac{4\pi^3 G_N}{45c^2}Q^4\Big)^{\frac{1}{2}} = 1.13\ \text{s}^{-1}\,, \qquad g_* = 10.75\,. \tag{319}$$

The scattering rate λ_{np} can be expressed in terms of the neutron decay time $\tau_n = 878.6 \pm 0.8$ s as follows (Bernstein 1988),

$$\lambda_{np} = \frac{255}{\tau_n \tau^5}(12 + 6\tau + \tau^2)\ [\mathtt{s}^{-1}]\,. \tag{320}$$

When one integrates Eq. (318), one finds that below $k_B T \simeq 0.5$ MeV, neutron abundance freezes at about, $X_n|_{\rm freeze} \simeq 0.15$. Below $k_B T = 0.1$ MeV, the reaction

$$n + p \ \leftrightarrow \ D + \gamma \tag{321}$$

starts depleting neutrons, producing deuterium and eventually helium. The final amount of helium produced can be estimated by assuming that all of the remaining neutrons have been converted into helium at $k_B T \simeq 0.07$ MeV, which is the energy scale at which the reaction (321) becomes effective. From the Friedmann equation one obtains for the age of the Universe,

$$t = 132\Big(\frac{0.1\ \text{MeV}}{k_B T}\Big)^2 [\mathtt{s}] \tag{322}$$

such that at that moment ($k_B T \simeq 0.07$ MeV), a fraction, $\mathrm{e}^{-t/\tau_n} \simeq 0.74$ have decayed through the neutron decay process (298). The final fraction of neutrons that survives is then, $0.15 \times 0.74 \simeq 0.11$. The helium abundance, which is defined as

$$Y_p = \frac{2n_n}{n_n + n_p} = 2X_n\,, \tag{323}$$

reads

$$Y_p \simeq 0.22\,. \tag{324}$$

This is perhaps surprisingly close the more accurate value, obtained by numerically solving the full set of equations of nuclear physics, and the result of which can be approximated by,

$$Y_p \approx 0.249 + 0.013 \ln\Big(\frac{\eta_b}{6 \times 10^{-10}}\Big) \tag{325}$$

Measurements of helium abundance in stellar systems show a mild dependence on the oxygen concentration. When extrapolated to the zero oxygen abundance, one obtains,

$$\begin{aligned} Y_p|_{\rm obs} &= 0.239 \pm 0.002 \qquad (\texttt{Luridiana et al. 2003},\ \texttt{Fields and Olive},\ \texttt{1998}) \\ Y_p|_{\rm obs} &= 0.245 \pm 0.002 \qquad (\texttt{Izotov et al. 1999})\,, \end{aligned} \tag{326}$$

which is marginally consistent with the primordial nucleosynthesis value (325), given that $\eta_b = 6.1 \pm 0.3 \times 10^{-10}$. The current situation concerning the BBN helium-4 abundance is illustrated in figure 58, where both observational bounds on helium-4 abundance quoted in (326) are marked. In order to

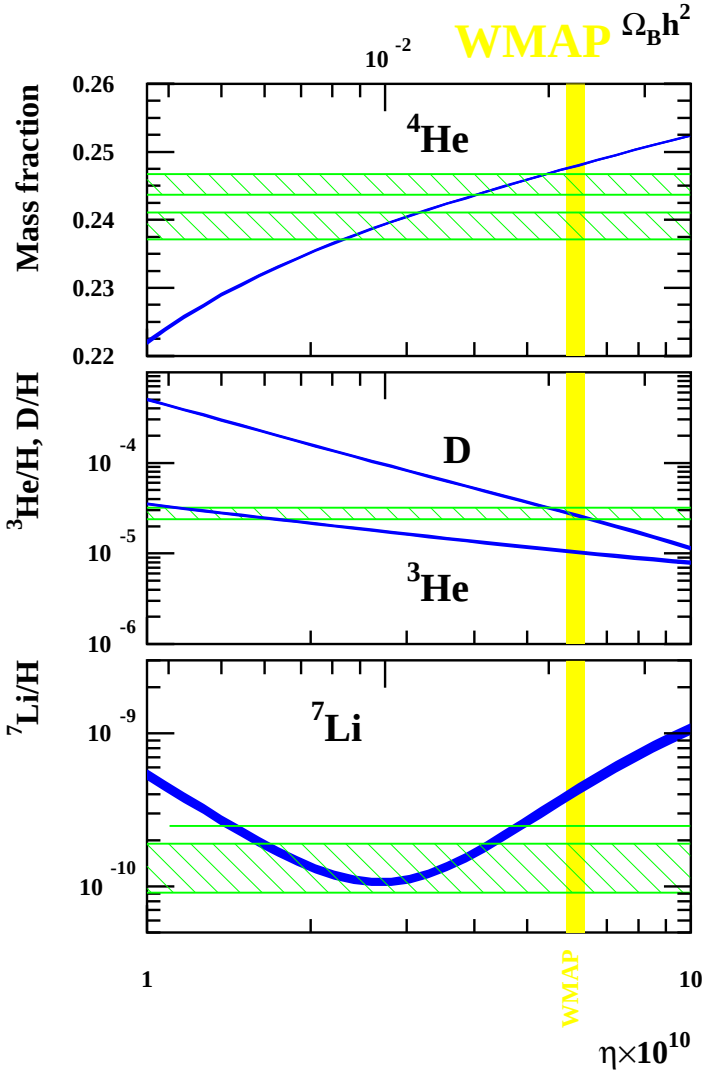


FIG. 58: Abundances of ^4He (mass fraction), D, ^3He and ^7Li (by number relative to ^1H) as a function of baryon-to-photon ratio, $\eta_b \times 10^{10} = (n_b/n_\gamma) \times 10^{10}$. The thickness of the lines corresponds to one standard deviation (68%) uncertainties. The hatched green areas represent observational uncertainties. We show two uncertainty bands for ^4He, the upper one from Izotov et al (1999) and the lower one from Luridiana et al. (2003). The observational unceratinty bands for D is from Kirkman et al. (2003) and for ^7Li from Ryan et al. (2000). The vertical yelow stripe corresponds to the one standard deviation band from Spergel et al. (WMAP 2003). The figure is taken from Coc et al. (2004).

remedy this rather unpleasant disagreement, Fields and Sarkar (in Particle Data Group, S. Eidelman et al., Phys. Lett. B 592, 1 (2004)) represent experimental bounds on the helium abundance with rather generous error bars, such that the WMAP constraint on η_b is about 2 standard deviations away from the ^4He abundance implied by the measurements.

This does not mean that the primordial nucleosynthesis is in a crisis, because the measured deuterium abundance agrees very well with the nucleosynthesis calculations and with the WMAP bound on η_b. Moreover, it is believed that systematic errors in the deuterium abundance measurements are better

under control then those of helium or lithium. The predicted abundance of deuterium is quite low because, as soon it is starts getting produced at $k_BT \sim 0.07$ MeV, deuterium is efficiently depleted by the reactions,

$$\begin{aligned} D + p &\leftrightarrow {}^3He + \gamma \\ D + D &\leftrightarrow {}^4He + \gamma \,. \end{aligned} \tag{327}$$

As can be seen from figure 58, according to the BBN calculations, for $\eta_b = 6.14 \pm 0.25 \times 10^{-10}$, deuterium freezes at,

$$n_D = 2.6 \pm 0.2 \times 10^{-5} n_H \,, \qquad (\texttt{Burles, Nollett, Turner 2001}) \,. \tag{328}$$

This is in perfect agreement with the baryon density

$$\eta_b|_{\text{Ly}-\alpha} = 5.9 \pm 0.5 \times 10^{-10} \tag{329}$$

or

$$\Omega_b h^2|_{\text{Ly}-\alpha} = 0.0214 \pm 0.0020 \,, \qquad , \tag{330}$$

implied by the measurement of the absorbtion lines of Lyman-α clouds

$$n_D = 2.8 \pm 0.5 \times 10^{-5} n_H \,, \qquad (\texttt{Kirkman, Tytler, Suzuki, O'Meara, Lubin, 2003}) \,. \tag{331}$$

The density of deuterium in Ly-α clouds can be inferred from the absorbtion lines in the radiation emitted by distant quasars. The absorbtion is due to the Ly-α transition in hydrogen and deuterium in Ly-α clouds at intermediate distances. Since Ly-α clouds are at cosmological distances, they have undergone very little evolution, and hence the density of deuterium in the clouds is very close to the primordial density produced during the BBN. The measurement discussed above is based on the absorbtion lines in the quasar Q1243+3047. The absorbtion takes place in a Ly-α cloud located at a redshift $z = 2.525659$. The absorbtion lines based on which the density of neutral hydrogen and neutral deuterium are inferred are around the wavelength, $\lambda \simeq 428.5$ nm. The analysed spectra are shown in figure 59.

Helium-3 abundance has not so far been measured with a satisfactory accuracy to be able to make a meaningful comparison with the BBN prediction, which can be inferred from figure 58,

$$n_{^3He} = 1.04 \pm 0.05 \times 10^{-5} n_H \,, \qquad (BBN) \,. \tag{332}$$

Just like ^4He abundance, the ^7Li-abundance does not fit very well the abundance predicted by the BBN,

$$n_{^7Li} \simeq 4 \pm 1 \times 10^{-10} \qquad (\texttt{BBN}) \,, \tag{333}$$

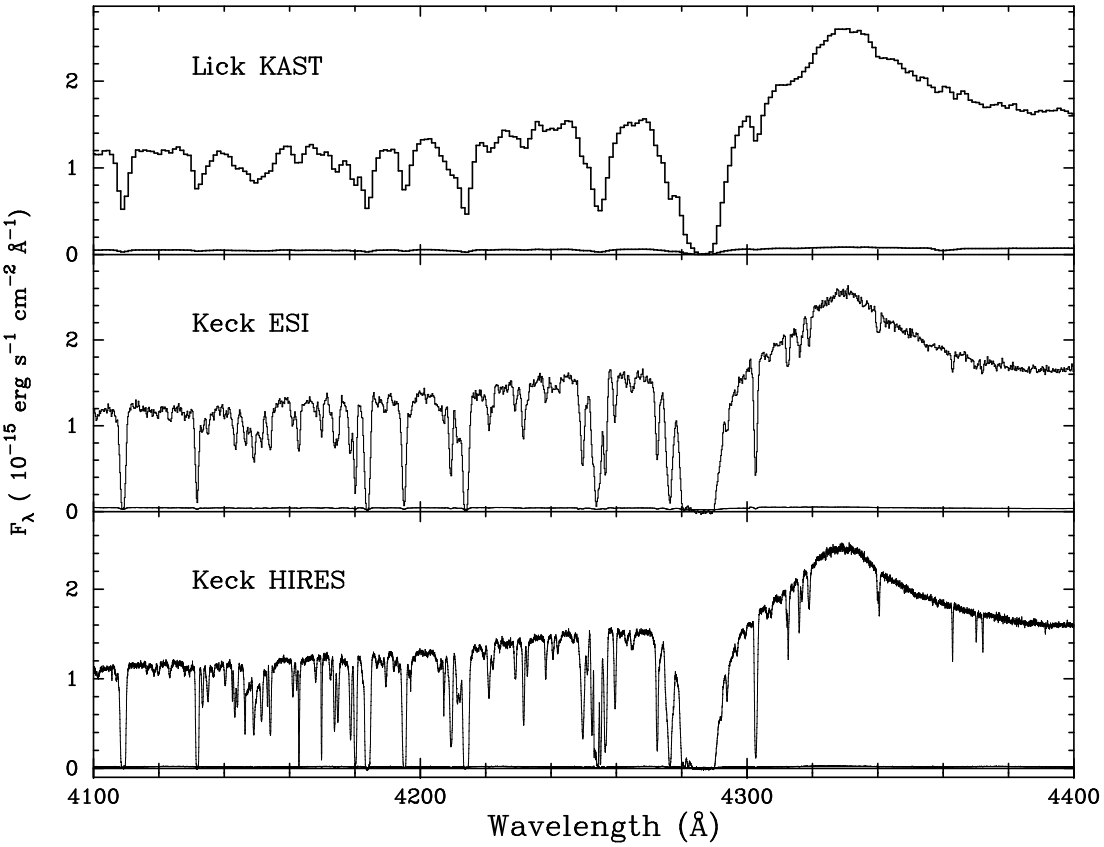


FIG. 59: The three absorbtion spectra of the quasar Q1243+3047 recorded by the telescopes Lick and Keck, based on which deuterium density is determined. The Ly-α absorbtion is located near $\lambda = 428.5$ nm. From Kirkman, Tytler, Suzuki, O'Meara and Lubin (2003).

which is to be compared with the observed balue,

$$n_{^7Li} = 1.2 \pm 0.7 \times 10^{-10} \qquad (\texttt{Ryan, Beers, Olive, Fields, Norris, 2000}) \tag{334}$$

This is usually not taken to be a serious problem for the following reasons. Firstly, systematic errors in the measurement of ^7Li are not very well understood. Secondly, there are no reliable data on some of the relevant interaction rates. In particular, the rate of the reactions, $^4He + {}^4He \leftrightarrow {}^6_3Li + D + \gamma$ and ${}^4_2He + {}^6_3Li \leftrightarrow {}^8_4Be + D + \gamma$, are important for the BBN prediction, but they have not been accurately measured.

Coincidences

Even though the strengths of the fundamental forces in nature are very different, all of the forces participate in primordial nucleosynthesis, and changing the strength of any interaction would change the primordial abundance of helium, deuterium and lithium. This curious fact lead some physicists (most notably Carter and Barrow) to propose the anthropic principle, according to which fundamental

constants in nature have such values that we can exist. One often resorts to philosophical arguments in order to defend or refute the anthropic principle. Here we content ourselves by simply stating curious coincidences which took place during primordial nucleosynthesis.

The amount of helium, deuterium and lithium produced at the BBN depends on:

- life time of neutron, $\tau_n = 878.6 \pm 0.8$ s, which is primarily determined by the *weak interaction* strength;
- the age of the Universe, $t \simeq 132(0.1 \text{ MeV}/(k_B T))^2 \mathtt{s} \propto G_N^{-1}$, which is determined by the strength of the *gravitational interaction*;
- deuterium binding energy, $B_D = 2.22$ MeV, determined by the strength of the *strong interaction*;
- the neutron-proton mass difference, $Q = 1.293$ MeV, which is primarily determined by the *electromagnetic interaction* strength.

Thus, all basic forces in Nature participate in the Big Bang Nucleosynthesis.

For example, increasing the strength of the *strong interaction*, would have as a consequence an increase in the binding energy of deuterium, which in turn would result in a more efficient generation of deuterium and thus also helium.

Next, increasing the strength of the weak interaction would result in a shorter neutron decay time, leading to a lesser deuterium and helium production.

A consequence of an increased strength of the electromagnetic interaction would be an increase in the mass difference between neutron and proton, increasing thus $Q = (m_n - m_p)c^2$, which in turn would lead to a reduced concentration of neutrons, and thus a lesser production of deuterium and helium.

Furthermore, increasing the strength of the gravitational interaction would result in a shorter age of the Universe, and thus less neutrons would decay by the same temperature, which would have as a consequence more deuterium synthesis and more helium production.

Finally, baryon density (and also lepton density) affects primordial nucleosynthesis. Indeed, Eq. (325) implies that, increasing $\eta_b = n_b/n_\gamma$, has as a consequence an increased production of helium. This can be understood as follows. Increasing η_b, increases equilibrium concentration of deuterium (η_b acts in the same sense as B_D), which in turn means a more efficient production of deuterium, and hence also helium. Moreover, increasing the number of relativistic degrees of freedom at nucleosynthesis decreases the age of the Universe, and hence has the same effect as increasing the strength of the gravitational interaction.

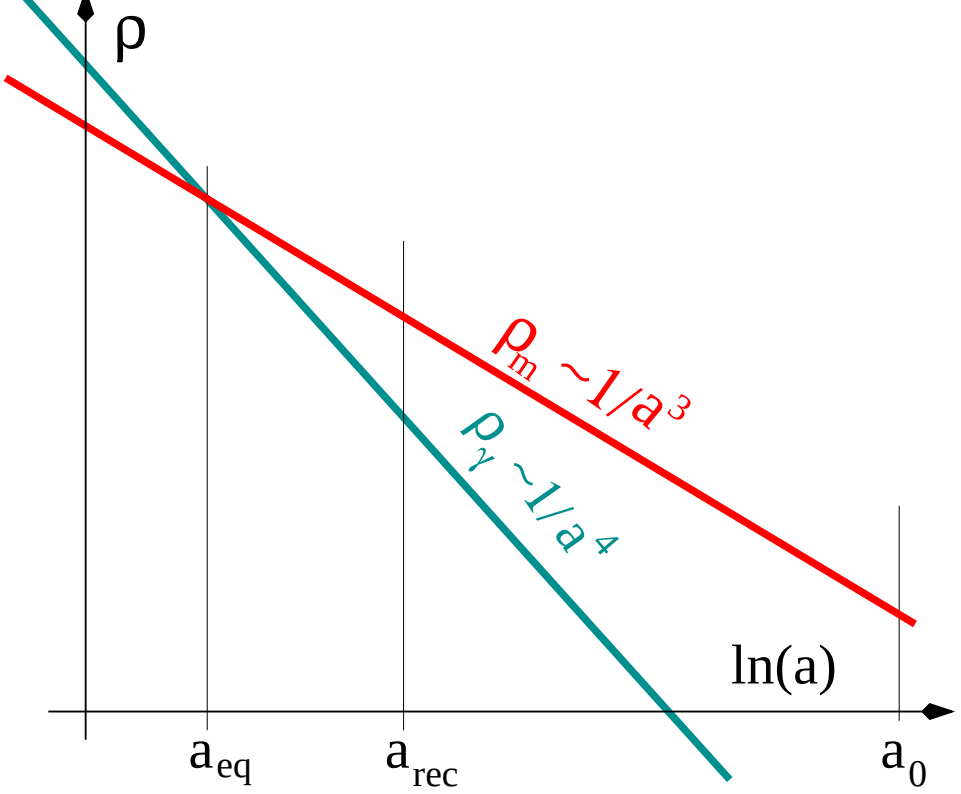


FIG. 60: The scaling of the radiation and matter energy densities around the epoch of equality, $z_{\rm eq} \sim 3230$ and decoupling at $z_{\rm dec} \simeq 1089$.

3. *Matter-radiation equality*

At a redshift,

$$z_{\rm eq} = \frac{a_0}{a_{\rm eq}} - 1 = 3230 \pm 210 \tag{335}$$

the energy density in matter and radiation is equal. The matter content of the Universe is composed of baryonic matter, ρ_b, and dark matter, $\rho_{\rm dm}$, $\rho_m = \rho_b + \rho_{dm}$. The present densities are,

$$\begin{aligned} \rho_b &= (0.042 \pm 0.003)\rho_{\rm cr} \\ \rho_m &= (0.24 \pm 0.03)\rho_{\rm cr} \end{aligned} \tag{336}$$

where $\rho_{\rm cr} = 3c^2H_0^2/(8\pi G_N)$ denotes the critical energy density. Eq. (336) implies that the amount of dark matter is about 5 times larger than the amount of baryonic matter,

$$\rho_{dm} \simeq 5\rho_b\,. \tag{337}$$

This is a tantalising hint that baryonic and dark matter have a related origin, or at least that they were created at the same epoch in the early Universe. The creation mechanisms are not known with certainty however.

The scaling of the energy densities in matter and radiation is shown in figure 60. Note that the epoch of equality, $z_{\rm eq} \sim 3230$, is very close to the epoch of electron-proton recombination, $z_{\rm rec} \simeq 1089$.

FIG. 61: A Compton scattering of a photon off an electron. A photon gets absorbed, and then reemitted at a different place. The strenght of the interaction is given by the fine structure constant, $\alpha_{\rm em} \simeq 1/137$. If the emitted photon has a larger energy than the absorbed photon, the process is known as the inverse Compton scattering.

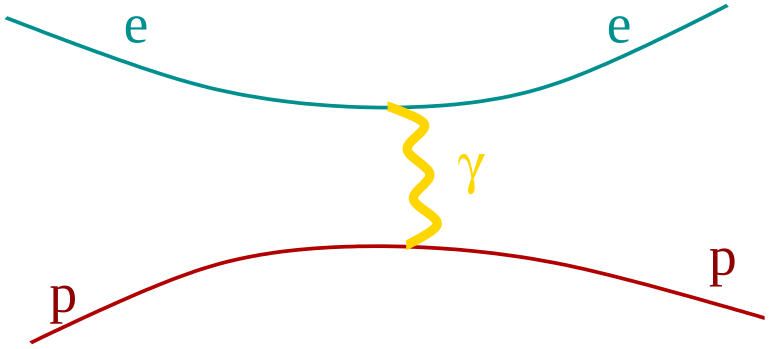


FIG. 62: Coulomb scattering of an electron off a proton *via* photon exchange. Through this process between equality at $z \sim 3230$ and decoupling at $z \simeq 1089$ protons remain tightly coupled to electrons. Just like Compton scattering, this interaction is of an electromagnetic strength.

During the short period between equality and recombination, the Universe is matter dominated, dark matter evolves freely, falling into the potential wells generated during primordial inflation, while baryonic matter remains tightly coupled to electrons and photons. It is only after recombination, when neutral hydrogen forms, and photons begin free-streaming, that baryons begin falling into the potential wells formed by dark matter, and the epoch of structure formation begins.

4. Decoupling and recombination

Up to the redshift,

$$z_{\rm dec} = 1089 \pm 1 \tag{338}$$

which corresponds to a temperature, $T_{\rm dec} \simeq 2970$ K, or $k_B T_{\rm dec} \simeq 0.256$ eV (recall that $k_B \simeq 8.6173 \times 10^{-5}$ eV/K), photons remain tightly coupled to electrons *via* Compton scattering, shown in figure 61, and electrons are tightly coupled to protons *via* Coulomb scattering, shown in figure 62. Below the temperature of decoupling, the production of neutral hydrogen, which is realised through the reaction,

$$p + e \leftrightarrow {}^1H + \gamma \qquad (B_H = 13.6 \text{ eV}) \tag{339}$$

devoids the Universe of free charges. Consequently, the Universe becomes electrically neutral and *transparent* to light, such that photons begin to free-stream. As we will see in a moment, electric neutrality is only approximate however.

Equilibrium consideration: Saha equation

Assuming that the reaction (339) is in chemical equilibrium, equation (293) holds, such that we have,

$$\frac{n_e n_p}{n_H} = \frac{n_{e0} n_{p0}}{n_{H0}}\,, \tag{340}$$

where we made use of $n_\gamma = n_{\gamma 0}$. Postulating electric neutrality of the Universe,

$$\frac{Q}{V} = n_p - n_e = 0 \tag{341}$$

results in

$$\frac{X_e^2}{1-X_e} = \frac{1}{n_b}\Big(\frac{m_e k_B T}{2\pi\hbar^2}\Big)^{3/2} \mathrm{e}^{-B_H/(k_B T)}\,, \qquad B_H = (m_e + m_p - m_H)c^2 = 13.6 \text{ eV}\,, \tag{342}$$

where we used $m_p \simeq m_H$ in the pre-factor, and we defined

$$X_e = \frac{n_e}{n_e + n_H} = \frac{n_p}{n_b} \tag{343}$$

with $n_b \simeq n_p + n_H$ (here we have for simplicity neglected helium) and we made use of,

$$n_{a0} = \Big(\frac{m_a}{2\pi\hbar^2}\Big)^{3/2} \mathrm{e}^{-m_a c^2/(k_B T)}\,. \tag{344}$$

Now taking account of $n_b = \eta_b n_\gamma$, $n_\gamma = [2\zeta(3)/\pi^2](k_B T/\hbar c)^3$, and $\eta_b = 6.1 \pm 0.3 \times 10^{-10}$, we obtain the following Saha equation,

$$\frac{X_e^2}{1-X_e} = \frac{1}{\eta_b}\frac{\pi^{1/2}}{2^{5/2}\zeta(3)}\Big(\frac{m_e c^2}{k_B T}\Big)^{3/2} \mathrm{e}^{-B_H/(k_B T)}\,, \tag{345}$$

By making use of the Friedmann equation, the temperature can be easily converted into the redshift z. When expressed in terms of redshift, one gets a curve shown in figure 63, which shows the electron concentration, $X_e = n_e/n_B$, as a function of the redshift, z. According to the Saha equilibrium, the concentration of electrons falls exponentially to zero as the redshift decreases. In reality, this does not happen, because at $X_e \sim 10^{-2}$, the Compton scattering rate drops below the expansion rate of the Universe, such that it is not fast enough to enforce chemical equilibrium. As a consequence, the density of electrons freezes at,

$$X_e|_{\text{freeze}} \sim 10^{-3}\,. \tag{346}$$

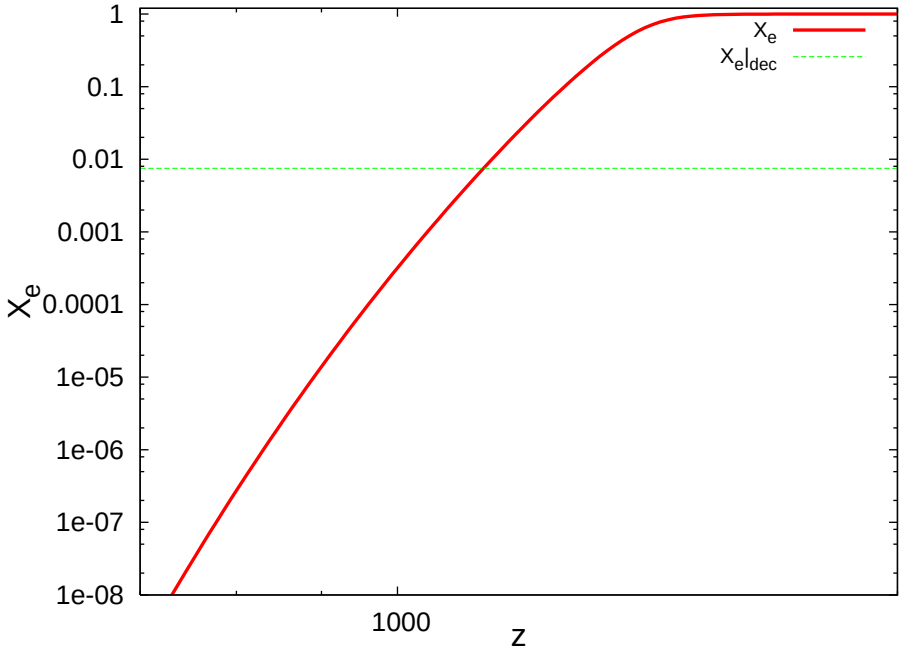


FIG. 63: The solution of the Saha equation (345) for the electron density, $X_e = n_e/n_b$ *vs.* reshift z (log-log plot). According to the Saha solution, the decoupling occurs at $z \simeq 1125$, which corresponds to $X_e = 0.0065$, below which the Compoton scattering rate falls below the expansion rate of the Universe (see below). This value is in fact somewhat higher than the reshift at decoupling, $z_{\rm dec} = 1089 \pm 1$, which is obtained by solving the full kinetic equation (347).

This represents the residual ionisation of the Universe, which survives up to today. One can obtain this result by applying the full kinetic equation (290) for the process (339), resulting in (see Problem 2.15),

$$\frac{dX_e}{dt} = -c\langle\sigma v\rangle\Big\{\eta_b n_\gamma X_e^2 - (1-X_e)\Big(\frac{m_e k_B T}{2\pi\hbar^2}\Big)^{3/2} {\rm e}^{-B_H/(k_B T)}\Big\}[{\rm s}^{-1}] \tag{347}$$

where the scattering cross section is of the form,

$$\langle\sigma v\rangle = 9.78\frac{\alpha_{\rm em}^2\hbar^2}{m_e^2 c^2}\Big(\frac{B_H}{k_B T}\Big)^{1/2} \ln\Big(\frac{B_H}{k_B T}\Big)[{\rm cm}^2]\,. \tag{348}$$

Here $\alpha_{\rm em} = 1/137$ is the fine structure constant. This equation can be further simplified by introducing a dimensionless time variable, $\tau = B_H/(k_B T)$, analogously as it was done in section C 2 in Eqs. (315–318).

Photon decoupling

Photons decouple when the rate to Compton scatter falls below the expansion rate of the Universe,

$$\Gamma_{\rm Compton} < H\,. \tag{349}$$

Taking account of the value of the Hubble parameter today, $\hbar H_0 = 2.1332h \times 10^{-33}$ eV, one easily finds,

$H_0 = 3.24h \times 10^{-18}$ s^{-1} $\simeq 2.30 \times 10^{-18}$ s^{-1}. On the other hand, the Compton scattering rate reads,

$$\Gamma_{\rm Compton} = cn_e\sigma_T = c\eta_b n_\gamma \sigma_T X_e \tag{350}$$

where $n_\gamma = 412(1+z)^3$ cm^{-3} denotes the density of photons as a function of the redshift, and

$$\sigma_T = 0.665 \times 10^{-24} \text{ cm}^2 \tag{351}$$

is the Thomson scattering cross section, such that the Compton scattering rate (350) becomes,

$$\Gamma_{\rm Compton} \simeq 5 \times 10^{-21} \text{ s}^{-1} \Big(\frac{\Omega_b h^2}{0.022}\Big)(1+z)^3 X_e \tag{352}$$

This is to be compared with the Hubble parameter, which we calculate by assuming that an early radiation era is followed by matter era and a recent epoch of acceleration (which for simplicity we assume to be caused by a cosmological term),

$$H = H_0 \Big(\Omega_m (1+z)^3 + \Omega_\Lambda + \frac{\Omega_m}{1+z_{\rm eq}}(1+z)^4\Big)^{1/2}, \tag{353}$$

with $\Omega_m \simeq 0.26$ and $\Omega_\Lambda \simeq 1 - \Omega_m$.

When considered at recombination, $z_{\rm rec} = 1089$, the condition,

$$\Gamma_{\rm Compton} \sim H \tag{354}$$

results in

$$0.0043(1+z_{\rm rec})^{3/2} X_e \sim \Big(1 + \frac{1+z_{\rm rec}}{3230}\Big)^{1/2}, \tag{355}$$

where we took account of the radiation contribution, which is at recombination about 1/3 of that in nonrelativistic matter. This means that below the electron concentration

$$n_e|_{\rm dec} \sim 7.5 \times 10^{-3} n_b\,, \tag{356}$$

the Universe becomes transparent at recombination, and photons begin to free-stream. Since the freeze-out density is $X_e|_{\rm freeze} \sim 10^{-3}$ is only a few times smaller than (356), the Universe becomes barely transparent at recombination. The electron density (356) is used to define the surface of last scattering, which corresponds to $z_{\rm dec} \simeq 1089$. At redshifts larger (smaller) than $z_{\rm dec}$, the Universe is opaque (transparent) to radiation.

This is illustrated by the conformal diagram in figure 64, in which we show an observer O who measures the photons of the CMBR. Beyond the last scattering surface at $z_{\rm dec} \simeq 1089$, to a good approximation the Universe is in thermal equilibirum and photons scatter rapidly. At $z \ll z_{\rm dec}$ the

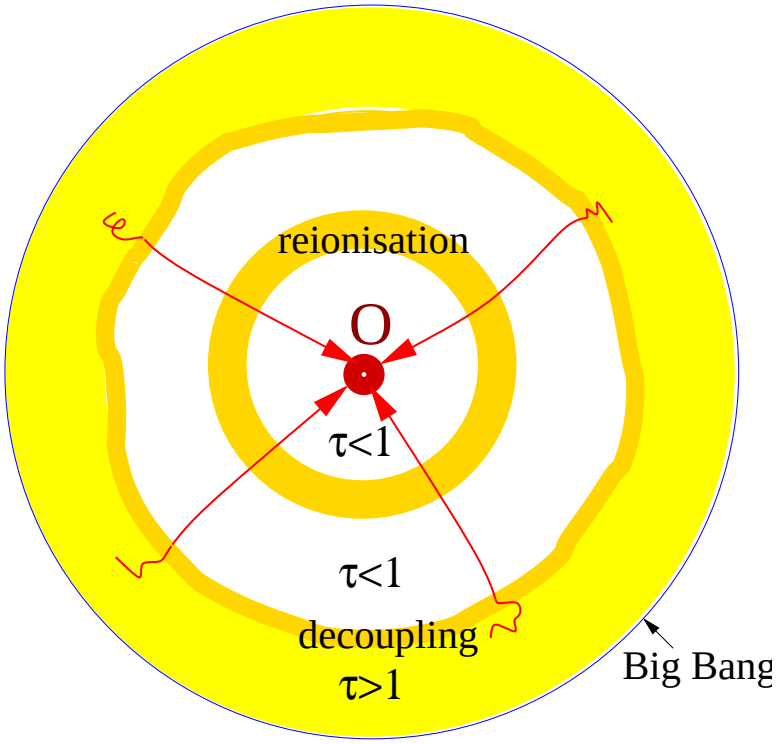


FIG. 64: A conformal diagram of an expanding Universe. The observer O is placed at the origin, $z = 0$. Photons arrive mostly uninterrupted (the optical depth $\tau < 1$) from the last scattering surface at $z_{\rm dec} \simeq 1089$ to us. At z 17 the universe reionises, and some photons do scatter, with the likelihood $p_{\rm scatter} = 1 - {\rm e}^{-\tau_{\rm reion}}$, with optical depth at the reionisation epoch, $\tau_{\rm reion} \sim 0.17$. At $z \gg 1089$ the Universe is a hot plasma, in which photons scatter readily, such that $\tau \gg 1$. The Big Bang is represented by the large outer circle at $z \to \infty$.

photons mostly free-stream. About 10% of the photons Compton scatter during the reionisation epoch around $z_{\rm reion} \sim 11$.

Imagine for a moment that there has never been any recombination. Then assuming a fully ionised Universe ($X_e = 1$), it follows from,

$$0.0043(1+z)^{3/2} X_e \sim 1 \tag{357}$$

that photons would begin free-streaming at about

$$z \simeq 37 \tag{358}$$

even if the Universe was fully ionized ($X_e = 1$), such that for $z < 37$ the Universe would in *any* case become transparent.

The polarisation measurements of the CMBR by the WMAP satellite suggest that the Universe was reionised at a redshift, $z_{\rm reion} = 10 \pm 3$. Assuming that the Universe was fully reionised at the redshift $z = 10$ and thereafter, and applying Eq. (355) to the epoch of reionisation, gives,

$$0.0042 \times (11)^{3/2} \simeq 0.15\,, \tag{359}$$

implying that about $\tau_{\rm reion} \sim 1/7$ of photons rescatter at $z \leq 10$. A more accurate analysis gives for the optical depth, $\tau_{\rm reion} \simeq 0.09$.

5. Dark ages, reionisation and large scale structures

The radio telescope LOFAR (LOw Frequency ARray) is under construction in the Netherlands (`http://www.lofar.nl/`) with the purpose to detect the signal from the $\lambda = 21$ cm hydrogen emission line from hydrogen gas clouds which permeate the integalactic medium at redshifts of about 10 and larger. This highly forbidden hydrogen emission line arises as a consequence of the hyperfine splitting between the spin triplet (in which the spins of the electron and proton are aligned) and the spin singlet hydrogen atom (in which the spins of the electron and proton are antiparallel). The corresponding frequency is $\nu = 1420.40575$ MHz and the transition rate is 2.9×10^{-15} s^{-1}. Due to the redshift the actual observed wavelength will be about 2m. These observations are expected to provide an invaluable information about reionisation and early formation of large scale structures at the end of *dark ages*, a period between recombination and reionisation from which as of yet no direct observations are available.

According to the polarisation measurement of the WMAP 3 year data (2006), the universe was completely reionised at a redshift,

$$z_{\rm reion} = 10 \pm 3 \tag{360}$$

with an optical depth

$$\begin{aligned} \tau_{\rm reion} &= \sigma_T \int n_e dl \\ &= 0.09 \pm 0.03 \end{aligned} \tag{361}$$

with $dl = cdt = (c/H)(da/a)$, and $\sigma_T = 0.665 \times 10^{-24}$ cm^2 being the Thomson cross section. The probability that a CMB photon gets Compton scattered can be expressed in terms of the optical depth at the reionisation epoch as follows,

$$p_{\rm scatter} = 1 - {\rm e}^{-\tau_{\rm reion}} \, . \tag{362}$$

The details of the reionisation history are not known, but the physics causing it is believed to be as follows. First over-densities in matter distribution formed population-three (Pop-III) stars, whose mass was much larger than the mass of the Sun, $M_{\rm III} \gg M_\odot$. These stars lived only very briefly (of the order $10^6 - 10^7$ years), before they burned their nuclear fuel and subsequently exploded as supernovae. These early supernovae, which are believed to have reionised the Universe, replenished it with first "metals" (carbon, nitrogen, oxygen, iron, *etc.*), essential for the formation of life, and hence these early stars grant a more careful investigation. Alternatively, the Universe could have been reionised by the active

early supermassive black holes which would sit in the centers of early quasars and which still furnish centers of some galaxies today.

At a redshift, $z \sim 10$, first galaxies, quasars, and clusters of galaxies were formed. Recall that quasars are believed to be QUAsiStellAR Objects (QSOs), whose diameter of the order one light day, with a supermassive black hole in its centrum. Even though quasars are so compact, they can emit as much radiation as a galaxy, which is sourced by the matter in-spiralling towards the central black hole. The recently completed Sloan Digital Sky Survey (SDSS) contains spectra of about 300000 quasars. Finally, about 4.6 billion years ago, the Solar System and the Earth formed.

D. Dark Energy Dominance

The distant supernovae observations by Perlmutter et al. and by Riess et al. (1998) suggest that the Universe is currently expanding at an accelerated pace, with the acceleration parameter,

$$-q_0 \equiv \frac{\ddot{a}_0}{H_0^2} \simeq 0.60 \pm 0.04\,, \tag{363}$$

where H_0 denotes the Hubble parameter today, and q_0 the deceleration parameter. Together with,

$$1 \simeq \Omega_m + \Omega_\Lambda + \Omega_k\,, \tag{364}$$

and the WMAP constraint,

$$\Omega_k \simeq 0\,, \tag{365}$$

the FLRW equation rewritten as

$$-q_0 = -\frac{1}{2}\Omega_m + \Omega_\Lambda \tag{366}$$

implies

$$q_0 \simeq -\frac{3}{2}\Omega_m - 1 \tag{367}$$

from where we infer,

$$\begin{aligned} \Omega_m &= \frac{2}{3}(1+q_0) \\ \Omega_\Lambda &= \frac{1}{3}(1-2q_0)\,. \end{aligned} \tag{368}$$

In Eq. (366) we neglected an eventual quintessence contribution, which is of the form, $-\frac{1}{2}(1+3w_Q)\Omega_Q$.

The redshift, at which the recent epoch of acceleration began, can be determined from the condition,

$$\Omega_m(1+z_\Lambda)^3 = \Omega_\Lambda \simeq 1 - \Omega_m\,, \tag{369}$$

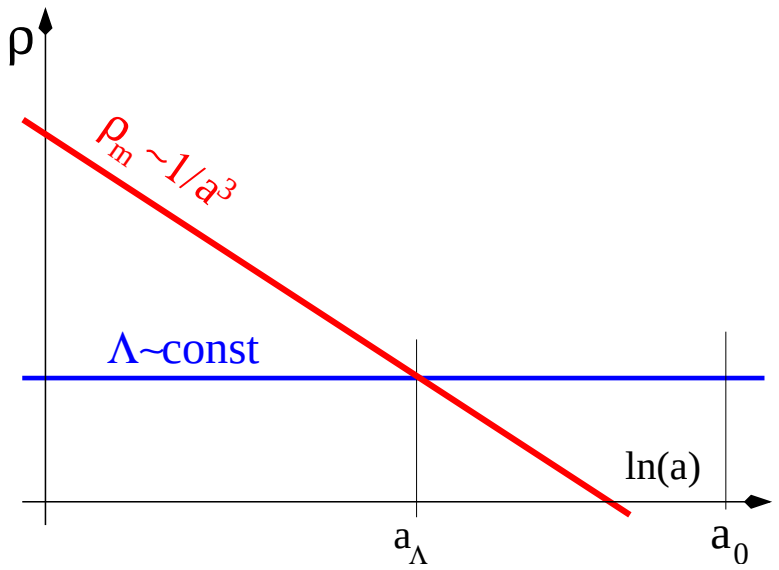


FIG. 65: The energy density scaling of nonrelativistic matter and dark energy (assumed to originate from a cosmological term) during recent epochs. The dark energy becomes a dominant component at a redshift, $z_\Lambda \simeq 0.4$.

from which one finds,

$$z_\Lambda \simeq \left(\frac{1-\Omega_m}{\Omega_m}\right)^{1/3} - 1 \simeq 0.4\,. \tag{370}$$

The scaling of the matter energy density and that in cosmological term are shown in figure 65. The fact that the two energy densities coincide at a recent epoch, $z_\Lambda \sim 0.4$, represents yet another coincidence in the Universe, which is poorly understood.