The violent Universe: the Big Bang

K. A. Olive
William I. Fine Theoretical Physics Institute, University of Minnesota, Minneapolis, MN 55455, USA

Abstract
Four lectures on Big Bang cosmology, including microwave background radiation, Big Bang nucleosynthesis, dark matter, inflation, and baryogenesis.

The Big Bang theory provides a detailed description of the history and evolution of the Universe. Direct experimental and observational evidence allows us to probe back to the first second after the initial state (bang) when the temperatures were of order 1 MeV and the light elements were created. Our understanding of the Standard Model of electroweak interactions allows us to push the description of the early universe back to about $10^{-10}$ s after the bang when we expect that the electroweak symmetry was restored. Indeed, it is possible to discuss events back to the Planck time ($10^{-44}$ s after the bang) albeit in a very model dependent way.

In these four lectures, I hope to give an overview of modern cosmology with an emphasis on particle interactions in cosmology. After a description of the standard FLRW model (including the microwave background radiation) in Lecture 1, I will cover both inflation and baryogenesis in Lecture 2. Lecture 3 will focus on Big Bang Nucleosynthesis (BBN) and Lecture 4 on dark matter.

1 Lecture 1: Standard Cosmology
1.1 The FLRW metric and its consequences

The standard Big Bang model assumes homogeneity and isotropy. As a result, one can construct the space-time metric by embedding a maximally symmetric three dimensional space in a four dimensional space-time (see, e.g., Ref. [1]). The most general form for a metric of this type is the Friedmann–Lemaître–Robertson–Walker metric which in co-moving coordinates is given by

$$ds^2 = dt^2 - R^2(t) \left[ \frac{dr^2}{(1 - kr^2)} + r^2 \left( d\theta^2 + \sin^2 \theta d\phi^2 \right) \right],$$

where $R(t)$ is the cosmological scale factor and $k$ the three-space curvature constant ($k = 0, +1, -1$ for a spatially flat, closed or open universe). $k$ and $R$ are the only two quantities in the metric which distinguish it from flat Minkowski space. It is also common to assume the perfect fluid form for the energy-momentum tensor

$$T^{\mu\nu} = pg^{\mu\nu} + (p + \rho)u^\mu u^\nu,$$

where $g_{\mu\nu}$ is the space-time metric described by Eq. (1), $p$ is the isotropic pressure, $\rho$ is the energy density and $u^\mu = (1, 0, 0, 0)$ is the velocity vector for the isotropic fluid. The (00) component of Einstein’s equation

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu} R - \Lambda g_{\mu\nu} = 8\pi G_N T_{\mu\nu}$$

yields the Friedmann equation

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

and the (ii) components give

$$\left( \frac{\ddot{R}}{R} \right) = \frac{\Lambda}{3} - \frac{4\pi G_N (\rho + 3p)}{3},$$

where $H$ is the Hubble parameter, $\Lambda$ is the cosmological constant, $\rho$ is the total energy density and $p$ is the total pressure.
where \( \Lambda \) is the cosmological constant. In addition, from \( T_{\mu \nu ; \nu} = 0 \), we obtain
\[
\dot{\rho} = -3H(\rho + p).
\]
(6)

Note that of these last three equations, only two are actually independent. These equations form the basis of the standard Big Bang model.

At early times \( (t < 10^5 \text{ yrs}) \) the Universe is thought to have been dominated by radiation so that the equation of state can be given by \( p = \rho/3 \). If we neglect the contributions to \( H \) from \( k \) and \( \Lambda \) (this is always a good approximation for small enough \( R \)) then we find that
\[
R(t) \sim t^{1/2}
\]
and \( \rho \sim R^{-4} \) so that \( t \sim (3/32\pi G_N \rho)^{1/2} \). Similarly for a matter or dust dominated universe with \( p = 0 \),
\[
R(t) \sim t^{2/3}
\]
and \( \rho \sim R^{-3} \). The Universe makes the transition between radiation and matter domination when \( \rho_{\text{rad}} = \rho_{\text{matter}} \) or when \( T \simeq \text{few} \times 10^3 \text{ K} \). In a vacuum or \( \Lambda \) dominated universe (that we are approaching today)
\[
R(t) \sim e^{\sqrt{\Lambda/3} t}.
\]
(9)

More general solutions for the behaviour of the scale factor can easily be found by defining a quantity \( Q \):
\[
Q = \frac{3k}{R^2} - 8\pi G_N \rho.
\]
(10)

If we further assume \( p = (\gamma - 1)\rho \) (with \( \gamma \) between 1 and 2), we have that \( \rho \sim R^{-3\gamma} \) and we can write
\[
\frac{\dot{R}}{R} = \pm \left[ \frac{\Lambda - Q}{3} \right]^{1/2},
\]
(11)

implying that \( Q \leq \Lambda \). For all choices of \( k \), the function \( Q \to -\infty \) as \( R \to 0 \). When \( k = +1 \), it is easy to see that \( Q \) has a maximum (with \( Q > 0 \)) and then tends to 0 as \( R \to \infty \). When \( k = 0 \) or \( -1 \), \( Q \) monotonically increases towards 0 as \( R \to \infty \). In Fig. 1, the function \( Q \) is plotted qualitatively as a function of the scale factor.

Let us first consider the more interesting case of \( k = +1 \). For a cosmological constant \( \Lambda > Q_{\text{max}} \), \( Q < \Lambda \) for all \( R \). In this case, two distinct solutions are possible: the universe may expand forever from a singularity at \( R = 0 \) to \( R = \infty \), or by choosing the lower sign in Eq. (11), the universe may collapse from infinity to a singularity.

There are also two possible solutions for \( 0 < \Lambda < Q_{\text{max}} \), the case depicted in Fig. 1. The universe may again start at a singularity at \( R = 0 \) and expand to the point where \( Q = \Lambda \) and then recollapse. Alternatively, the Universe may start from infinity and collapse to the point where \( Q = \Lambda \) (at larger \( R \)), bounce back and expand to infinity. When \( \Lambda = Q_{\text{max}} \) there are a total of five solutions. At the value of \( R \) such that \( Q = Q_{\text{max}} \), we have the Einstein Static Universe. This is the only solution which is neither expanding nor contracting. The remaining four solutions either asymptotically expand or contract towards or away from the Einstein Static case. Finally for \( \Lambda < 0 \), there is only one solution for which the universe expands and subsequently recollapses.

Finding the solutions when \( k = 0 \) or \( -1 \) is similar, and the behaviour of \( R \) depends only on whether \( \Lambda \) is positive or negative. For \( \Lambda \geq 0 \) there are two solutions, for \( \Lambda < 0 \), there is only one. When \( \Lambda = 0 \), we obtain the standard notion that closed universes recollapse, while open and flat universes expand forever. When \( \Lambda \neq 0 \), these simple associations are spoiled as a closed universe can expand forever (for large enough \( \Lambda \)) and open and flat universes can recollapse (for \( \Lambda < 0 \)).
Fig. 1: The qualitative behaviour of the scale factor $R$ as obtained using the condition $Q \leq \Lambda$ for the three choices of the curvature constant $k$.

Exact solutions to the equations of motion can be obtained relatively easily in terms of conformal coordinates. We can, for example, rewrite the metric as

$$ds^2 = dt^2 - R^2(t) \left[ d\chi^2 + f^2(\chi) \left( d\theta^2 + \sin^2 \theta d\phi^2 \right) \right],$$

where

$$f(\chi) = \begin{cases} \sinh \chi & k = -1 \\ \chi & k = 0 \\ \sin \chi & k = +1 \end{cases}.$$  \hspace{1cm} (13)

We can go further and define a conformal time coordinate using $Rd\eta = dt$ so that

$$ds^2 = R^2(\eta) \left[ dt^2 - d\chi^2 + f^2(\chi) \left( d\theta^2 + \sin^2 \theta d\phi^2 \right) \right].$$

In terms of these coordinates, the Friedmann equation becomes

$$R'' + kR = \frac{4\pi G}{3} (\rho - 3p)R^3,$$  \hspace{1cm} (15)

where $'$ denotes a derivative with respect to $\eta$ and is easily solved

$$R \propto \begin{cases} \cosh \eta - 1 & k = -1 \\ \eta^2/2 & k = 0 \\ 1 - \cos \eta & k = +1 \end{cases}$$ \hspace{1cm} \begin{cases} \sinh \eta - \eta & k = -1 \\ \eta^3/6 & k = 0 \\ \eta - \sin \eta & k = +1 \end{cases}$$

for $p = 0$ and

$$R \propto \begin{cases} \sinh \eta & k = -1 \\ \eta & k = 0 \\ \sin \eta & k = +1 \end{cases}$$ \hspace{1cm} \begin{cases} \cosh \eta - 1 & k = -1 \\ \eta^2/2 & k = 0 \\ 1 - \cos \eta & k = +1 \end{cases}$$

for $p = \rho/3$. 

3
In the absence of a cosmological constant, one can define a critical energy density \( \rho_c \) such that \( \rho = \rho_c \) for \( k = 0 \)

\[
\rho_c = \frac{3H^2}{8\pi G_N}.
\tag{18}
\]

In terms of the present value of the Hubble parameter this is

\[
\rho_c = 1.88 \times 10^{-29} h_0^2 \text{ g cm}^{-3},
\tag{19}
\]

where

\[
h_0 = H_0/(100 \text{ km Mpc}^{-1} \text{ s}^{-1}).
\tag{20}
\]

The cosmological density parameter is then defined by

\[
\Omega \equiv \frac{\rho}{\rho_c}.
\tag{21}
\]

It is useful to also define a deceleration parameter

\[
q_0 = -\frac{\ddot{R}_0}{R_0^2}
\tag{22}
\]

This standard definition was formulated under the presumption that the expansion rate of the Universe is in fact slowing down. As noted above, and discussed further below, modern measurements indicate the opposite. That is, the expansion is accelerating (meaning that \( q_0 < 0 \)).

The Friedmann equation also lends itself to integration to determine the age of the Universe. Note that for a given equation of state, we can write \( \rho = \rho_0\left(R_0/R\right)^{3\gamma} \), and

\[
H_0t = \int_0^1 \frac{dx}{[1 - \Omega_0 - \Omega_\Lambda + \Omega_0 x^{2-3\gamma} + \Omega_\Lambda x^2]^{1/2}},
\tag{27}
\]

where \( \Omega_\Lambda = \Lambda/3H^2_0 \). When \( \Lambda = 0 \) and \( \Omega_0 = 1 \), this is easily integrated to give

\[
t = \frac{2}{3H_0}, \quad \gamma = 1
\tag{28}
\]
Fig. 2: The evolution of the cosmological density parameter $\Omega$ as a function of the scale factor for a closed, open and spatially flat Universe

or

$$t = \frac{1}{2H_0} \gamma = \frac{4}{3}. \quad (29)$$

Because of the finite age of the Universe in the Big Bang model, there is a particle horizon corresponding to the maximum distance traversed by light. In general, we can write the proper distances between a location specified by comoving coordinate $t_1, r_1$ and our location at $t$ and $r = 0$ in terms of the metric (1)

$$d_p = R(t) \int_0^{r_1} \frac{dr'}{\sqrt{1 - kr'^2}}. \quad (30)$$

For light paths ($ds^2 = 0$)

$$\int_{t_1}^{t} \frac{dt'}{R(t')} = \int_0^{r_1} \frac{dr'}{\sqrt{1 - kr'^2}}. \quad (31)$$

As $t_1 \to 0$, $r_1$ becomes the maximum coordinate distance from which we can receive a signal. Thus the particle horizon is defined by

$$d_H = R(t) \int_0^{r_H} \frac{dr'}{\sqrt{1 - kr'^2}} = R(t) \int_0^{t} \frac{dt'}{R(t')}. \quad (32)$$

For $\Lambda = 0$ and $\Omega_0 = 1$, we again obtain very simple solutions:

$$d_H = 3t \quad \gamma = 1 \quad (33)$$

or

$$d_H = 2t \quad \gamma = \frac{4}{3}. \quad (34)$$

Note that because $d_H/R$ grows with time, we see more of the Universe as time goes on. That is, new information is continuously entering our particle horizon.
1.2 The hot thermal Universe

The epoch of recombination occurs when electrons and protons form neutral hydrogen through $e^- + p \rightarrow H + \gamma$ at a temperature $T_R \sim \text{few} \times 10^3$ K $\sim 1$ eV. For $T < T_R$, photons are decoupled while for $T > T_R$, photons are in thermal equilibrium and at higher temperatures, the Universe is radiation dominated and the content of the radiation plays a very important role. Today, the content of the microwave background consists of photons with $T_0 = 2.725 \pm 0.001$ K [3]. We can calculate the energy density of photons from

$$\rho_\gamma = \int E_\gamma dn_\gamma, \quad (35)$$

where the density of states is given by

$$dn_\gamma = \frac{g_\gamma}{2\pi^2} [\exp(E_\gamma/T) - 1]^{-1} q^2 dq, \quad (36)$$

and $g_\gamma = 2$ simply counts the number of degrees of freedom for photons, $E_\gamma = q$ is just the photon energy (momentum). (I am using units such that $\hbar = c = k_B = 1$ and will do so through the remainder of these lectures.) Integrating Eq. (35) gives

$$\rho_\gamma = \frac{\pi^2}{15} T^4, \quad (37)$$

which is the familiar blackbody result. In addition, we also have

$$p = \frac{1}{3} \rho \quad s = \frac{4}{3} \frac{\rho}{T} \quad n_\gamma = \frac{2\zeta(3)}{\pi^2} T^3. \quad (38)$$

In general, at very early times, at very high temperatures, other particle degrees of freedom join the radiation background when $T \sim m_i$ for each particle type $i$ if that type is brought into thermal equilibrium through interactions. In equilibrium, the energy density of a particle type $i$ is given by

$$\rho_i = \int E_i dn_{q_i}, \quad (39)$$

and

$$dn_{q_i} = \frac{g_i}{2\pi^2} [\exp[(E_{q_i} - \mu_i)/T] \pm 1]^{-1} q^2 dq, \quad (40)$$

where again $g_i$ counts the total number of degrees of freedom for type $i$,

$$E_{q_i} = (m_i^2 + q_i^2)^{1/2}, \quad (41)$$

$\mu_i$ is the chemical potential if present and $\pm$ corresponds to either Fermi or Bose statistics.

In the limit that $T \gg m_i$ the total energy density can be conveniently expressed by

$$\rho = \left( \sum_B g_B + \sum_F g_F \right) \frac{\pi^2}{30} T^4 \equiv \frac{\pi^2}{30} N(T) T^4, \quad (42)$$

where $g_{B(F)}$ are the total number of boson (fermion) degrees of freedom and the sum runs over all boson (fermion) states with $m \ll T$. The factor of 7/8 is due to the difference between the Fermi and Bose integrals. Equation (42) defines $N(T)$ by taking into account new particle degrees of freedom as the temperature is raised.

In the radiation dominated epoch, Eq. (6) can be integrated (neglecting the $T$-dependence of $N$) giving us a relationship between the age of the Universe and its temperature

$$t = \left( \frac{90}{32\pi^4 G_N N(T)} \right)^{1/2} T^{-2}. \quad (43)$$
Table 1: Effective numbers of degrees of freedom in the Standard Model

<table>
<thead>
<tr>
<th>Temperature</th>
<th>New particles</th>
<th>$4N(T)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T &lt; m_{\gamma}$</td>
<td>$\gamma$'s + $\nu$'s</td>
<td>29</td>
</tr>
<tr>
<td>$m_{\gamma} &lt; T &lt; m_{\mu}$</td>
<td>$e^\pm$</td>
<td>43</td>
</tr>
<tr>
<td>$m_{\mu} &lt; T &lt; m_{\pi}$</td>
<td>$\mu^\pm$</td>
<td>57</td>
</tr>
<tr>
<td>$m_{\pi} &lt; T &lt; T_c^*$</td>
<td>$\pi$'s</td>
<td>69</td>
</tr>
<tr>
<td>$T_c &lt; T &lt; m_{charm}$</td>
<td>$\pi$'s + $u, \bar{u}, d, \bar{d}, s, \bar{s}$ + gluons</td>
<td>247</td>
</tr>
<tr>
<td>$m_{c} &lt; T &lt; m_{\tau}$</td>
<td>$c, \bar{c}$</td>
<td>289</td>
</tr>
<tr>
<td>$m_{\tau} &lt; T &lt; m_{\text{bottom}}$</td>
<td>$\tau^\pm$</td>
<td>303</td>
</tr>
<tr>
<td>$m_{b} &lt; T &lt; m_{W,Z}$</td>
<td>$b, \bar{b}$</td>
<td>345</td>
</tr>
<tr>
<td>$m_{W,Z} &lt; T &lt; m_{\text{Higgs}}$</td>
<td>$W^\pm, Z$</td>
<td>381</td>
</tr>
<tr>
<td>$m_{H} &lt; T &lt; m_{\text{top}}$</td>
<td>$H^0$</td>
<td>385</td>
</tr>
<tr>
<td>$m_{t}$ &lt; $T$</td>
<td>$t, \bar{t}$</td>
<td>427</td>
</tr>
</tbody>
</table>

*Tc corresponds to the confinement–deconfinement transition between quarks and hadrons. $N(T)$ is shown in Fig. 3 for $T_c = 150$ and $400$ MeV.

Put into a more convenient form

$$t^2 T_{\text{MeV}}^2 = 2.4 [N(T)]^{-1/2}. \quad (44)$$

where $t$ is measured in seconds and $T_{\text{MeV}}$ in units of MeV.

The value of $N(T)$ at any given temperature depends on the particle physics model. In the standard $SU(3) \times SU(2) \times U(1)$ model, we can specify $N(T)$ up to temperatures of order 100 GeV. The value of $N$ in the Standard Model can be seen in Table 1.

At higher temperatures, $N(T)$ will be model dependent. For example, in the minimal $SU(5)$ model, one needs to add to $N(T)$, 24 states for the X and Y gauge bosons, another 24 from the adjoint Higgs, and another 6 (in addition to the 4 already counted in $W^\pm, Z$ and $H$) from the 5 of Higgs. Hence for $T > M_X$ in minimal $SU(5)$, $N(T) = 160.75$. In a supersymmetric model this would at least double, with some changes possibly necessary in the table if the lightest supersymmetric particle has a mass below $m_t$.

The notion of equilibrium also plays an important role in the standard Big Bang model. If, for example, the Universe were not expanding, then given enough time, each particle state would come into equilibrium with every other. Because of the expansion of the Universe, certain rates might be too slow indicating, for example, in a scattering process that the two incoming states might never find each other to bring about an interaction. Depending on their rates, certain interactions may pass in and out of thermal equilibrium during the course of the Universal expansion. Qualitatively, for each particle $i$, we will require that some rate $\Gamma_i$ involving that type be larger than the expansion rate of the Universe or

$$\Gamma_i > H, \quad (45)$$

in order to be in thermal equilibrium.

A good example for a process in equilibrium at some stage and out of equilibrium at others is that of neutrinos. If we consider the standard neutral or charged-current interactions such as $e^+ + e^- \leftrightarrow \nu + \bar{\nu}$ or $e + \nu \leftrightarrow e + \nu$ etc., very roughly the rates for these processes will be

$$\Gamma = n \langle \sigma v \rangle, \quad (46)$$

where $\langle \sigma v \rangle$ is the thermally averaged weak interaction cross section

$$\langle \sigma v \rangle \sim \mathcal{O}(10^{-2}) T^2 / M_W^4, \quad (47)$$
and $n \sim T^3$ is the number density of leptons. Hence the rate for these interactions is

$$\Gamma_{\text{wk}} \sim 0(10^{-2})T^5/M_W^3.$$  \hfill{(48)}

The expansion rate, on the other hand, is just

$$H = \left(\frac{8\pi G_N \rho}{3}\right)^{1/2} = \left(\frac{8\pi^3}{90}N(T)\right)^{1/2} T^2/M_P \sim 1.66N(T)^{1/2}T^2/M_P.$$  \hfill{(49)}

The Planck mass $M_P = G^{-1/2}_N = 1.22 \times 10^{19}$ GeV.

Neutrinos will be in equilibrium when $\Gamma_{\text{wk}} > H$ or

$$T > (500M_W^4/M_P)^{1/3} \sim 1 \text{ MeV}.$$  \hfill{(50)}

The temperature at which these rates are equal is commonly referred to as the decoupling or freeze-out temperature and is defined by

$$\Gamma(T_d) = H(T_d).$$  \hfill{(51)}

For temperatures $T > T_d$, neutrinos will be in equilibrium, while for $T < T_d$ they will not. Basically, in terms of their interactions, the expansion rate is just too fast and they never ‘see’ the rest of the matter in the Universe (nor themselves). Their momenta will simply redshift and their effective temperature (the shape of their momenta distribution is not changed from that of a blackbody) will simply fall with $T \sim 1/R$.

The relation $RT \sim \text{const}$ is a direct consequence of the energy conservation equation (6). Indeed, using $s = dp/dT$, this equation can be rewritten as

$$\frac{d}{dt}(R^3 s) = 0,$$  \hfill{(52)}

making it more a statement of conservation of (comoving) entropy than energy (which is not conserved in comoving coordinates).

Soon after neutrino decoupling, the $e^\pm$ pairs in the thermal background begin to annihilate (when $T \lesssim m_e$). Because the neutrinos are decoupled, the energy released heats up the photon background.
relative to the neutrinos. The change in the photon temperature can easily be computed from entropy conservation. The neutrino entropy must be conserved separately from the entropy of interacting particles. If we denote \( T_\gamma > \) the temperature of photons, and \( e^\pm \) before annihilation, we also have \( T_\nu = T_\gamma \) as well. The entropy density of the interacting particles at \( T = T_\gamma \) is just

\[
s_\gamma = \frac{4}{3} \rho_\gamma = \frac{4}{3} \left( 2 + \frac{7}{2} \right) \left( \frac{\pi^2}{30} \right) T_\gamma^3,
\]

while at \( T = T_\nu < \), the temperature of the photons just after \( e^\pm \) annihilation, the entropy density is

\[
s_\nu = \frac{4}{3} \rho_\nu = \frac{4}{3} \left( 2 + \frac{7}{2} \right) \left( \frac{\pi^2}{30} \right) T_\nu^3,
\]

and by conservation of entropy \( s_\nu = s_\gamma \) and

\[
(T_\nu/T_\gamma)^3 = \frac{11}{4}.
\]

Thus, the photon background is at higher temperature than the neutrinos because the \( e^\pm \) annihilation energy could not be shared among the neutrinos, and

\[
T_\nu = (4/11)^{1/3} T_\gamma \simeq 1.9 \, \text{K}.
\]

For further reading on standard Big Bang cosmology see Refs. [4–9].

1.3 The Cosmic Microwave Background

There has been a great deal of progress in the last several years concerning the determination of both \( \Omega_m \) and \( \Omega_\Lambda \). Cosmic Microwave Background (CMB) anisotropy experiments have been able to determine the curvature (i.e., the sum of \( \Omega_m \) and \( \Omega_\Lambda \)) to better than one per cent, while observations of type Ia supernovae at high redshift and baryon acoustic oscillations provide information on (nearly) orthogonal combinations of the two density parameters.

The CMB is of course deeply rooted in the development and verification of the Big Bang model and Big Bang Nucleosynthesis (BBN) [10]. Indeed, it was the formulation of BBN that led to the prediction of the microwave background [11]. The argument is rather simple. BBN requires temperatures greater than 100 keV, which according to the Standard Model time–temperature relation, Eq. (44), corresponds to timescales less than about 200 s. The typical cross section for the first link in the nucleosynthetic chain is

\[
\sigma v (p + n \rightarrow D + \gamma) \simeq 5 \times 10^{-20} \, \text{cm}^3/\text{s}.
\]

This implies that it was necessary to achieve a density

\[
n \sim \frac{1}{\sigma v t} \sim 10^{17} \, \text{cm}^{-3},
\]

for nucleosynthesis to begin. The density in baryons today is known approximately from the density of visible matter to be \( n_{B0} \sim 10^{-7} \, \text{cm}^{-3} \) and since we know that the density \( n \) scales as \( R^{-3} \sim T^3 \), the temperature today must be

\[
T_0 = (n_{B0}/n)^{1/3} T_{BBN} \sim 10 \, \text{K}
\]

thus linking two of the most important tests of the Big Bang theory.

Of course it was not until many years later that the microwave background radiation was discovered by Penzias and Wilson [12] while perfecting a radio antenna to track the Echo satellite. They found a background noise which could not be eliminated corresponding to a temperature of \( 3.5 \pm 1 \, \text{K} \). One of the most important papers on modern cosmology was published with the title “A measurement of excess
antenna temperature at 4080-Mc/s". This was followed by the seminal paper by Dicke, Peebles, Roll, and Wilkinson [13] putting this observation in a cosmological context. Subsequently, there have been many observations of the CMB culminating in the COBE observation [14] which determined the temperature to an unprecedented level, set aside any lingering doubts about the true black body nature of the CMB, and discovered the intrinsic anisotropies in the background.

An enormous amount of cosmological information is encoded in the angular expansion of the CMB temperature

$$T(\theta, \phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta, \phi).$$

(60)

The monopole term characterizes the mean background temperature of $T_\gamma = 2.725 \pm 0.001$ K as determined by COBE [3], whereas the dipole term can be associated with the Doppler shift produced by our peculiar motion with respect to the CMB. In contrast, the higher order multipoles are directly related to energy density perturbations in the early Universe. When compared with theoretical models, the higher order anisotropies can be used to constrain several key cosmological parameters. In the context of simple adiabatic cold dark matter (CDM) models, there are nine of these: the cold dark matter density, $\Omega_m h^2$; the baryon density, $\Omega_B h^2$; the curvature — characterized by $\Omega_{\text{total}}$; the hubble parameter, $h$; the optical depth, $\tau$; the spectral indices of scalar and tensor perturbations, $n_s$ and $n_t$; the ratio of tensor to scalar perturbations, $r$; and the overall amplitude of fluctuations, $Q$.

Microwave background anisotropy measurements have made tremendous advances in the last few years. The power spectrum [15–23] has been measured relatively accurately out to multipole moments corresponding to $\ell \sim 2000$. A compilation of recent data is shown in Fig. 4 [24], where the power at
Fig. 5: Two-dimensional confidence regions in the \((\Omega_m, \Omega_\Lambda)\) plane. The coloured Monte Carlo points derive from WMAP [21] and show that the CMB alone requires a flat Universe \(\Omega_\Lambda + \Omega_m \approx 1\) if the Hubble constant is not too high. The SNe Ia results [27] very nearly constrain the orthogonal combination \(\Omega_\Lambda - \Omega_m\). Also shown is the region obtained from baryon acoustic oscillations [29]. The intersection of these constraints is the most direct (but far from the only) piece of evidence favouring a flat model with \(\Omega_m \approx 0.25\).

As indicated above, the details of this spectrum enable one to make accurate predictions of a large number of fundamental cosmological parameters. The results of the WMAP data (with other information concerning the power spectrum) is shown in Table 2. For details see Refs. [2, 24, 25].

<table>
<thead>
<tr>
<th>(\Omega_m h^2)</th>
<th>WMAP alone</th>
<th>WMAP + BAO + SN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.133 ± 0.006</td>
<td>0.136 ± 0.004</td>
<td></td>
</tr>
<tr>
<td>(\Omega_B h^2)</td>
<td>0.0227 ± 0.0006</td>
<td>0.0227 ± 0.0006</td>
</tr>
<tr>
<td>(h)</td>
<td>0.72 ± 0.03</td>
<td>0.705 ± 0.013</td>
</tr>
<tr>
<td>(n_s)</td>
<td>0.963±0.014</td>
<td>0.960 ± 0.013</td>
</tr>
<tr>
<td>(\tau)</td>
<td>0.087 ± 0.017</td>
<td>0.084 ± 0.016</td>
</tr>
<tr>
<td>(\Omega_\Lambda)</td>
<td>0.74 ± 0.03</td>
<td>0.726 ± 0.015</td>
</tr>
</tbody>
</table>

Of particular interest to us here is the CMB determination of the total density \(\Omega_{\text{tot}}\) as well as the matter density \(\Omega_m\). There is strong evidence that the Universe is flat or very close to it. As noted earlier, the best determination of \(\Omega_{\text{tot}}\) is 1.006 ± 0.006. Furthermore, the matter density is significantly larger than the baryon density implying the existence of cold dark matter. Also, the baryon density, as we will see below, is consistent with the BBN production of D/H and its abundance in quasar absorption systems. The apparent discrepancy between the CMB value of \(\Omega_{\text{tot}}\) and \(\Omega_m\), though not conclusive on its own, is a sign that a contribution from the vacuum energy density or cosmological constant, is also required. The preferred region in the \(\Omega_m - \Omega_\Lambda\) plane is shown in Fig. 5.

The presence or absence of a cosmological constant is a long standing problem in cosmology. We
know that the cosmological term is at most a factor of a few times larger than the current mass density. Thus from Eq. (4), we see that the dimensionless combination $G_N\Lambda \lesssim 10^{-121}$. Also shown in Fig. 5, are the results from SN Ia [26, 27] and baryon acoustic oscillations [28]. Taken together, we are led to a seemingly conclusive picture. The Universe is nearly flat with $\Omega_{\text{tot}} \simeq 1$. However, the density in matter makes up only 23% of this total, with the remainder in a cosmological constant or some other form of dark energy.

2 Lecture 2: Inflation and Baryogenesis

Despite the successes of the standard Big Bang model, there are a number of unanswered questions that appear difficult to explain without imposing what may be called unnatural initial conditions. The resolution of these problems may lie in a unified theory of gauge interactions or possibly in a theory which includes gravity. For example, prior to the advent of grand unified theories (GUTs), the baryon-to-photon ratio, could have been viewed as being embarrassingly small. Although we still do not know the precise mechanism for generating the baryon asymmetry, many quite acceptable models are available. In a similar fashion, it is hoped that a field theoretic description of inflation may resolve the problems outlined below.

2.1 Cosmological problems

2.1.1 The curvature problem

As noted above, the determined value of $\Omega$ in Eq. (26) is curious since at the present time we do not know even the sign of the curvature term in the Friedmann equation (4), i.e., we do not know if the Universe is open, closed or spatially flat.

The curvature problem (or flatness problem or age problem) can manifest itself in several ways. For a radiation dominated gas, the entropy density $s \sim T^3$ and $R \sim T^{-1}$. Thus assuming an adiabatically expanding Universe, the quantity $\dot{k} = k/R^2T^2$ is a dimensionless constant. If we now apply the limit in Eq. (26) to Eq. (24) (with $\Lambda = 0$) we find

$$\dot{k} = \frac{k}{R^2T^2} = \frac{(\Omega_0 - 1)H_0^2}{T_0^2} \lesssim 10^{-60}. \quad (61)$$

This limit on $\dot{k}$ represents an initial condition on the cosmological model. The problem then becomes what physical processes, if any, in the early Universe produced a value of $\dot{k}$ so extraordinarily close to zero (or $\Omega$ close to one). A more natural initial condition might have been $\dot{k} \sim 0(1)$. In this case the Universe would have become curvature dominated at $T \sim 10^{-1}M_P$. For $k = +1$, this would signify the onset of recollapse. As already noted earlier, one would naturally expect the effects of curvature (seen in Fig. 2 by the separation of the three curves) to manifest themselves at times on the order of the Planck time as gravity should provide the only dimensionful scale in this era. If we view the evolution of $\Omega$ in Fig. 2 as a function of time, then it would appear that the time $t_0 = 13.7 \text{ Gyr} = 8 \times 10^{60}M_P^{-1} (\sim$ the current age of the Universe) appears at the far left of the x-axis, i.e., before the curves separate. Why then has the Universe lasted so long before revealing the true sign of $k$?

2.1.2 The horizon problem

Because of the cosmological principle, all physical length scales grow as the scale factor $R(t) \sim t^{2/3\gamma}$, with $\gamma$ defined by $p = (\gamma - 1)\rho$. However, as we have seen, there is a particle horizon $d_H(t) \sim t$ as defined in Eq. (32). For $\gamma > \frac{2}{3}$, scales originating outside of the horizon will eventually become part of our observable Universe. Hence we would expect to see anisotropies on large scales [30].

In particular, let us consider the microwave background today. The photons we observe have been decoupled since recombination at $T_d \sim 3000 \text{ K}$. At that time, the horizon volume was simply
$V_d \propto t_d^3$, where $t_d$ is the age of the Universe at $T = T_d$. Then $t_d = t_0(T_0/T_d)^{3/2} \sim 2 \times 10^5$ yrs, where $T_0 = 2.725$ K [3] is the present temperature of the microwave background. Our present horizon volume $V_0 \propto t_0^3$ can be scaled back to $t_d$ (corresponding to that part of the Universe which expanded to our present visible Universe) $V_0(t_d) \propto V_0(T_0/T_d)^3$. We can now compare $V_0(t_d)$ and $V_d$. The ratio

$$\frac{V_0(t_d)}{V_d} \propto \frac{V_0(t_0^3)}{V_d t_d^{-3}} \propto \frac{t_0^3 T_0^3}{t_d^3 T_d^3} \sim 5 \times 10^4$$

(62)

corresponds to the number of horizon volumes or causally distinct regions at decoupling which are encompassed in our present visible horizon.

In this context, it is astonishing that the microwave background appears highly isotropic on large scales with $\Delta T/T = 1.1 \pm 0.2 \times 10^{-5}$ at angular separations of $10^\circ$ [14]. The horizon problem, therefore, is the lack of an explanation as to why nearly $10^5$ causally disconnected regions at recombination all had the same temperature to within one part in $10^{-5}$.

### 2.1.3 Density perturbations

Although it appears that the Universe is extremely isotropic and homogeneous on very large scales (in fact the Standard Model assumes complete isotropy and homogeneity) it is very inhomogeneous on small scales. In other words, there are planets, stars, galaxies, clusters, etc. On these small scales there are large density perturbations namely $\delta \rho/\rho$ background limits force $\delta \rho/\rho \propto 1$ on smaller scales.

In an expanding Universe, density perturbations evolve with time [4]. The evolution of the Fourier transformed quantity $\delta \rho(k, t)$ depends on the relative size of the wavelength $\lambda \propto k^{-1}$ and the horizon scale $H^{-1}$. For $k \ll H$, (always true at sufficiently early times) $\delta \rho/\rho \propto t$ while for $k \gg H$, $\delta \rho/\rho$ is constant (or grows moderately as $\ln t$) assuming a radiation dominated Universe. In a matter dominated Universe, on scales larger than the Jeans length scale (determined by $k_J = \frac{4\pi Gm_{\text{matter}}}{v_s^2}$, $v_s =$ sound speed) perturbations grow with the scale factor $R$. Because of the growth in $\delta \rho/\rho$, the microwave background limits force $\delta \rho/\rho$ to be extremely small at early times.

Consider a perturbation with wavelength on the order of a galactic scale. Between the Planck time and recombination, such a perturbation would have grown by a factor of $O(10^{57})$ and the anisotropy limit of $\delta \rho/\rho \lesssim 10^{-5}$ implies that $\delta \rho/\rho \lesssim 10^{-61}$ on the scale of a galaxy at the Planck time. One should compare this value with that predicted from purely random (or Poisson) fluctuations of $\delta \rho/\rho \sim 10^{-40}$ (assuming $10^{90}$ particles (photons) in a galaxy) [31]. The extent of this limit is of course related to the fact that the present age of the Universe is so great.

An additional problem is related to the formation time of the perturbations. A perturbation with a wavelength large enough to correspond to a galaxy today must have formed with wavelength modes much greater than the horizon size if the perturbations are primordial, as is generally assumed. This is due to the fact that the wavelengths red-shift as $\lambda \sim R \sim t^{1/2}$ while the horizon size grows linearly. It would appear that a mechanism for generating perturbations with acausal wavelengths is required.

### 2.1.4 The magnetic monopole problem

In addition to the much desired baryon asymmetry produced by grand unified theories, a less favourable aspect is also present; GUTs predict the existence of magnetic monopoles. Monopoles will be produced [32] whenever any simple group [such as $SU(5)$] is broken down to a gauge group which contains a $U(1)$ factor [such as $SU(3) \times SU(2) \times U(1)$]. The mass of such a monopole would be

$$M_m \sim m_{\text{GUT}}/\alpha_{\text{GUT}} \sim 10^{16} \text{ GeV}.$$  

(63)
The basic reason monopoles are produced is that in the breaking of SU(5), the Higgs adjoint needed to break SU(5) cannot align itself over all space \[33\]. On scales larger than the horizon, for example, there is no reason to expect the direction of the Higgs field to be aligned. Because of this randomness, topological knots are expected to occur and these are the magnetic monopoles. We can then estimate that the minimum number of monopoles produced \[34\] would be roughly one per horizon volume or causally connected region at the time of the SU(5) phase transition \( t_c \),

\[
n_m \sim (2t_c)^{-3},
\]

resulting in a monopole-to-photon ratio expressed in terms of the transition temperature of

\[
\frac{n_m}{n_\gamma} \sim \left(\frac{10T_c}{M_T}\right)^3.
\]

The overall mass density of the Universe can be used to place a constraint on the density of monopoles. For \( M_m \sim 10^{16} \text{ GeV} \) and \( \Omega_m h^2 < 1 \) we have that

\[
\frac{n_m}{n_\gamma} < 0(10^{-25}).
\]

The predicted density, however, from Eq. (65) for \( T_c \sim M_{\text{GUT}} \) is

\[
\frac{n_m}{n_\gamma} \sim 10^{-9}.
\]

Hence we see that standard GUTs and cosmology have a monopole problem.

### 2.2 Inflation

All of the problems discussed above can be neatly resolved if the Universe underwent a period of cosmological inflation \[35, 36\]. That is, if the Universe at some stage becomes dominated by the vacuum as could be the case during a phase transition, our assumptions of an adiabatically expanding Universe may not be valid. Indeed, we expect several cosmological phase transitions to have occurred including the breakdown of a Grand Unified symmetry such as SU(5) \( \rightarrow \) SU(3)c \( \times \) SU(2)L \( \times \) U(1)Y or the electroweak transition SU(2)L \( \times \) U(1)Y \( \rightarrow \) U(1)em, or possibly some other non-gauged transition.

During a phase transition, the motion of a scalar field will be described by a scalar potential. If the solution to the equations of motion for the scalar field leads to a slowly evolving scalar field (this will depend on the details of the potential), the Universe may become dominated by the vacuum energy density associated with the potential near the initial field value, say \( \phi \approx 0 \). The energy density of the symmetric vacuum \( V(0) \) acts as a cosmological constant with

\[
\Lambda = \frac{8\pi V(0)}{M_P^2}.
\]

During this period of slow evolution, the energy density due to radiation or matter will fall below the vacuum energy density, \( \rho \ll V(0) \). When this happens, the expansion rate will be dominated by the constant \( V(0) \) and from Eq. (4) we find the exponentially expanding solution given in Eq. (9). When the field evolves towards the global minimum it will begin to oscillate about the minimum, energy will be released during its decay and a hot thermal Universe will be restored. If released fast enough, it will produce radiation at a temperature \( T_R \approx V(0) \). In this reheating process, entropy has been created and \( (RT)_f > (RT)_i \). Thus we see that during a phase transition, the relation \( RT \sim \text{constant} \) need not hold true and our dimensionless constant \( \hat{k} \) may actually not have been constant.

If during the phase transition the value of \( RT \) changed by a factor of \( O(10^{30}) \), the cosmological problems discussed above would be solved. The isotropy would in a sense be generated by the immense
expansion; one small causal region would get blown up and our entire visible Universe would have been at one time in thermal contact. In addition, the parameter $k$ could have started out $O(1)$ and have been driven small by the expansion. The wavelengths of density perturbations would have been stretched by the expansion $\lambda \sim R$ making it appear that $\lambda \gg H^{-1}$ or that the perturbations have left the horizon. Rather, it is the size of the causally connected region that is no longer simply $H^{-1}$. However, not only does inflation offer an explanation for large scale perturbations, it also offers a source for the perturbations themselves [37]. Monopoles would also be diluted away.

The cosmological problems could be solved if

$$H \tau > 65$$  \hspace{1cm} (69)

where $\tau$ is the duration of the phase transition. In a successful theory, density perturbations are produced and do not exceed the limits imposed by the microwave background anisotropy, the vacuum energy density was converted to radiation so that the reheated temperature is sufficiently high, and baryogenesis is realized.

In the original (old) inflationary scenario, the phase transition determined by a potential with a large barrier separating the false and true vacua proceeds via the formation of bubbles [38]. The Universe reheats with the release of entropy which must occur through bubble collisions and the transition is completed when the bubbles fill up all of space. It is known [39], however, that the requirement for a long timescale $\tau$ is not compatible with the completion of the phase transition. The Universe as a whole remains trapped in the exponentially expanding phase containing only a few isolated bubbles of the broken $SU(3) \times SU(2) \times U(1)$ phase.

The well-known solution to the dilemma of old inflation is called the new inflationary scenario [40]. New inflation (it was at the time) was originally based on symmetry breaking using a flat potential of the Coleman–Weinberg form Ref. [41]. Instead of proceeding by tunnelling and the formation of bubbles, the transition takes place more or less uniformly on large scales. The details of the inflationary transition are determined from the equations of motion.

A Lagrangian for a scalar field which includes a scalar potential may be incorporated into the total action including gravity

$$I = \int d^4x \sqrt{g} \left( \frac{R}{2\kappa^2} - \frac{1}{2} \partial_{\mu} \phi \partial^{\mu} \phi - V(\phi) \right).$$ \hspace{1cm} (70)

The equation of motion for a scalar field $\phi$ can be derived from the energy-momentum tensor

$$T_{\mu\nu} = \partial_{\mu} \phi \partial_{\nu} \phi - \frac{1}{2} g_{\mu\nu} \partial_{\rho} \phi \partial^{\rho} \phi - g_{\mu\nu} V(\phi).$$ \hspace{1cm} (71)

By associating $\rho = T_{00}$ and $p = R^{-2}(t)T_{ii}$ we have

$$\rho = \frac{1}{2} \dot{\phi}^2 + \frac{1}{2} R^{-2}(t)(\nabla \phi)^2 + V(\phi),$$ \hspace{1cm} (72)

$$p = \frac{1}{2} \dot{\phi}^2 - \frac{1}{6} R^{-2}(t)(\nabla \phi)^2 - V(\phi),$$ \hspace{1cm} (73)

and from Eq. (6) we can write the equation of motion (by considering a homogeneous region, we can ignore the gradient terms)

$$\ddot{\phi} + 3H \dot{\phi} = -\partial V/\partial \phi.$$ \hspace{1cm} (74)

Consider now the approximation $\partial V/\partial \phi \sim (\partial^2 V/\partial \phi^2)\dot{\phi}$; the equation of motion becomes [42]

$$\ddot{\phi} + 3H \dot{\phi} + m^2(\phi)\dot{\phi} = 0,$$ \hspace{1cm} (75)
with \( m^2(\phi) = \partial^2 V/\partial \phi^2 < 0 \). The solution when \( |m^2| \gg H^2 \) grows exponentially as \( \phi \sim e^{m|t|} \) while for \( |m^2| \ll H^2 \) the scalar field grows as \( \phi \sim e^{m^2t/3H} \). In the latter case the field moves very slowly during a time period
\[
\tau \sim 3H/|m|^2. \tag{76}
\]
This approximation is known as the slow-rollover approximation.

If the scalar mass is tuned somewhat, \( m_\phi \sim 10^9 \text{ GeV} \), a significant amount of inflation is possible. From Eq. (76) one sees that
\[
H \tau \sim \frac{H^2}{|m|^2} \sim \frac{v^4}{M_P^2|m|^2} \sim 10^4, \tag{77}
\]
for \( v \sim 10^{15} \text{ GeV} \). Reheating no longer occurs via the collisions of bubbles, but by the decay of scalar field oscillations. As the scalar field settles to its minimum, the solution to the equations of motion look like
\[
\phi(t) \sim \frac{v}{mt} \sin mt \tag{78}
\]
and the reheat temperature is
\[
T_R \sim \left( \Gamma_D M_P \right)^{1/2} \quad \Gamma_D < H_I, \tag{79}
\]
where \( \Gamma_D \) is the scalar field decay rate and \( H_I \) is the value of \( H \) during inflation.

In addition to producing \( \Omega = 1 \), which is clearly seen from Eq. (61) as \( \hat{k} \to 0 \), new inflation is capable of producing scale invariant density perturbations [37] of the type preferred for galaxy formation models. However, the original [40] new inflationary models based on a Coleman–Weinberg [41] type of \( SU(5) \) breaking produced density fluctuations with magnitude \( \delta \rho/\rho \sim O(10^2) \) rather than \( \delta \rho/\rho \sim 10^{-5} \) as needed to remain consistent with microwave background anisotropies. Other more technical problems [43] concerning slow rollover and the effects of quantum fluctuations also pass doom on this original model.

General models of inflation can be described by a few so-called slow roll parameters \( \epsilon \) and \( \eta \). These are given by
\[
\epsilon = 3 \left( \frac{p}{\rho} + 1 \right) = \frac{4\pi}{M_P^2} \left( \frac{\dot{\phi}}{H} \right)^2 \approx \frac{M_P^2}{16\pi} \left( \frac{V''(\phi)}{V(\phi)} \right)^2, \tag{80}
\]
and
\[
\eta = -\frac{\dddot{\phi}}{H \dot{\phi}} \simeq \left( \frac{V''(\phi)}{V(\phi)} \right). \tag{81}
\]
For sufficient inflation both of these parameters must be small. The amount of inflation (given by the total number of e-foldings, \( N \)) is given by \( \epsilon \),
\[
N = -\int H dt = \int \frac{H}{\dot{\phi}} d\phi = \frac{2\sqrt{\pi}}{M_P} \int \frac{d\phi}{\sqrt{\epsilon}}. \tag{82}
\]
As noted above, inflation leads to a nearly scale-free spectrum of density fluctuations, \( \Delta(k) \propto \delta \rho/\rho \) with a power spectrum of the form \( \Delta^2(k) \propto k^{n-1} \). The spectral index is determined by the inflationary potential
\[
n \simeq 1 - 6\epsilon + 2\eta. \tag{83}\]
In addition to the scalar perturbations, tensor perturbations (gravity waves) will also be produced during inflation with spectral index \( n_g \simeq -2\epsilon \). The ratio of the amplitudes of the tensor to scalar perturbations is an important observable also given in terms of the slow roll parameter, \( \epsilon, \eta \simeq 16\epsilon \).

A generic model of inflation can be described by a potential of the form
\[
V(\phi) = \mu^4 P(\phi), \tag{84}
\]
Table 3: Slow roll parameters for chaotic inflationary models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon$</td>
<td>1/120</td>
<td>1/60</td>
</tr>
<tr>
<td>$\eta$</td>
<td>1/120</td>
<td>1/40</td>
</tr>
<tr>
<td>$n$</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>$r$</td>
<td>0.13</td>
<td>0.27</td>
</tr>
</tbody>
</table>

where $\phi$ is the scalar field driving inflation, the inflaton, $\mu$ is an as yet unspecified mass parameter, and $P(\phi)$ is a function of $\phi$ which possesses the features necessary for inflation, but contains no small parameters. That is, $P(\phi)$ takes the form

$$P(\phi) = P(0) + m_\phi^2 \phi^2 + \lambda \phi^4 + \ldots,$$

where all of the couplings in $P$ are $O(1)$ and $\ldots$ refers to possible non-renormalizable terms. Most of the useful inflationary potentials can be put into the form of Eq. (84).

The requirements for successful inflation boil down to: 1) enough inflation; and 2) density perturbations of the right magnitude. The latter reduces approximately to

$$\frac{\delta \rho}{\rho} \sim O(100) \frac{\mu^2}{M_P^2}.$$  

(86)

For large scale fluctuations of the type measured by COBE [14], we can use Eq. (86) to fix the inflationary scale $\mu$ of the inflaton potential [44]:

$$\frac{\mu^2}{M_P^2} = \text{few} \times 10^{-8}.$$  

(87)

Fixing $(\mu^2/M_P^2)$ has immediate general consequences for inflation [45]. For example, the Hubble parameter during inflation, $H^2 \sim (8\pi/3)(\mu^4/M_P^2)$ so that $H \sim 10^{-7} M_P$. The duration of inflation is $\tau \simeq M_P^2/\mu^4$, and the number of e-foldings of expansion is $H \tau \sim 8\pi(M_P^2/\mu^2) \sim 10^9$. If the inflaton decay rate goes as $\Gamma \sim m_\phi^3/\mu^2$, the Universe recovers at a temperature $T_R \sim (\Gamma M_P)^{1/2} \sim \mu^3/M_P^2 \sim 10^{-11} M_P \sim 10^9 \text{ GeV}.$

Two commonly studied potentials are associated with a form of inflation known as chaotic inflation [46]. In these models it is assumed that as part of an initially chaotic state $\phi > M_P$ with $V(\phi) \sim M_P^4$. Once these assumptions are made, chaotic models of inflation are by far the simplest. Typical models for chaotic inflation in terms of a single scalar field are described by the following scalar potentials [46, 47]

$$V = \frac{1}{4} \lambda \phi^4$$  

(88)

or

$$V = \frac{1}{2} m_\phi^2 \phi^2.$$  

(89)

That’s all! Nothing more complicated is necessary. It is assumed that at the Planck time, all fields $\phi(x)$ satisfy $V(\phi) < M_P^4$ and $(\partial_\mu \phi)^2 < M_P^4$. It is also assumed that there exist domains sufficiently large $l > H^{-1}$ with $\phi$ homogeneous and $\phi > M_P$.

For suitably large initial values of the scalar field $\phi$, the Universe expands quasiexponentially. Sufficient inflation requires only that $\phi_0 > \text{few} \times M_P$. Although this is not a strong constraint on chaotic models, a stronger constraint is derivable from the consideration of density perturbations. One finds that $\delta \rho/\rho \lesssim 10^{-5}$ for $\lambda < 10^{-14}$ or $m < 10^{-6}M_P$. The slow roll parameters are easily determined for these two models and are given in Table 3.

Finally, CMB measurements can be used to test inflationary models by determining or limiting the slow roll parameters. For example, WMAP [2] is able to set a constraint in the $r, n$ plane as shown. For
example, at $n = 0.95$, the 68% (95%) CL upper limit on $r$ is $< 0.07 (0.17)$, while at $n = 0.97$, $r < 0.18 (0.28)$. As one can see, while the $m^2 \phi^2$ model is well within the constraints, the $\lambda \phi^4$ model is not.

### 2.3 Baryogenesis

It appears that there is apparently very little antimatter in the Universe. To date, the only antimatter observed is the result of a high-energy collision, either in an accelerator or in a cosmic-ray collision in the atmosphere. There has been no sign to date of any primary antimatter, such as an anti-helium found in cosmic-rays. In addition, the number of photons greatly exceeds the number of baryons. Indeed, the value of $\Omega_{baryons}$ is

\[ \frac{n_B}{n_\gamma} \approx 274 \Omega_B h^2 = (6.23 \pm 0.17) \times 10^{-10}. \] (90)

In the Standard Model, the entropy density today is related to $n_\gamma$ by

\[ s = \frac{2\pi^4}{90\zeta(3)}(2 + \frac{4}{21})n_\gamma = 7.04n_\gamma, \] (91)

so that Eq. (90) implies $n_B/s \approx 8.8 \times 10^{-11}$. In the absence of baryon number violation or entropy production this ratio is conserved, however, and hence represents a potentially undesirable initial condition.

Let us for the moment assume that in fact $\eta = 0$. We can compute the final number density of nucleons left over after annihilations of baryons and antibaryons have frozen out. At very high temperatures (neglecting a quark–hadron transition) $T > 1$ GeV, nucleons were in thermal equilibrium with the photon background and $n_B = n_B = (3/2)n_\gamma$ (a factor of 2 accounts for neutrons and protons and the factor 3/4 for the difference between Fermi and Bose statistics). As the temperature fell below $m_N$ annihilations kept the nucleon density at its equilibrium value $\langle n_B/n_\gamma \rangle = \left( \frac{\pi^{1/2}(m_N/T)^{3/2}/2\zeta(3)}{\exp(-m_N/T)} \right)$ until the annihilation rate $\Gamma_A \approx n_B m_N^{-2}$ fell below the expansion rate. This occurred at $T \approx 20$ MeV. However, at this time the nucleon number density had already dropped to

\[ n_B/n_\gamma \approx 10^{-18}, \] (92)

which is eight orders of magnitude too small [48] aside from the problem of having to separate the baryons from the antibaryons. If any separation did occur at higher temperatures (so that annihilations were as yet incomplete) the maximum distance scale on which separation could occur is the causal scale related to the age of the Universe at that time. At $T = 20$ MeV, the age of the Universe was only $t = 2 \times 10^{-3}$ s. At that time, a causal region (with distance scale defined by $2ct$) could only have contained $10^{-5} M_\odot$ which is very far from the galactic mass scales, $10^{12} M_\odot$, we are asking for separations to occur.

#### 2.3.1 The out-of-equilibrium decay scenario

The production of a net baryon asymmetry requires baryon number violating interactions, C and CP violation and a departure from thermal equilibrium [49]. The first two of these ingredients are contained in GUTs, the third can be realized in an expanding Universe where, as we have seen, it is not uncommon that interactions come in and out of equilibrium. In SU(5), the fact that quarks and leptons are in the same multiplets allows for baryon non-conserving interactions such as $e^- + d \leftrightarrow \bar{u} + \bar{u}$, etc., or decays of the supermassive gauge bosons $X$ and $Y$ such as $X \rightarrow e^- + d, \bar{u} + \bar{u}$. Although today these interactions are very ineffective because of the very large masses of the $X$ and $Y$ bosons, in the early Universe when $T \sim M_X \sim 10^{15}$ GeV these types of interactions should have been very important. C and CP violation is very model dependent. In the minimal SU(5) model, as we will see, the magnitude of C and CP violation

18
is too small to yield a useful value of $\eta$. The C and CP violation in general comes from the interference between tree level and first loop corrections.

The departure from equilibrium is very common in the early Universe when interaction rates cannot keep up with the expansion rate. In fact, the simplest (and most useful) scenario for baryon production makes use of the fact that a single decay rate goes out of equilibrium. It is commonly referred to as the out-of-equilibrium decay scenario [50]. The basic idea is that the gauge bosons $X$ and $Y$ (or Higgs bosons) may have a lifetime long enough to insure that the inverse decays have already ceased so that the baryon number is produced by their free decays.

More specifically, let us call $X$, either the gauge boson or Higgs boson which produces the baryon asymmetry through decays. Let $\alpha$ be its coupling to fermions. For $X$ a gauge boson, $\alpha$ will be the GUT fine structure constant, while for $X$ a Higgs boson, $(4\pi\alpha)^{1/2}$ will be the Yukawa coupling to fermions. The decay rate for $X$ will be

$$\Gamma_X \simeq \alpha M_X.$$  \hspace{1cm} (93)

However, decays can only begin occurring when the age of the Universe is longer than the $X$ lifetime $\Gamma_X^{-1}$, i.e., when $\Gamma_X > H$

$$\alpha M_X \gtrsim N(T)^{1/2} T^2 / M_P,$$  \hspace{1cm} (94)

or at a temperature

$$T^2 \lesssim \alpha M_X M_P N(T)^{-1/2}.$$  \hspace{1cm} (95)

Scatterings, on the other hand, proceed at a rate $\Gamma_S \sim \alpha^2 T^3 / M_X^2$ and hence are not effective at lower temperatures. To be in equilibrium, decays must have been effective as $T$ fell below $M_X$ in order to track the equilibrium density of $X$’s (and $\bar{X}$’s). Therefore, the out-of-equilibrium condition is that at $T = M_X, \Gamma_D < H$

$$M_X \gtrsim \alpha M_P (N(M_X))^{-1/2} \sim 10^{18} \alpha \text{ GeV}.$$  \hspace{1cm} (96)

In this case, we would expect a maximal net baryon asymmetry to be produced.

To see the role of C and CP violation, consider the two channels for the decay of an $X$ gauge boson: $X \to (1)\bar{u}u, (2)e^-d$. Suppose that the branching ratio into the first channel with baryon number $B = -2/3$ is $r$ and that of the second channel with baryon number $B = +1/3$ is $1 - r$. Suppose in addition that the branching ratio for $\bar{X}$ into $(\bar{1})uu$ with baryon number $B = +2/3$ is $\bar{r}$ and into $(\bar{2})e^+\bar{d}$ with baryon number $B = -1/3$ is $1 - \bar{r}$. Though the total decay rates of $X$ and $\bar{X}$ (normalized to unity) are equal as required by CPT invariance, the differences in the individual branching ratios signify a violation of C and CP conservation.

Denote the parity (P) of the states (1) and (2) by $\uparrow$ or $\downarrow$, then we have the following transformation properties:

Under CPT : $\Gamma(X \to 1 \uparrow) = \Gamma(\bar{1} \downarrow \to \bar{X})$

Under CP : $\Gamma(X \to 1 \uparrow) = \Gamma(X \to \bar{1} \downarrow)$

Under C : $\Gamma(X \to 1 \uparrow) = \Gamma(\bar{X} \to \bar{1} \uparrow)$. \hspace{1cm} (97)

We can now denote

$$r = \Gamma(\bar{X} \to \bar{1} \uparrow) + \Gamma(\bar{X} \to \bar{1} \downarrow)$$  \hspace{1cm} (98)

$$\bar{r} = \Gamma(\bar{X} \to \bar{1} \uparrow) + \Gamma(\bar{X} \to \bar{1} \downarrow).$$  \hspace{1cm} (99)

The total baryon number produced by an $X, \bar{X}$ decay is then

$$\Delta B = \frac{2}{3} r + \frac{1}{3} (1 - r) + \frac{2}{3} \bar{r} - \frac{1}{3} (1 - \bar{r})$$

$$\Delta B = \bar{r} - r = \Gamma(\bar{X} \to \bar{1} \uparrow) + \Gamma(\bar{X} \to \bar{1} \downarrow) - \Gamma(X \to 1 \uparrow) - \Gamma(X \to 1 \downarrow).$$  \hspace{1cm} (100)

One sees clearly therefore, that from Eqs. (97) if either C or CP are good symmetries, $\Delta B = 0$.

19
The schematic view presented above can be extended to a complete calculation given a specific model [51, 52], see also Ref. [53] for reviews.

The time evolution for the generation of a baryon asymmetry is shown in Fig. 6. As one can see, for large values of $M_X$, i.e., values which satisfy the lower limit given in Eq. (96), the maximal value for the baryon asymmetry $n_B/s \sim 10^{-2} \epsilon$ is achieved. This confirms numerically the original out-of-equilibrium decay scenario [50]. For smaller values of $M_X$ an asymmetry is still produced which, however, is smaller due to partial equilibrium maintained by inverse decays. The growth of the asymmetry as a function of time is now damped, and it reaches its final value when inverse decays freeze out. Finally, by studying different initial conditions, one can show that the result for the final baryon asymmetry is in fact largely independent of the initial baryon asymmetry.

From Eq. (101) it is clear that a complete calculation of $n_B/s$ will require a calculation of the CP violation in the decays (summed over parities) which we can parametrize by

$$\epsilon = \bar{\epsilon} - \epsilon = \frac{\Gamma(\bar{X} \to \bar{1}) - \Gamma(X \to 1)}{\Gamma(X \to 1) + \Gamma(X \to 1)} \sim \frac{\text{Im} \Gamma}{\text{Re} \Gamma}.$$  

At the tree level, as one can see, $\Gamma(X \to 1) \propto g^4_{55}$ is real and there is no C or CP violation. At the one loop level for gauge boson decay, there is also no net contribution to $\epsilon$, and we must turn to Higgs decay.
At least two Higgs five-plets are required to generate sufficient C and CP violation. With two five-plets, $H$ and $H'$, the interference of diagrams of the type in Fig. 7 will yield a non-vanishing $\epsilon$ [54],

$$
\epsilon \propto \text{Im}(a'^\dagger ab'^\dagger b) \neq 0,
$$

if the couplings $a \neq a'$ and $b \neq b'$.

There are of course many alternative methods to generate the baryon asymmetry, though each makes use of the same three ingredients. For example, a supersymmetric mechanism proposed by Affleck and Dine [55] makes use of flat directions in the scalar potential. There are many such flat directions, and some of these yield an non-vanishing expectation value to GUT baryon number violating operators. Supersymmetry breaking perturbs the flatness, leading to the cosmological evolution of the scalar which oscillates about the global (charge and colour conserving) minimum of the potential. Baryon number is stored in these oscillations and a net asymmetry is produced as these fields decay.

Another mechanism to generate the baryon asymmetry employs the heavy right-handed neutrinos used in the see-saw mechanism to generate neutrino masses [56]. The simplest of such mechanisms is based on the decay of a right-handed neutrino-like state [57]. This mechanism is certainly novel in that it does not require grand unification at all. By simply adding to the Lagrangian a Dirac and Majorana mass term for a new right-handed neutrino state,

$$
L \ni M \nu^c \nu^c + \lambda H L \nu^c,
$$

the out-of-equilibrium decays $\nu^c \rightarrow L + H^*$ and $\nu^c \rightarrow L^* + H$ will generate a non-zero lepton number $L \neq 0$. The out-out-equilibrium condition for these decays translates to $10^{-3} \lambda^2 M_P < M$ and $M$ could be as low as $O(10)$ TeV. (Note that once again in order to have a non-vanishing contribution to the C and CP violation in this process at 1-loop, at least two flavours of $\nu^c$ are required. For the generation of masses of all three neutrino flavors, three flavours of $\nu^c$ are required.) Electroweak sphaleron effects [58] can transfer this lepton asymmetry into a baryon asymmetry. If sphalerons are in equilibrium, the baryon number can be expressed in terms of $B - L$

$$
B = \frac{28}{79} (B - L).
$$

In the absence of a primordial $B - L$ asymmetry, the baryon number is erased by equilibrium processes. Right-handed neutrinos produce a net lepton asymmetry and hence a net $B - L$ yielding the final baryon asymmetry given by Eq. (105).

### 3 Lecture 3: Big Bang Nucleosynthesis

The standard model [59] of Big Bang Nucleosynthesis (BBN) is based on the relatively simple idea of including an extended nuclear network into a homogeneous and isotropic cosmology. Apart from the
input nuclear cross sections, the theory contains only a single parameter, namely the baryon-to-photon ratio, $\eta$, and even that has been fixed by WMAP [2]. The theory then allows one to make predictions (with well-defined uncertainties) of the abundances of the light elements, D, $^3$He, $^4$He, and $^7$Li [60].

Conditions for the synthesis of the light elements were attained in the early Universe at temperatures $T \gtrsim 1$ MeV. In the early Universe, the energy density was dominated by radiation with

$$\rho = \frac{\pi^2}{30} (2 + \frac{7}{2} + \frac{7}{4} N_\nu) T^4,$$

from the contributions of photons, electrons and positrons, and $N_\nu$ neutrino flavours (at higher temperatures, other particle degrees of freedom should be included as well). At these temperatures, weak interaction rates were in equilibrium. In particular, the processes

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

(107)

fix the ratio of number densities of neutrons to protons. At $T \gg 1$ MeV, $(n/p) \approx 1$.

As we have seen in the case of neutrino interactions, the weak interactions do not remain in equilibrium at lower temperatures. Freeze-out occurs when the weak interaction rate $\Gamma_{wk} \sim G_F^2 T^5$ falls below the expansion rate which is given by the Hubble parameter $H \sim T^2/M_P$. The $\beta$-interactions in Eq. (107) freeze out at about 0.8 MeV. As the temperature falls and approaches the point where the weak interaction rates are no longer fast enough to maintain equilibrium, the neutron-to-proton ratio is given approximately by the Boltzmann factor,

$$\left(\frac{n}{p}\right) \approx e^{-\Delta m/T} \sim 1/5,$$

where $\Delta m$ is the neutron–proton mass difference. After freeze-out, free neutron decays drop the ratio slightly to about $1/7$ before nucleosynthesis begins. A useful semi-analytic description of freeze-out has been given [61, 62].

The nucleosynthesis chain begins with the formation of deuterium by the process,

$$p + n \rightarrow D + \gamma.$$

However, because of the large number of photons relative to nucleons, $\eta^{-1} = n_\gamma/n_B \sim 10^{10}$, deuterium production is delayed past the point where the weak interaction rates are no longer fast enough to maintain equilibrium, the deuterium binding energy, $E_B = 2.2$ MeV (the average photon energy in a blackbody is $\bar{E}_\gamma \approx 2.7 T$). This is because there are many photons in the exponential tail of the photon energy distribution with energies $E > E_B$ despite the fact that the temperature or $\bar{E}_\gamma$ is less than $E_B$. The degree to which deuterium production is delayed can be found by comparing the qualitative expressions for the deuterium production and destruction rates,

$$\Gamma_p \approx n_B \sigma v$$
$$\Gamma_d \approx n_\gamma \sigma v e^{-E_B/T}.$$ 

When the quantity $\eta^{-1} \exp(-E_B/T) \sim 1$, the rate for deuterium destruction $(D + \gamma \rightarrow p + n)$ finally falls below the deuterium production rate and the nuclear chain begins at a temperature $T \sim 0.1$ MeV.

The dominant product of Big Bang nucleosynthesis is $^4$He and its abundance is very sensitive to the $(n/p)$ ratio

$$Y_p = \frac{2(n/p)}{[1 + (n/p)]} \approx 0.25,$$

(109)
i.e., an abundance of close to 25% by mass. Lesser amounts of the other light elements are produced: D and $^3$He at the level of about $10^{-5}$ by number, and $^7$Li at the level of $10^{-10}$ by number. The gap at $A = 8$ prevents the production of other isotopes in any significant quantity. The nuclear chain is shown in Fig. 8.

Historically, BBN as a theory explaining the observed element abundances was nearly abandoned due its inability to explain all element abundances. Subsequently, stellar nucleosynthesis became the leading theory for element production [63]. However, two key questions persisted. 1) The abundance of
$^4\text{He}$ as a function of metallicity is nearly flat and no abundances are observed to be below about 23% as exaggerated in Fig. 9. In particular, even in systems in which an element such as oxygen which traces stellar activity is observed at extremely low values (compared with the solar value of $\text{O/H} \approx 4.9 \times 10^{-4}$), the $^4\text{He}$ abundance is nearly constant. This is very different from all other element abundances (with the exception of $^7\text{Li}$ as we shall see below). For example, in Fig. 10, the N/H vs. O/H correlation is shown [64]. As one can clearly see, the abundance of N/H goes to 0, as O/H goes to 0, indicating a stellar source for nitrogen. 2) Stellar sources cannot produce the observed abundance of D/H. Indeed, stars destroy deuterium and no astrophysical site is known for the production of significant amounts of deuterium [65]. Thus we are led back to BBN for the origins of D, $^3\text{He}$, $^4\text{He}$, and $^7\text{Li}$.

### 3.1 Abundance predictions

Because standard BBN theory rests upon the Standard Model of particle physics, the electroweak aspects of the calculation are very well-determined and do not introduce an appreciable uncertainty. Instead, the major uncertainties come from the thermonuclear reaction rates. There are 11 key strong rates (as well as the neutron lifetime) which dominate the uncertainty budget [66–69]. In contrast to the situation for much of stellar nucleosynthesis, BBN occurs at high enough temperatures that laboratory data exist at and even below the relevant energies, so that no extrapolation is needed. Monte Carlo techniques [66,67] are used to determine the best-fit abundances, and their uncertainties, at each $\eta$.

Recently the input nuclear data have been carefully reassessed [66,68–71], leading to improved
precision in the abundance predictions. In addition, polynomial fits to the predicted abundances and the error correlation matrix have been given [72, 73]. The NACRE Collaboration presented an updated nuclear compilation [70]. For example, notable improvements include a reduction in the uncertainty in the rate for $^3\text{He}(n, p)T$ from 10% [74] to 3.5% and for $T(\alpha, \gamma)^7\text{Li}$ from $\sim 23$–$30\%$ [74] to $\sim 4\%$. Since then, new data and techniques have become available, motivating new compilations. Within the last year, several new BBN compilations have been presented [73, 75–77].

The light element abundances are shown in Fig. 11 as a function of $\eta$ [77]. The plot shows the abundance of $^4\text{He}$ by mass $Y$ and the abundances of the other three isotopes by number. The curves indicate the central predictions from BBN, while the bands correspond to the uncertainty in the predicted abundances. The uncertainty range in $^4\text{He}$ reflects primarily the $1\sigma$ uncertainty in the neutron lifetime.

In standard BBN with $N_\nu = 3$, the only free parameter is the density of baryons (strictly speaking, nucleons), which sets the rates of the strong reactions. Because standard BBN is a one-parameter theory, any abundance measurement determines $\eta$, while additional measurements overconstrain the theory and thereby provide a consistency check. BBN has thus historically been the premier means of determining the cosmic baryon density.
Fig. 11: The predictions of standard BBN [77] showing the primordial abundances as a function of the baryon-to-photon ratio $\eta$. Abundances are quantified as ratios to hydrogen, except for $^4$He which is given in baryonic mass fraction $Y_p = \rho_{^4\text{He}}/\rho_B$. The lines give the mean values, and the surrounding bands give the $1\sigma$ uncertainties.

The release of the first-year WMAP results on the anisotropy spectrum of the CMB were a landmark event for all of cosmology, but particularly for BBN. As discussed above, the value of $\eta$ has been fixed by CMB measurements as given by Eq. (90). Thus, within the context of the Standard Model, BBN becomes a zero-parameter theory, and the light element predictions are completely determined to within the uncertainties in $\eta$ and the BBN theoretical errors. Comparison with light element observations then can be used to restate the test of BBN–CMB consistency, or to turn the problem around and test the astrophysics of post-BBN light element evolution [78].

3.2 Light element observations and comparison with theory

BBN theory predicts the abundances of D, $^3$He, $^4$He, and $^7$Li, which are essentially determined at $t \sim 180$ s. Abundances are, however, observed at much later epochs, after stellar nucleosynthesis has commenced. The ejecta from stellar processing can alter the light element abundances from their primordial values, but also produce heavy elements such as C, N, O, and Fe (‘metals’). Thus one seeks astrophysical sites with low metal abundances, in order to measure light element abundances which are
closer to primordial. For all of the light elements, systematic errors are an important and often dominant limitation to the precision of the primordial abundances.

### 3.2.1 D/H

In recent years, high-resolution spectra have revealed the presence of D in high-redshift, low-metallicity quasar absorption systems (QAS), via its isotope-shifted Lyman-α absorption. These are the first measurements of light element abundances at cosmological distances. It is believed that there are no astrophysical sources of deuterium [65], so any measurement of D/H provides a lower limit to primordial D/H and thus an upper limit on \( \eta \). Recent observations by FUSE show a wide dispersion in the deuterium abundance in local gas seen via its absorption, \((D/H)_{\text{local gas}} = (0.5 - 2.2) \times 10^{-5}\) [79]. This surprisingly large spread, taken together with the positive correlation of D/H with temperature and metallicity along various sightlines, led [79] to suggestions that deuterium may suffer significant and preferential depletion onto dust grains. In this case the true local interstellar D/H value would lie at the upper limit of the observed values, giving \((D/H)_{\text{ISM}} \gtrsim (2.31 \pm 0.24) \times 10^{-5}\). However, extracting a primordial deuterium value requires a Galactic chemical evolution model (e.g., Ref. [80]), whose model dependences yield uncertainties in the determination of the primordial deuterium abundance. Many of these models do not predict significant D/H depletion [81] at high redshift, and in this case the high-redshift measurements are expected to recover the primordial deuterium abundance.

The deuterium abundance at low metallicity has been measured in several quasar absorption systems [82]. The weighted mean value of the seven systems with reliable abundance determinations is \( \log (D/H)_{\text{local gas}} = -4.55 \pm 0.03 \) where the error includes a scale factor of 1.72 and corresponds to \( D/H = (2.82 \pm 0.21) \times 10^{-5} \). These are shown in Fig. 12. Since the D/H shows considerable scatter it is likely that systematic errors dominate the uncertainties. In this case it may be more appropriate to derive the uncertainty using sample variance (see, for example, Ref. [66]) which gives a more conservative range \( \log (D/H) = -4.55 \pm 0.08 \) or \( D/H = (2.82 \pm 0.53) \times 10^{-5} \).

Using the WMAP value for the baryon density (90), the primordial D/H abundance is predicted to be [77]

\[
(D/H)_p = (2.49 \pm 0.17) \times 10^{-5}.
\]

As one can see from Fig. 11, this is in good agreement with the average of the seven best determined quasar absorption system abundances noted above, particularly when systematic uncertainties are taken into account.

### 3.2.2 \(^4\text{He}\)

We observe \(^4\text{He}\) in clouds of ionized hydrogen (H II regions), the most metal-poor of which are in dwarf galaxies. There is now a large body of data on \(^4\text{He}\) and CNO in these systems [83]. These data confirm that the small stellar contribution to helium is positively correlated with metal production. Recently a careful study of the systematic uncertainties in \(^4\text{He}\), particularly the role of underlying absorption, has led to a higher value for the primordial abundance of \(^4\text{He}\) [84]. Using a subset of the highest quality from the data of Izotov and Thuan [83], all of the physical parameters listed above including the \(^4\text{He}\) abundance were determined self-consistently with Monte Carlo methods [85]. The extrapolated \(^4\text{He}\) abundance was determined to be \( Y_p = 0.249 \pm 0.009 \). Conservatively, it would be difficult at this time to exclude any value of \( Y_p \) inside the range 0.232–0.258.

At the WMAP value for \( \eta \), the \(^4\text{He}\) abundance is predicted to be [77]

\[
Y_p = 0.2486 \pm 0.0002.
\]

This is in excellent agreement with the most recent analysis of the \(^4\text{He}\) abundance [84]. Note also that the large uncertainty ascribed to this value indicates that while \(^4\text{He}\) is certainly consistent with the WMAP
3.2.3 \(^7\)Li

The systems best suited for Li observations are metal-poor stars in our Galaxy. Observations have long shown [86] that Li does not vary significantly in Pop II stars with metallicities \(< 1/30\) of solar — the ‘Spite plateau’. Precision data suggest a small but significant correlation between Li and Fe [87] which can be understood as the result of Li production from Galactic cosmic rays [88]. Extrapolating to zero metallicity one arrives at a primordial value [89] \(\text{Li}/\text{H} \mid_p = (1.23 \pm 0.34) \times 10^{-10}\).

Figure 13 shows the different Li components for a model with \((\text{Li}/\text{H})_p = 1.23 \times 10^{-10}\) as a function of the iron abundance expressed as the log of Fe/H relative to the solar value. The linear slope produced by the model is independent of the input primordial value. The model [90] includes, in addition to primordial \(^7\)Li, lithium produced in Galactic cosmic-ray nucleosynthesis (primarily \(\alpha + \alpha\) fusion), and \(^7\)Li produced by the \(\nu\)-process during type II supernovae. As one can see, these processes are not sufficient to reproduce the population I abundance of \(^7\)Li (at near solar \([\text{Fe/H}] \lesssim 0\)), and additional production sources are needed.

---

**Fig. 12:** D/H abundances shown as a function of [Si/H]. Labels denote the background quasi-stellar objects (QSO), except for the local interstellar value (LISM; [79]).

determination of the baryon density, it does not provide for a highly discriminatory test of the theory at this time.
Fig. 13: Contributions to the total predicted lithium abundance from the adopted Galactic chemical evolution model of Ref. [90], compared with low metallicity stars and a sample of high metallicity stars. The solid curve is the sum of all components.

A recent reanalysis of the $^3\text{He}(\alpha, \gamma)^7\text{Be}$ reaction, which is the most important $^7\text{Li}$ production process in BBN, was considered in detail in Ref. [91]. When the new rate is used a high $^7\text{Li}$ abundance is found [77] at the WMAP value of $\eta_{^7\text{Li}}/\text{H} = (5.24^{+0.71}_{-0.62}) \times 10^{-10}$. This represents a 23% increase in $^7\text{Li}$ over previous calculations [73]. The increase is primarily due to an increase in the $^3\text{He}(\alpha, \gamma)^7\text{Be}$ cross section. Newer data [91] implies 17% increase in this reaction leading to a 16% increase in $^7\text{Li}$. In addition, the 1.5% increase in $\eta$ from the 3-year to 5-year WMAP data [2] leads to a 3% increase in $^7\text{Li}$ and finally another 1% increase is due to updated $pn$ rates. In addition, the uncertainty in the BBN $^7\text{Li}$ abundance is roughly a factor of 2 times smaller than previous determinations. This value for primordial $^7\text{Li}$ is in clear contradiction with most estimates of the primordial Li abundance. Several attempts at explaining this discrepancy by adjusting some of the key nuclear rates have proved unsuccessful [71, 92, 93].

An important source for systematic error lies in the derived effective temperature of the star. $[\text{Li}] = \log(^7\text{Li}/\text{H}) + 12$ is very sensitive to the temperature, with $\partial[\text{Li}] / \partial T_{\text{eff}} \simeq 0.065$ – 0.08. Unfortunately there is no standard for determining effective temperatures, and, for a given star, there is considerable range depending on the method used. This spread in temperatures was made manifest in the recent work of Melendez and Ramirez [94] using the infra-red flux method (IRFM) which showed differences for very low metallicities ([Fe/H] < -3) by as much as 500 K, with typical differences of ~ 200 K with respect to that of Ref. [87]. As a consequence the derived $^7\text{Li}$ abundance was significantly higher with $^7\text{Li}/\text{H}_{p} = (2.34 \pm 0.32) \times 10^{-10}$ [94, 95].

Recently a dedicated set of observations were performed with the specific goal of determining the effective temperature in metal-poor stars [96]. Using a large set of Fe I excitation lines (~ 100 lines per star), the Boltzmann equation was used with the excitation energies, $\chi_i$, to determine the temperature through the distribution of excited levels. Again, there was no evidence for the high temperatures reported in Ref. [94], rather, temperatures were found to be consistent with previous determinations. The mean $^7\text{Li}$ abundance found in Ref. [96] was $^7\text{Li}/\text{H} = (1.3 - 1.4 \pm 0.2) \times 10^{-10}$, consistent with the bulk of prior abundance determinations.
There are of course other possible sources of systematic uncertainty in the $^7\text{Li}$ abundance. It is possible that some of the surface $^7\text{Li}$ has been depleted if the outer layers of the stars have been transported deep enough into the interior, and/or mixed with material from the hot interior; this may occur due to convection, rotational mixing, or diffusion. Estimates for possible depletion factors are in the range $\sim 0.2–0.4$ dex, i.e., factors of $1.6–2.5$ [97]. Recent attempts to deplete the $^7\text{Li}$ abundance through diffusion introduce a source of turbulence tuned to fit the abundances of heavy elements in NGC6397 [98]. Once parameters are set, the degree of lithium depletion becomes a prediction of the model. For this cluster, a depletion factor of $0.25$ dex is found, i.e., a factor of $1.8$. Note that while this depletion factor would bring the previous BBN result of $^7\text{Li}/\text{H} = 4.26 \times 10^{-10}$ [73] to a value close the value observed in that cluster [99] ($2.2 \times 10^{-10}$), a larger depletion factor is needed with the new BBN value for $^7\text{Li}$ given above. It is also not clear whether this mechanism will work for the wide range of stellar parameters seen in the field. As noted above, the Li data show a negligible intrinsic spread in Li. Any mechanism which reduces significantly the abundance of $^7\text{Li}$ must do so uniformly over a wide range of stellar parameters (temperature, surface gravity, metallicity, rotational velocity, etc.).

It is also possible that the lithium discrepancy is a sign of new physics beyond the Standard Model. One possibility is the cosmological variation of the fine structure constant. Varying $\alpha$ would induce a variation in the deuterium binding energy and could yield a decrease in the predicted abundance of $^7\text{Li}$ [100]. A potential solution to both lithium problems is particle decay after BBN which could lower the $^7\text{Li}$ abundance (and produce some $^6\text{Li}$ as well) [101]. This has been investigated in the framework of the constrained minimal supersymmetric Standard Model if the lightest supersymmetric particle is assumed to be the gravitino [102] and indeed, some models have been found which accomplish these goals [103].

3.2.4 $^3\text{He}$

Since $^3\text{He}$ is also quite sensitive to the baryon density, one might hope that it too could be used as a baryometer. Observations of H II regions in our own Galaxy yield values of the $^3\text{He}/\text{H}$ ratio that are compatible with calculations of the primordial value [104, 105]. However, the extrapolation from the observations to a primordial abundance is complicated by the unknown chemical evolution of $^3\text{He}$. Indeed, one does not even know whether $^3\text{He}/\text{H}$ is increasing or decreasing with cosmic time. Thus a primordial extrapolation yields only an order-of-magnitude range of allowable values of $^3\text{He}/\text{H}$ [106].

3.3 Beyond the Standard Model

Given the simple physics underlying BBN, it is remarkable that it still provides one of the most effective tests for physics beyond the Standard Model. Limits on particle physics beyond the Standard Model come mainly from the observational bounds on the $^4\text{He}$ abundance. As discussed earlier, the neutron-to-proton ratio is fixed by its equilibrium value at the freeze-out of the weak interaction rates at a temperature $T_f \sim 1$ MeV modulo the occasional free neutron decay. Furthermore, freeze-out is determined by the competition between the weak interaction rates and the expansion rate of the Universe

$$G_F^2 T_f^5 \sim \Gamma_{\text{weak}}(T_f) = H(T_f) \sim \sqrt{G_N NT_f^2}.$$  \hspace{1cm} (113)

In the Standard Model, the number of relativistic particle species at 1 MeV is $N = \frac{1}{2} + \frac{7}{2} N_\nu$. The presence of additional neutrino flavours (or any other relativistic species) at the time of nucleosynthesis increases the overall energy density of the Universe and hence the expansion rate leading to a larger value of $T_f$, $(n/p)$, and ultimately $Y_p$. Because of the form of Eq. (113) it is clear that just as one can place limits [107] on $N_\nu$, any changes in the weak or gravitational coupling constants can be similarly constrained (for a discussion see Ref. [108]).

The helium curve in Fig. 11 was computed taking $N_\nu = 3$; the computed abundance scales as $\Delta Y_{\text{BBN}} \simeq 0.013 \Delta N_\nu$ [61]. The dependence of the light element abundances on $N_\nu$ is shown in
Fig. 14: BBN abundance predictions [78] as a function of the baryon-to-photon ratio $\eta$, for $N_\nu = 2$–7. The bands show the 1$\sigma$ error bars. Note that for the isotopes other than Li, the error bands are comparable in width to the thickness of the abundance curve shown. All bands are centred on $N_\nu = 3$.

Fig. 14 [78]. For a fixed value of $\eta = (6.14 \pm 0.25) \times 10^{-10}$ (slightly below the current WMAP value) and $Y_p = 0.249 \pm 0.009$, the likelihood distribution for $N_\nu$ is shown by the shaded region in Fig. 15 [109]. Also shown for comparison are the likelihood distribution based the WMAP value of $\eta$ using D/H alone, $Y_p$ and D/H, and the result based on BBN alone. Despite the increased uncertainty in the He abundance, it still provides the strongest constraint on $N_\nu$. D/H is nonetheless becoming competitive in its ability to set limits on $N_\nu$.

The 95% CL upper limits to $N_\nu$ are given in Table 4. In all cases the preferred values for $N_\nu$ are consistent with $N_\nu$, and in many cases are much closer to $N_\nu$ than 1$\sigma$. This restates the overall consistency among standard BBN theory, D and $^3$He observations, and CMB anisotropies. It also constrains departures from this scenario. The combined limit using BBN + light elements + CMB limit is [109]

$$2.67 \leq N_\nu \leq 3.85$$

at 68% CL.

4 Lecture 4: Dark Matter

Evidence for dark matter in the Universe is available from a wide range of observational data. As discussed many times above, the analysis of the cosmic microwave background anisotropies leads to the conclusion that the curvature of the Universe is close to zero indicating that the sum of the fractions of critical density, $\Omega$, in matter and a cosmological constant (or dark energy) is very close to one [2]. When combined with a variety of data including results from the analysis of type Ia supernovae observations [26,27] and baryon acoustic oscillations [28] one is led to the concordance model where $\Omega_m \sim 0.23$ and $\Omega_\Lambda \sim 0.73$ with the remainder (leading to $\Omega_{tot} = 1$) in baryonic matter. This is in addition to the
Fig. 15: The likelihood distribution for $N_\nu$ based on the WMAP value of $\eta$ and $Y_p$ (shaded), WMAP and D/H (dashed), WMAP and both $Y_p$ and D/H $A$ (dotted). Also shown is the result without imposing the WMAP value for $\eta$ (long dashed).

Table 4: This table shows constraints placed on $N_\nu$ and $\eta$ by various combinations of observations. Shown are the 68% confidence limits determined by marginalizing the 2-D likelihood distribution. Also shown are the 95% upper limits on $\delta N_\nu = N_\nu - 3$, given that $\delta N_\nu > 0$.

<table>
<thead>
<tr>
<th>Observations</th>
<th>$\eta_{10} \equiv 10^{10}\eta$</th>
<th>$N_\nu$</th>
<th>$\delta N_{\nu,max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_p + D/H_A$</td>
<td>$5.94^{+0.56}_{-0.50}$</td>
<td>$3.14^{+0.70}_{-0.65}$</td>
<td>1.59</td>
</tr>
<tr>
<td>$Y_p + \eta_{CMB}$</td>
<td>$6.14 \pm 0.25$</td>
<td>$3.08^{+0.74}_{-0.68}$</td>
<td>1.63</td>
</tr>
<tr>
<td>$D/H_A + \eta_{CMB}$</td>
<td>$6.16 \pm 0.25$</td>
<td>$3.59^{+1.14}_{-1.04}$</td>
<td>2.78</td>
</tr>
<tr>
<td>$Y_p + D/H_A + \eta_{CMB}$</td>
<td>$6.10^{+0.24}_{-0.22}$</td>
<td>$3.24^{+0.61}_{-0.57}$</td>
<td>1.44</td>
</tr>
</tbody>
</table>

The classic evidence from galactic rotation curves [110], which indicate that nearly all spiral galaxies are embedded in a large galactic halo of dark matter leading to rather constant rotational velocities at large distances from the centre of the galaxy (in contrast to the expected $v^2 \sim 1/r$ behaviour in the absence of dark matter). Other dramatic pieces of evidence can be found in combinations of X-ray observations and weak lensing showing the superposition of dark matter (from lensing) and ordinary matter from X-ray gas [111] and from the separation of baryonic and dark matter after the collision of two galaxies as seen in the Bullet cluster [112]. For a more complete discussion see Ref. [113].

From the first column of Table 2, we can obtain the density of cold dark matter from the difference between the total matter density and the baryon density [2]

$$\Omega_{CDM} h^2 = 0.1099 \pm 0.0062$$  \hspace{1cm} (115)$$

or a $2\sigma$ range of 0.0975–0.1223 for $\Omega_{CDM} h^2$. 

31
4.1 Neutrinos

Dark matter must be both long-lived or stable and electrically and colour neutral. As a result, once baryons and neutrinos are eliminated as candidates, one must look beyond the Standard Model. From WMAP, we already know that the baryon density is far below the requisite amount in cold dark matter. Light neutrinos \( (m \lesssim 30 \text{ eV}) \) are a long-time standard when it comes to non-baryonic dark matter [114]. Light neutrinos are, however, ruled out as a dominant form of dark matter because they produce too much large scale structure [115]. The energy density of light neutrinos with \( m_\nu \lesssim 1 \text{ MeV} \) can be expressed at late times as \( \rho_\nu = \frac{3}{11} m_\nu n_\gamma \). Imposing the constraint \( \Omega_\nu h^2 < 0.12 \), translates into a strong constraint (upper bound) on Majorana neutrino masses [116]:

\[
m_{\text{tot}} = \sum m_\nu < 11 \text{ eV,}
\]

(116)

where the sum runs over neutrino mass eigenstates. The limit for Dirac neutrinos depends on the interactions of the right-handed states. The limit (116) and the corresponding initial rise in \( \Omega_\nu h^2 \) as a function of \( m_\nu \) is displayed in Fig. 16. Much stronger limits on the sum of neutrino masses are possible when combining the WMAP data with large scale structure surveys. A typical limit is \( m_{\text{tot}} < 0.7 \text{ eV} \) or \( \Omega_\nu h^2 < 0.0076 \) [2].

The calculation of the relic density for neutrinos more massive than \( \sim 1 \text{ MeV} \), is substantially more involved. The relic density is now determined by the freeze-out of neutrino annihilations which occur at \( T \lesssim m_\nu \), after annihilations have begun to seriously reduce their number density [118]. For particles which annihilate through approximate weak scale interactions, annihilations freeze out when \( T \sim m_\chi/20 \).

Based on the leptonic and invisible width of the \( Z \) boson, experiments at LEP have determined that the number of neutrinos is \( N_\nu = 2.994 \pm 0.012 \) [119]. Thus, LEP excludes additional neutrinos (with
standard weak interactions) with masses \( m_\nu \lesssim 45 \text{ GeV}. \) The mass density of ordinary heavy neutrinos is found to be very small, \( \Omega_\nu h^2 < 0.001 \) for masses \( m_\nu > 45 \text{ GeV} \) up to \( m_\nu \sim \mathcal{O}(100) \text{ TeV} \) [118]. Laboratory constraints for Dirac neutrinos are available [120], excluding neutrinos with masses between 10 GeV and 4.7 TeV. This is significant, since it precludes the possibility of neutrino dark matter based on an asymmetry between \( \nu \) and \( \overline{\nu} \) [121].

### 4.2 Axions

Owing to space limitations, the discussion of axions as a dark matter candidate will be very brief. Axions are pseudo-Goldstone bosons which arise in solving the strong CP problem [122, 123] via a global U(1) Peccei–Quinn symmetry. The invisible axion [123] is associated with the flat direction of the spontaneously broken PQ symmetry. Because the PQ symmetry is also explicitly broken (the CP violating \( U(1) \) Peccei–Quinn symmetry. The invisible axion [123] is associated with the flat direction of the spontaneous breaking of PQ symmetry. Because the PQ symmetry is also explicitly broken (the CP violating \( U(1) \) Peccei–Quinn symmetry. The invisible axion [123] is associated with the flat direction of the spontaneous breaking of PQ symmetry. Because the PQ symmetry is also explicitly broken (the CP violating Peccei–Quinn symmetry. The invisible axion [123] is associated with the flat direction of the spontaneously broken PQ symmetry. Because the PQ symmetry is also explicitly broken (the CP violating Peccei–Quinn symmetry. The invisible axion [123] is associated with the flat direction of the spontaneously broken PQ symmetry. Because the PQ symmetry is also explicitly broken (the CP violating Peccei–Quinn symmetry. The invisible axion [123] is associated with the flat direction of the spontaneously broken PQ symmetry. Because the PQ symmetry is also explicitly broken (the CP violating Peccei–Quinn symmetry. The invisible axion [123] is

\[ a \sim m_\pi f_\pi / f_a \text{ where } f_a, \text{ the axion decay constant, is the vacuum expectation value of the PQ current and can be taken to be quite large. If we write the axion field as } a = f_a \theta, \text{ near the minimum, the potential produced by QCD instanton effects looks like } V \sim m_a^2 \theta^2 f_a^2. \]

The axion equations of motion lead to a relatively stable oscillating solution. The energy density stored in the oscillations exceeds the critical density [124] unless \( f_a \lesssim 10^{12} \text{ GeV} \).

Axions may also be emitted from stars and supernovae [125]. In supernovae, axions are produced via nucleon–nucleon bremsstrahlung with a coupling \( g_A N \propto m_N / f_a \). As was noted above, the cosmological density limit requires \( f_a \lesssim 10^{12} \text{ GeV}. \) Axion emission from red giants imply [126] \( f_a \gtrsim 10^{10} \text{ GeV} \) (though this limit depends on an adjustable axion–electron coupling), the supernova limit requires [127] \( f_a \gtrsim 2 \times 10^{11} \text{ GeV} \) for naive quark model couplings of the axion to nucleons. Thus only a narrow window exists for the axion as a viable dark matter candidate.

### 4.3 Neutralinos

Supersymmetry is one of the best-motivated proposals for physics beyond the Standard Model. It is well known that supersymmetry could help stabilize the mass scale of electroweak symmetry breaking by cancelling the quadratic divergences in the radiative corrections to the mass-squared of the Higgs boson [128]. In addition, including supersymmetric partners of Standard Model particles in the renormalization-group equations (RGEs) for the gauge couplings of the Standard Model would permit them to unify [129], whereas unification would not occur if only the Standard Model particles were included in the RGEs.

To construct the supersymmetric Standard Model [130] we start with the complete set of chiral fermions needed in the Standard Model, and add a scalar superpartner to each Weyl fermion so that each field in the Standard Model corresponds to a chiral multiplet. Similarly we must add a gaugino for each of the gauge bosons in the Standard Model making up the gauge multiplets. The Minimal Supersymmetric Standard Model (MSSM) [131] is defined by its minimal field content (which accounts for the known Standard Model fields) and minimal superpotential necessary to account for the known Yukawa mass terms. As such we define the MSSM by the superpotential

\[ W = \epsilon_{ij} \left[ y_e H_1^1 L^i e^c + y_d H_2^1 Q^j d^c + y_u H_3^1 Q^j u^c \right] + \epsilon_{ij} \mu H_1^i H_2^j. \]  

(117)

In Eq. (117), the indices \( \{ij\} \) are SU(2)_L doublet indices. The Yukawa couplings \( y \) are all \( 3 \times 3 \) matrices in generation space. Note that there is no generation index for the Higgs multiplets. Colour and generation indices have been suppressed in the above expression. There are two Higgs doublets in the MSSM. This is a necessary addition to the Standard Model which can be seen as arising from the holomorphic property of the superpotential. That is, there would be no way to account for all of the Yukawa terms for both up-type and down-type multiplets with a single Higgs doublet. To avoid a massless Higgs state, a mixing term must be added to the superpotential.
In defining the MSSM, we have limited the model to contain a minimal field content: the only new fields are those which are \textit{required} by supersymmetry. Consequently, apart from superpartners, only the Higgs sector was enlarged from one doublet to two. Moreover, in writing the superpotential (117), we have also made a minimal choice regarding interactions. We have limited the types of interactions to include only the minimal set required in the Standard Model and its supersymmetric generalization. There are, however, additional superpotential terms which are consistent with gauge invariance. These would lead to rapid baryon and/or lepton number violation and can be eliminated by imposing a discrete symmetry on the theory called \textit{R}-parity [132]. This can be represented as
\[
R = (-1)^{3B+L+2s},
\]
where \(B, L,\) and \(s\) are the baryon number, lepton number, and spin, respectively. It is easy to see that, with the definition (118), all the known Standard Model particles have \textit{R}-parity +1. For example, the electron has \(B = 0, L = -1,\) and \(s = 1/2,\) and the photon has \(B = L = 0\) and \(s = 1,\) so in both cases \(R = 1.\) Similarly, it is clear that all superpartners of the known Standard Model particles have \(R = -1,\) since they must have the same value of \(B\) and \(L\) as their conventional partners, but differ by 1/2 unit of spin. If \(R\)-parity is exactly conserved, then the additional superpotential terms must be absent from the theory. An immediate result of imposing \(R\)-parity is the stability of the lightest \(R = -1\) sparticle making it a potential dark matter candidate. Possible choices for the lightest supersymmetric particle (LSP) are the neutralino, sneutrino, and gravitino. Here, I will focus only on the former.

There are four neutralinos, each of which is a linear combination of the \(R = -1\) neutral fermions [133]: the wino \(\tilde{W}^\beta,\) the partner of the third component of the \(SU(2)\) gauge boson; the bino, \(\tilde{B};\) and the two neutral Higgsinos, \(\tilde{H}_1\) and \(\tilde{H}_2.\) The mass and composition of the LSP are determined by the gaugino masses, \(\mu,\) and \(\tan\beta.\) In general, neutralinos can be expressed as a linear combination
\[
\chi = \alpha\tilde{B} + \beta\tilde{W}^\beta + \gamma\tilde{H}_1 + \delta\tilde{H}_2.
\]
The solution for the coefficients \(\alpha, \beta, \gamma\) and \(\delta\) for neutralinos that make up the LSP can be found by diagonalizing the mass matrix
\[
(W^3, \tilde{B}, \tilde{H}_1^0, \tilde{H}_2^0) \rightarrow \begin{pmatrix} M_2 & 0 & -g_2v_1 \sqrt{2} & 2g_2v_2 \sqrt{2} \\ 0 & M_1 & 0 & 0 \\ -g_2v_1 \sqrt{2} & 0 & -\mu & 0 \\ 2g_2v_2 \sqrt{2} & \sqrt{2} & -\mu & 0 \end{pmatrix} \begin{pmatrix} \tilde{W}^3 \\ \tilde{B} \\ \tilde{H}_1^0 \\ \tilde{H}_2^0 \end{pmatrix},
\]
where \(M_1(M_2)\) is a soft supersymmetry breaking term giving mass to the \(U(1) (SU(2))\) gaugino(s).

The relic density of neutralinos depends on additional parameters in the MSSM beyond \(M_1, M_2, \mu,\) and \(\tan\beta.\) These include the sfermion masses \(m_f\) and the Higgs pseudo-scalar mass \(m_A.\) To determine the relic density it is necessary to obtain the general annihilation cross-section for neutralinos. In much of the parameter space of interest, the LSP is a bino and the annihilation proceeds mainly through sfermion exchange.

In its generality, the minimal supersymmetric standard model (MSSM) has over 100 undetermined parameters. There are good arguments based on grand unification [129] and supergravity [134] which lead to a strong reduction in the number of parameters. I will assume several unification conditions placed on the supersymmetric parameters. In all models considered, the gaugino masses are assumed to be unified at the GUT scale with value \(m_{1/2}\) as are the trilinear couplings with value \(A_0.\) Also common to all models considered here is the unification of all soft scalar masses set equal to \(m_0\) at the GUT scale. With this set of boundary conditions at the GUT scale, we can use the radiative electroweak symmetry breaking conditions by specifying the ratio of the two Higgs vacuum expectation values, \(\tan\beta,\) and the mass \(M_Z\) to predict the values of the Higgs mixing mass parameter \(\mu\) and Higgs pseudoscalar mass.
The sign of $\mu$ remains free. This class of models is often referred to as the constrained MSSM (CMSSM) \cite{135–139}. In the CMSSM, the solutions for $\mu$ generally lead to a lightest neutralino which is very nearly a pure $\tilde{\chi}^0_1$.

In Fig. 17, an example of the renormalization group running of the mass parameters in the CMSSM is shown. Here, we have chosen $m_{1/2} = 250$ GeV, $m_0 = 100$ GeV, $\tan \beta = 3$, $A_0 = 0$, and $\mu < 0$. Indeed, it is rather amazing that, from so few input parameters, all of the masses of the supersymmetric particles can be determined. The characteristic features that one sees in the figure are, for example, that the coloured sparticles are typically the heaviest in the spectrum. This is due to the large positive correction to the masses due to $\alpha_3$ in the RGEs. Also, one finds that the $\tilde{\chi}^0_1$ is typically the lightest sparticle. But most importantly, notice that one of the Higgs mass squared goes negative triggering electroweak symmetry breaking \cite{140}. (The negative sign in the figure refers to the sign of the mass squared, even though it is the mass of the sparticles which is depicted.)

For a given value of $\tan \beta$, $A_0$, and $sgn(\mu)$, the resulting regions of parameter space with acceptable relic density and which satisfy the phenomenological constraints can be displayed on the $m_{1/2} - m_0$ plane. In Fig. 18(a), the light shaded region corresponds to that portion of the CMSSM plane with $\tan \beta = 10$, $A_0 = 0$, and $\mu > 0$ such that the computed relic density yields the WMAP value given in Eq. (115) \cite{138}. The bulk region at relatively low values of $m_{1/2}$ and $m_0$, tapers off as $m_{1/2}$ is increased. At higher values of $m_0$, annihilation cross sections are too small to maintain an acceptable relic density and $\Omega_\chi h^2$ is too large. Although sfermion masses are also enhanced at large $m_{1/2}$ (due to RGE running), co-annihilation processes between the LSP and the next lightest sparticle (in this case the $\tilde{\tau}$) enhance the annihilation cross section and reduce the relic density. This occurs when the LSP and NLSP are nearly degenerate in mass. The dark shaded region has $m_{\tilde{\tau}} < m_\chi$ and is excluded. The effect of co-annihilations is to create an allowed band about 25–50 GeV wide in $m_0$ for $m_{1/2} \lesssim 950$ GeV, or $m_{1/2} \lesssim 400$ GeV, which tracks above the $m_{\tilde{\tau}1} = m_\chi$ contour \cite{141}.

Also shown in Fig. 18(a) are the relevant phenomenological constraints. These include the LEP limits on the chargino mass: $m_{\chi^\pm} > 104$ GeV \cite{142}; on the selectron mass: $m_{\tilde{e}} > 99$ GeV \cite{143}; and on the Higgs mass: $m_h > 114$ GeV \cite{144}. The former two constrain $m_{1/2}$ and $m_0$ directly via
the sparticle masses, and the latter indirectly via the sensitivity of radiative corrections to the Higgs mass to the sparticle masses, principally $m_{\tilde{t},\tilde{b}}$. FeynHiggs [145] is used for the calculation of $m_h$. The Higgs limit imposes important constraints principally on $m_{1/2}$ particularly at low $\tan \beta$. Another constraint is the requirement that the branching ratio for $b \rightarrow s\gamma$ be consistent with the experimental measurements [146]. These measurements agree with the Standard Model, and therefore provide bounds on MSSM particles [147] such as the chargino and charged Higgs masses, in particular. Typically, the $b \rightarrow s\gamma$ constraint is more important for $\mu < 0$, but it is also relevant for $\mu > 0$, particularly when $\tan \beta$ is large. The constraint imposed by measurements of $b \rightarrow s\gamma$ also excludes small values of $m_{1/2}$.

Finally, there are regions of the $(m_{1/2}, m_0)$ plane that are favoured by the BNL measurement [148] of $g_{\mu} - 2$ at the 2-$\sigma$ level, corresponding to a deviation from the Standard Model calculation [149].

Another mechanism for extending the allowed regions in the CMSSM to large $m_\chi$ is rapid annihilation via a direct-channel pole when $m_\chi \sim \frac{1}{2} m_A$ [135, 137]. Since the heavy scalar and pseudoscalar Higgs masses decrease as $\tan \beta$ increases, eventually $2m_\chi \simeq m_A$ yielding a ‘funnel’ extending to large $m_{1/2}$ and $m_0$ at large $\tan \beta$, as seen in Fig. 18(b). As one can see, the impact of the Higgs mass constraint is reduced (relative to the case with $\tan \beta = 10$) while that of $b \rightarrow s\gamma$ is enhanced.

Shown in Fig. 19 are the WMAP lines [138] of the $(m_{1/2}, m_0)$ plane for $\mu > 0$ and values of $\tan \beta$ from 5 to 55, in steps $\Delta(\tan \beta) = 5$. We notice immediately that the strips are considerably narrower than the spacing between them, though any intermediate point in the $(m_{1/2}, m_0)$ plane would be compatible with some intermediate value of $\tan \beta$. The right (left) ends of the strips correspond to the maximal (minimal) allowed values of $m_{1/2}$ and hence $m_\chi$. The lower bounds on $m_{1/2}$ are due to the Higgs mass constraint for $\tan \beta \leq 23$, but are determined by the $b \rightarrow s\gamma$ constraint for higher values of $\tan \beta$.

Finally, there is one additional region of acceptable relic density known as the focus-point region [150], which is found at very high values of $m_0$. An example showing this region is found in
Fig. 19: The strips display the regions of the \((m_{1/2}, m_0)\) plane that are compatible with the WMAP determination of \(\Omega_\chi h^2\) and the laboratory constraints for \(\mu > 0\) and \(\tan \beta = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55\). The parts of the strips compatible with \(g_\mu - 2\) at the 2-\(\sigma\) level have darker shading.

Fig. 20, plotted for \(\tan \beta = 10, \mu > 0\), and \(m_t = 172.4\) GeV. As \(m_0\) is increased, the solution for \(\mu\) at low energies as determined by the electroweak symmetry breaking conditions eventually begins to drop. When \(\mu \lesssim m_{1/2}\), the composition of the LSP gains a strong Higgsino component and as such the relic density begins to drop precipitously. As \(m_0\) is increased further, there are no longer any solutions for \(\mu\). This occurs in the shaded region in the upper left corner of Fig. 20. The position of the focus point strip is very sensitive to the value of \(m_t\) [151].

As seen in Fig. 18, the relic density constraint is compatible with relatively large values of \(m_{1/2}\) and \(m_0\). However, all values of \(m_{1/2}\) and \(m_0\) are not equally viable when the available phenomenological and cosmological constraints are taken into account. A global likelihood analysis enables one to pin down the available parameter space in the CMSSM. One can avoid the dependence on priors by performing a pure likelihood analysis as in Ref. [152], or a purely \(\chi^2\)-based fit as done in Refs. [153, 154]. Here we present results from one such analysis [155, 156], which used a Markov–Chain Monte Carlo (MCMC) technique to explore efficiently the likelihood function in the parameter space of the CMSSM. A full list of the observables and the values assumed for them in this global analysis are given in Ref. [154], and updated in Refs. [155, 156].

The 68\% and 95\% confidence-level (CL) regions in the \((m_{1/2}, m_0)\) plane of the CMSSM is shown in Fig. 21 [155]. Also shown for comparison are the physics reaches of ATLAS and CMS with 1/\(fb\) of integrated luminosity [157, 158]. (MET stands for missing transverse energy, SS stands for same-sign dilepton pairs, and the sensitivity for finding the lightest Higgs boson in cascade decays of supersymmetric particles is calculated for 2/\(fb\) of data.) The likelihood analysis assumed \(\mu > 0\), as motivated by the sign of the apparent discrepancy in \(g_\mu - 2\), but sampled all values of \(\tan \beta\) and \(A_0\): the experimental sensitivities were estimated assuming \(\tan \beta = 10\) and \(A_0 = 0\), but are probably not very sensitive to these assumptions. The global maximum of the likelihood function (indicated by the black dot) is at \(m_{1/2} = 310\) GeV, \(m_0 = 60\) GeV, \(A_0 = 240\) GeV, \(\tan \beta = 11\), and \(\chi^2/N_{dof} = 20.4/19\) (37\% probability). It is encouraging that the best-fit point lies well within the LHC discovery range, as do the 68\% and
most of the 95% CL regions.

Improvements in sensitivity have made it possible for direct detection experiments [159, 160] to be competitive with other phenomenological constraints. The elastic cross section for $\chi$ scattering on a nucleus can be decomposed into a scalar (spin-independent) and a spin-dependent part. Each of these can be written in terms of the cross sections for elastic scattering off individual nucleons. The scalar part of the cross section can be written as

$$\sigma_{SI} = \frac{4m_r^2}{\pi} [Z f_p + (A - Z)f_n]^2,$$

(121)

where $m_r$ is the $\chi$-nuclear reduced mass and

$$\frac{f_N}{m_N} = \sum_{q=u,d,s} f_{T_q}^{(N)} \frac{\alpha_{3q}}{m_q} + \frac{2}{2\pi} f_{TG} \sum_{q=c,b,t} \frac{\alpha_{3q}}{m_q},$$

(122)

for $N = p$ or $n$. The parameters $f_{T_q}^{(N)}$ are defined by

$$m_N f_{T_q}^{(N)} \equiv \langle N | m_q \bar{q}q | N \rangle \equiv m_q B_q^{(N)},$$

(123)

and the $\alpha_{3q}$ contain the individual quark-neutralino scattering cross sections, see Refs. [161–163] for further details regarding the calculation of the cross section.

The elastic scattering of neutralinos on nucleons is very sensitive to the strangeness contribution to the nucleon mass and can be characterized by the parameter $y$ which is also related to the $\pi$-nucleon sigma term $\Sigma_{\pi N}$ by

$$y \equiv \frac{2B_s}{B_u + B_d} = 1 - \frac{\sigma_0}{\Sigma_{\pi N}},$$

(124)

where $\sigma_0$ is the change in the nucleon mass due to non-zero $u$ and $d$ masses and is estimated from octet baryon mass differences to be $\sigma_0 = 36$ MeV [164], and the latest determination of $\Sigma_{\pi N} = 64$ MeV.

The effects of varying these assumptions are discussed in the context of the CMSSM in Refs. [162,
Fig. 21: The \((m_0, m_{1/2})\) plane in the CMSSM showing the regions favoured in a likelihood analysis at the 68\% (blue) and 95\% (red) confidence levels [155]. The best-fit point is shown by the black point. Also shown are the discovery contours in different channels for the LHC with 1/fb (2/fb for the Higgs search in cascade decays of sparticles) [157, 158].

Lattice calculations are now reaching the stage where they may also provide useful information on \(\Sigma_{\pi N}\) [165], and a recent analysis would suggest a lower value \(\Sigma_{\pi N} < 40\) [166].

In Fig. 22(a) we show CMSSM spin-independent neutralino-nucleon cross section, as obtained in a scan over all CMSSM parameters with \(5 \leq \tan \beta \leq 55\), \(100 \leq m_{1/2} \leq 2000\) GeV, \(0\) GeV \(\leq m_0 \leq 2000\) GeV, and \(-3m_{1/2} \leq A_0 \leq 3m_{1/2}\) [167]. We also allow both positive and negative \(\mu\), except for large \(\tan \beta > 30\), where convergence becomes difficult in the \(\mu < 0\) case. At low \(m_\chi < 300\) GeV, cross sections generally exceed \(10^{-7}\) pb, and the largest scalar cross sections, which occur for \(m_\chi \sim 100\) GeV, are already excluded by CDMS II [159] and/or XENON10 [160]. These exclusions occur primarily in the focus-point region at large \(\tan \beta\). On the other hand, for \(m_\chi \gtrsim 400\) GeV, scalar cross sections are well below \(10^{-9}\) pb, and come from the co-annihilation strip or the rapid-annihilation funnel that appears at large \(\tan \beta\) in the CMSSM. The effective cross sections shown are suppressed for points with \(\Omega_\chi \ll \Omega_{CDM}\), and there may be cancellations at larger \(m_\chi\) that suppress the cross sections substantially. These regions of parameter space will not be probed by direct detection experiments in the near future [168, 169]. The corresponding 68\% and 95\% CL regions in the cross section–neutralino mass plane from the frequentist analysis of Ref. [156] are shown in Fig. 22(b).

Acknowledgements

This work was supported in part by DOE grant DE-FG02-94ER-40823

References

Fig. 22: (a) The entire potential range of neutralino–nucleon cross sections as functions of neutralino mass for the CMSSM, with $5 \leq \tan \beta \leq 55$, $0 \leq m_{1/2} \leq 2000$ GeV, $100$ GeV $\leq m_0 \leq 2000$ GeV, and $-3m_{1/2} \leq A_0 \leq 3m_{1/2}$. Also shown are upper limits on the spin-independent dark matter scattering cross section from CDMS II [159] (solid black line) and XENON10 [160] (solid pink line), as well as the expected sensitivities for XENON100 [168] (dashed pink line) and SuperCDMS at the Soudan Mine [169] (dashed black line). Taken from Ref. [167]. (b) The 68% and 95% CL areas in the cross section–mass plane from the frequentist analysis of Ref. [156].


