Chapter 12

Supernovae of Type Ia

12.1 Standard candles and distances

12.1.1 The principle

- Before starting with the details of supernovae, their type Ia and their cosmological relevance, let us set the stage with a few illustrative considerations.

- Suppose we had a standard candle whose luminosity, L, we know precisely. Then, according to the definition of the luminosity distance in (2.16), the distance can be inferred from the measured flux, f, through

$$D_{\mathrm{lum}} = \sqrt{\frac{L}{4\pi f}}\ . \tag{12.1}$$

- Besides the redshift z, the luminosity distance will depend on the cosmological parameters,

$$D_{\mathrm{lum}} = D_{\mathrm{lum}}(z; \Omega_{\mathrm{m}0}, \Omega_{\Lambda 0}, H_0, \ldots)\ , \tag{12.2}$$

  which can be used in principle to determine cosmological parameters from a set of distance measurements from a class of standard candles.

- For this to work, the standard candles must be at a suitably high redshift for the luminosity distance to depend on the cosmological model. As we have seen in (2.17), all distance measures tend to

$$D \approx \frac{cz}{H_0} \tag{12.3}$$

  at low redshift and lose their sensitivity to all cosmological parameters except H_0.

- In reality, we rarely know the absolute luminosity L even of cosmological standard candles. The problem is that they need to be calibrated first, which is only possible from a flux measurement once the distance is known by other means, such as from parallaxes in case of the Cepheids.

- Supernovae, however, which are the subject of this chapter, are typically found at distances which are way too large to allow direct distance measurements. Therefore, the only way out is to combine distant supernovae with local ones, for which the approximate distance relation (12.3) holds.

- Any measurement of flux f_i and redshift z_i of the i-th standard candle in a sample then yields an estimate for the luminosity L in terms of the squared inverse Hubble constant,

$$L = 4\pi f_i \left(\frac{cz_i}{H_0}\right)^2 . \tag{12.4}$$

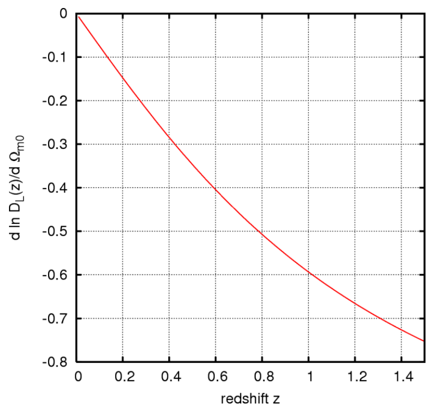
  Since all cosmological distance measures are proportional to the Hubble length c/H_0, the dependences on H_0 on both sides of (12.1) cancels, and the determination of cosmological parameters *other than* the Hubble constant becomes possible. Thus, the first lesson to learn is that cosmology from distant supernovae requires a sample of nearby supernovae for calibration.

- Of course, this nearby sample must satisfy the same criterion as the distance indicators used for the determination of the Hubble constant: their redshifts must be high enough for the peculiar velocities to be negligible, thus $z \gtrsim 0.02$. On the other hand, the redshifts must be low enough for the linear appoximation (12.3) to hold.

- It is important to note that it is not necessary to know the absolute luminosity L even up to the uncertainty in H_0. If L is truly independent of redshift, cosmological parameters could still be determined through (12.1) from the *shape* of the measured relation between flux and redshift even though its precise *amplitude* may be unknown. It is only important that the objects used *are* standard candles, but not how bright they are.

12.1.2 Requirements and degeneracies

- Let us now collect several facts about cosmological inference from standard candles. Since we aim at the determination of cosmological parameters, say Ω_{m0}, it is important to estimate the accuracy that we can achieve from measurements of the luminosity distance.



Logarithmic derivative of the luminosity distance with respect to Ω_{m0}.

- Suppose we restrict the attention to spatially flat cosmological models, for which $\Omega_{\Lambda 0} = 1 - \Omega_{m0}$. Then, because the dependence on the Hubble constant was canceled, Ω_{m0} is the only remaining relevant parameter. We estimate the accuracy through first-order Taylor expansion,

$$\Delta D_{\rm lum} \approx \frac{\mathrm{d}D_{\rm lum}}{\mathrm{d}\Omega_{\rm m0}} \Delta\Omega_{\rm m0} \; , \tag{12.5}$$

about a fiducial model, such as a ΛCDM model with $\Omega_{m0} = 0.3$.

- At a fiducial redshift of $z \approx 0.8$, we find numerically

$$\frac{\mathrm{d}\ln D_{\rm lum}}{\mathrm{d}\Omega_{\rm m0}} \approx -0.5 \; , \tag{12.6}$$

which shows that a *relative* distance accuracy of

$$\frac{\Delta D_{\rm lum}}{D_{\rm lum}} \approx -0.5\Delta\Omega_{\rm m0} \tag{12.7}$$

is required to achieve an absolute accuracy of $\Delta\Omega_{m0}$. For $\Delta\Omega_{m0} \approx 0.02$, say, distances thus need to be known to $\approx 1\%$.

- This accuracy requires sufficiently large supernova samples. Assuming Poisson statistics for simplicity and distance measurements to N supernovae, the combined accuracy is

$$|\Delta\Omega_{\rm m0}| \approx \frac{2}{\sqrt{N}} \frac{\Delta D_{\rm lum}}{D_{\rm lum}} \; . \tag{12.8}$$

That is, an accuracy of $\Delta\Omega_{m0} \approx 0.02$ can be achieved from ≈ 100 supernovae whose individual distances are known to $\approx 10\%$.

- Anticipating physical properties of type-Ia supernovae, their intrinsic peak luminosities in blue light are $L \approx 3.3 \times 10^{43}\,\mathrm{erg\,s^{-1}}$, with a relative scatter of order 10%. (As we shall see later, type-Ia supernovae are *standardisable* rather than standard candles, and the standardising procedure is currently not able to reduce the scatter further.)

- Given uncertainties in the luminosity L and in the flux measurement, error propagation on (12.1) yields the distance uncertainty

$$D_{\rm lum} = \left[\left(\frac{\mathrm{d}D_{\rm lum}}{\mathrm{d}L}\right)^2 \Delta L^2 + \left(\frac{\mathrm{d}D_{\rm lum}}{\mathrm{d}f}\right)^2 \Delta f^2\right]^{1/2} \; , \tag{12.9}$$

or the relative uncertainty

$$\frac{\Delta D_{\rm lum}}{D_{\rm lum}} = \frac{1}{2}\left[\left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta f}{f}\right)^2\right]^{1/2} \; . \tag{12.10}$$

Even if the flux could be measured precisely, the intrinsic luminosity scatter currently forbids distance determinations to better than 10%.

- Fluxes have to be inferred from photon counts. For various reasons to be clarified later, supernova light curves should be determined until ~ 35 days after the peak, when the luminosity has typically dropped to $\approx 2.5 \times 10^{42}\,\mathrm{erg\,s^{-1}}$. The luminosity distance to $z \approx 0.8$ is $\approx 5\,\mathrm{Gpc}$, which implies fluxes $f \approx 1.1 \times 10^{-14}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ at peak and $f \approx 8.7 \times 10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ 35 days later.

- Dividing by an average photon energy of 5×10^{-12} erg, multiplying with the area of a typical telescope mirror with 4 m diameter, and assuming a total quantum efficiency of 30%, we find detected photon fluxes of $f_\gamma \approx 85\,\mathrm{s^{-1}}$ at peak and $f_\gamma \approx 7\,\mathrm{s^{-1}}$ 35 days afterwards. These fluxes are typically distributed over a few CCD pixels.

- Supernovae occur in galaxies, which means that their fluxes need to be measured on the background of the galactic light. On the area of a distant supernova image, the photon flux from the host galaxy is comparable to the flux from the supernova. Therefore, an estimate for the signal-to-noise ratio for the detection is

$$\frac{\mathrm{S}}{\mathrm{N}} \approx \frac{N}{2\sqrt{N}} = \frac{\sqrt{N}}{2}\,, \tag{12.11}$$

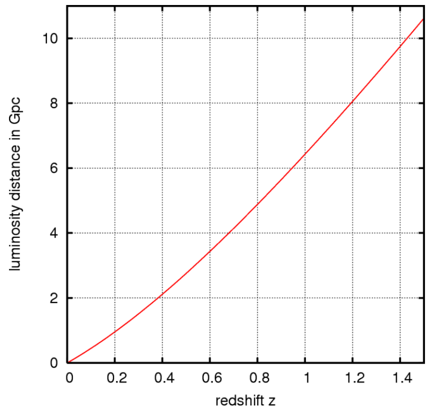
  where N is the number of photons *per pixel* detected from supernova and host galaxy during the exposure time. Signal-to-noise ratios of $\gtrsim 10$ up to 35 days after the maximum thus require $N \approx 400$ photons per pixel. Assuming that the supernova appears on typically ~ 4 pixels, this implies exposure times of order $4 \times 400/7 \approx 230\,\mathrm{s}$, or a few minutes. Typical exposure times are of order $15\ldots30$ minutes to capture supernovae out to redshifts $z \sim 1$. Then, the photometric error around peak luminosity is certainly less than the remaining scatter in the intrinsic luminosity, and relative distance accuracies of order 10% are within reach.

- However, a major difficulty is the fact that the identification of type-Ia supernovae requires spectroscopy. Sufficiently accurate spectra typically require long exposures on the world's largest telescopes, such as ESO's Very Large Telescope which consists of four individual mirrors with 8 m diameter each.

- In order to see what we can hope to constrain by measuring angular-diameter distances, we form the gradient of D_lum in the Ω_m0-$\Omega_{\Lambda0}$ plane,

$$\vec{g} \equiv \left(\frac{\partial D_\mathrm{lum}}{\partial\Omega_\mathrm{m0}}, \frac{\partial D_\mathrm{lum}}{\partial\Omega_{\Lambda0}}\right)^t\,, \tag{12.12}$$

  at a fiducial ΛCDM model with $\Omega_\mathrm{m0} = 0.3$. When normalised to



The luminosity distance in a universe with $\Omega_\mathrm{m0} = 0.3$ and $\Omega_{\Lambda0} = 0.7$ with Hubble constant $h = 0.72$.

unit length, it turns out to point into the direction

$$\vec{g} = \begin{pmatrix} -0.76 \\ 0.65 \end{pmatrix} . \tag{12.13}$$

- This vector rotated by 90° then points into the direction in the Ω_{m0}-$\Omega_{\Lambda 0}$ plane along which the luminosity distance does *not* change. Thus, near the fiducial ΛCDM model, the parameter combination

$$P \equiv \vec{g} \cdot \begin{pmatrix} \Omega_{m0} \\ \Omega_{\Lambda 0} \end{pmatrix} = -0.76\,\Omega_{m0} + 0.65\,\Omega_{\Lambda 0} \tag{12.14}$$

is degenerate. The degeneracy direction, characterised by the vector $\mathcal{R}(\pi/2)\,\vec{g} = (0.65, 0.76)^t$, points under an angle of $\arctan(0.76/0.65) = 49.5°$ with the Ω_{m0} axis, almost along the diagonal from the lower left to the upper right corner of the parameter plane. Thus, it is almost perpendicular to the degeneracy direction obtained from the curvature constraint due to the CMB. This illustrates how parameter degeneracies can very efficiently be broken by *combining* suitably different types of measurement.

12.2 Supernovae

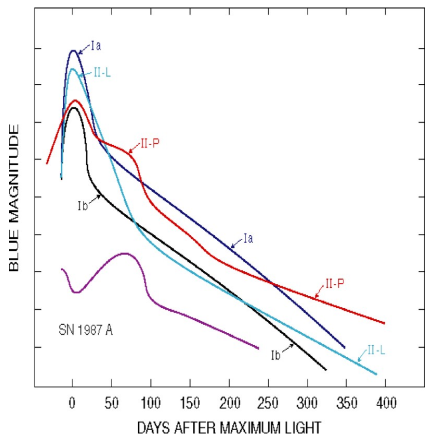
12.2.1 Types and classification

- Supernovae are "eruptively variable" stars. A sudden rise in brightness is followed by a gentle decline. They are unique events which at peak brightness reach luminosities comparable to those of an entire galaxy, or $10^{10} \ldots 10^{11}\, L_\odot$. They reach their maxima within days and fade within several months.

- Supernovae are traditionally characterised according to their early spectra. If hydrogen lines are missing, they are of type I, otherwise of type II. Type-Ia supernovae show silicon lines, unlike type-Ib/c supernovae, which are distinguished by the prominence of helium lines. Normal type-II supernovae have spectra dominated by hydrogen. They are subdivided according to their lightcurve shape into type-IIL and type-IIP. Type-IIb supernova spectra are dominated by helium instead.

- Except for type-Ia, supernovae arise due to the collapse of a massive stellar core, followed by a thermonuclear explosion which disrupts the star by driving a shock wave through it. Core-collapse supernovae of type-I (i.e. types Ib/c) arise from stars with masses between $8 \ldots 30\, M_\odot$, those of type-II from more massive stars.

Supernova 1994d in its host galaxy.



Lightcurves of supernovae of different types.

- Type-Ia supernovae, which we are dealing with here, arise when a white dwarf is driven over the Chandrasekhar mass limit by mass overflowing from a companion star. In a binary system, the more massive star evolves faster and can reach its white-dwarf stage before its companion leaves the main sequence and becomes a red giant. When this happens, and the stars are close enough, matter will flow from the expanding red giant on the white dwarf.

- Electron degeneracy pressure can stabilise white dwarfs up to the Chandrasekhar mass limit of $\sim 1.4\,M_\odot$. When the white dwarf is driven over that limit, it collapses, starts a thermonuclear runaway and explodes. Since this type of explosion involves an approximately fixed amount of mass, it is physically plausible that the explosion releases a fixed amount of energy. Thus, the Chandrasekhar mass limit is the main responsible for type-Ia supernovae to be approximate standard candles.

- The thermonuclear runaway in type-Ia supernovae converts the carbon and oxygen in the core of the white dwarf into ^{56}Ni, which later decays through ^{56}Co into the stable ^{56}Fe. According to detailed numerical explosion models, the nuclear fusion is started at random points near the centre of the white dwarf.

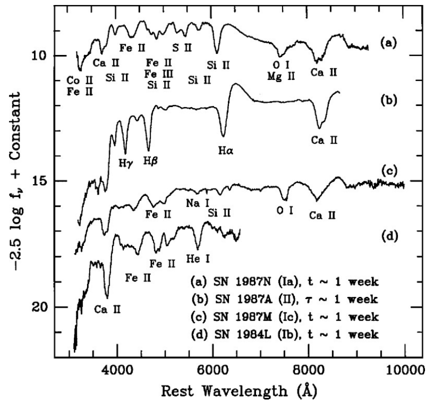
- Since the core material is degenerate, its pressure is independent of its temperature. The mass accreted from the companion star increases the pressure. Once it exceeds the Fermi pressure, inverse beta decay sets in,

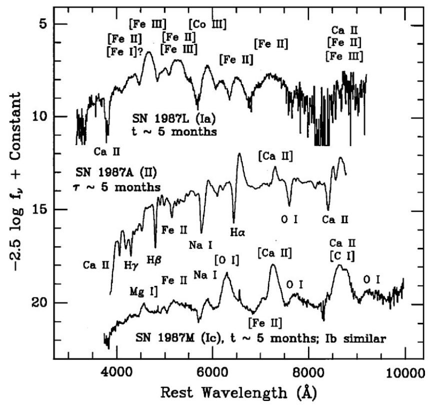
$$p + e^- \to n + \nu_e + \gamma\,, \tag{12.15}$$

  which suddenly removes the degenerate electrons. Under the high pressure, the temperature rises dramatically and ignites the fusion. The neutrinos carry away much of the explosion energy unnoticed because they can leave the supernova essentially without further interaction.

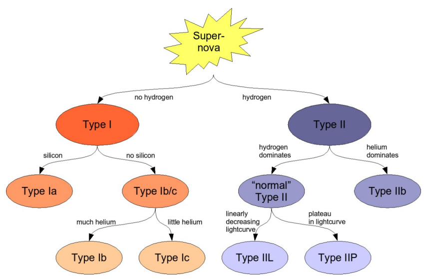
- The presence of silicon lines in the type-Ia spectra indicates that not all of the white dwarf's material is converted into ^{56}Ni. This shows that there is no explosion, but a deflagration, in which the flame front propagates at velocities below the sound speed. The deflagration can burn the material fast enough if it is turbulent, because the turbulence dramatically increases the surface of the flame front and thus the amount of material burnt per unit time. Typically, $\sim 0.5\,M_\odot$ of ^{56}Ni is produced in theoretical models.

- The peak brightness is reached when the deflagration front reaches the former white dwarf's surface and drives it as a rapidly expanding envelope into the surrounding space. The γ photons released in the nuclear fusion processes are redshifted by scattering

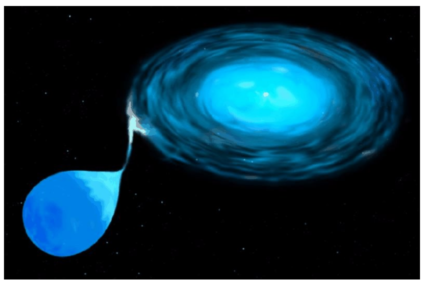




Early (top) and late spectra of different supernova types.



Supernova classification.



Type-Ia supernovae occur when white dwarfs are driven over the Chandrasekhar mass limit by mass flowing from a companion star.

with the expanding material and finally leave the explosion site as X-ray, UV, optical and infrared photons.

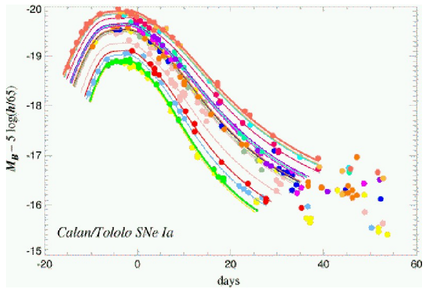
- Once the thermonuclear fusion has ended, additional energy is released by the β decay of ^{56}Co into ^{56}Fe with a half life of 77.12 days. The exponential nature of the radioactive decay causes the typical exponential decline phase in supernova light curves.
- Since the supernova light has to propagate through the expanding envelope before we can see it, the opacity of the envelope and thus its metallicity are important for the appearance of the supernova.

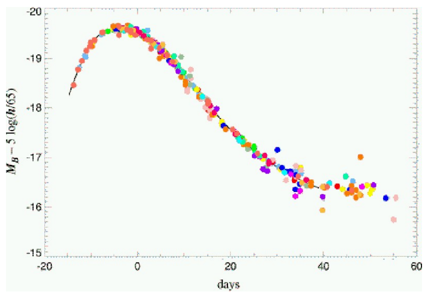
12.2.2 Observations

- Since supernovae are transient phenomena, they can only be detected by sufficiently frequent monitoring of selected areas in the sky. Typically, fields are selected by their accessibility for the telescope to be used and the least degree of absorption by the Galaxy. Since a type-Ia supernova event lasts for about a month, monitoring is required every few days.
- Supernovae are then detected by differential photometry, in which the average of all preceding images is subtracted from the last image taken. Since the seeing varies, the images appear convolved with point-spread functions of variable width even if they are taken with identical optics, thus the objects on them appear more or less blurred. Before they can be meaningfully subtracted, they therefore have to be convolved with the same effective point-spread function. This causes several complications in the later analysis procedure, in particular with the photometry.
- Of course, this detection procedure returns many variable stars and supernovae of other types, which are not standard candles and have to be removed from the sample. Pre-selection of type-Ia candidates is done by colour and the light-curve shape, but the identification of type-Ia supernovae requires spectroscopy in order to identify the decisive silicon lines at 6347 Å and 6371 Å. Since these lines move out of the optical spectrum for redshifts $z \gtrsim 0.5$, near-infrared observations are crucially important for the high-redshift supernovae relevant for cosmology.
- Nearby supernovae, which we have seen need to be observed for calibration, show that type-Ia supernovae are *not* standard candles but show a substantial scatter in luminosity. It turned out that there is an empirical relation between the duration of the supernova event and its peak brightness in that brighter supernovae last longer.

- This relation between the light-curve shape and the brightness can be used to *standardise* type-Ia supernovae. It was seen as a major problem for their cosmological interpretation that the origin for this relation was unknown, and that its application to high-redshift supernovae was based on the untested assumption that the relation found and calibrated with local supernovae would also hold there. Recent simulations indicate that the relation is an opacity effect: brighter supernovae produce more ^{56}Ni and thus have a higher metallicity, which causes the envelope to be more opaque, the energy transport through it to be slower, and therefore the supernova to last longer.

- Thus, before a type-Ia supernova can be used as a standard candle, its duration must be determined, which requires the light-curve to be observed over sufficiently long time. It has to be taken into account here that the cosmic expansion leads to a time dilation, due to which supernovae at redshift z appear longer by a factor of $(1 + z)$. We note in passing that the confirmation of this time dilation effect indirectly confirms the cosmic expansion. *After* the standardisation, the scatter in the peak brightnesses of nearby supernovae is substantially reduced. This encourages (and justifies) their use as *standardisable* candles for cosmology.





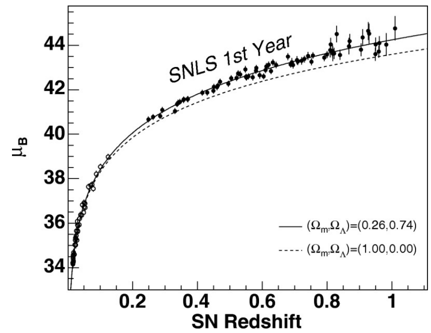
Lightcurves of type-Ia supernovae before (top) and after correction.

- The remaining relative uncertainty is now typically between $10\dots15\%$ for individual supernovae. Since, as we have seen following (12.7), we require relative distance uncertainties at the per cent level, of order a hundred distant supernovae are required before meaningful cosmological constraints can be placed, which justifies the remark after (12.8).

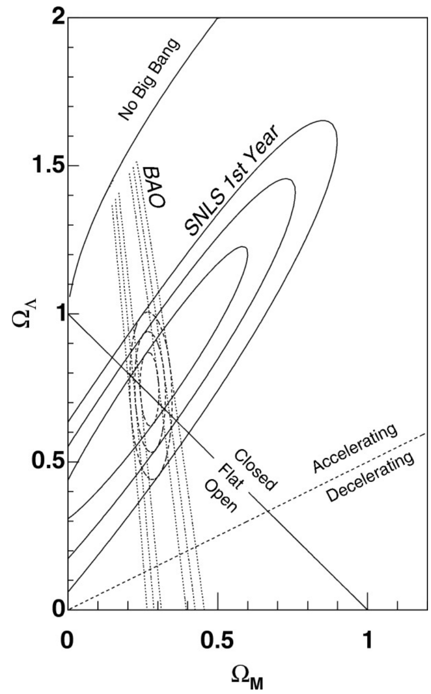
- An example for the several currently ongoing supernova surveys is the *Supernova Legacy Survey* (SNLS) in the framework of the Canada-France-Hawaii Legacy Survey (CFHTLS), which is carried out with the 4-m Canada-France-Hawaii telescope on Mauna Kea. It monitors four fields of one square degree each five times during the 18 days of dark time between two full moons (lunations).

- Differential photometry is performed to find out variables, and candidate type-Ia supernovae are selected by light-curve fitting after removing known variable stars. Spectroscopy on the largest telescopes (mostly ESO's VLT, but also the Keck and Gemini telescopes) is then needed to identify type-Ia supernovae. To give a few characteristic numbers, the SNLS has taken 142 spectra of type-Ia candidates during its first year of operation, of which 91 were identified as type-Ia supernovae.

- The light curves of these objects are observed in several different filter bands. This is important to correct for interstellar absorption. Any dimming by intervening material makes supernovae appear fainter, and thus more distant, and will bias the cosmological results towards faster expansion. Since the intrinsic colours of type-Ia supernovae are characteristic, any deviation between the observed and the intrinsic colours signals interstellar absorption which is corrected by adapting the amount of absorption such that the observed is transformed back into the intrinsic colour.



Distances to type-Ia supernovae (in logarithmic units) as a function of their redshift, as measured by the *Supernova Legacy Survey*.



Cosmological parameter constraints derived from the same data.

- This correction procedure is expected to work well unless there is material on the way which absorbs equally at all wavelengths, so-called "grey dust". This could happen if the absorbing dust grains are large compared to the wavelength. Currently, it is quite difficult to concusively rule out grey dust, although it is implausible based on the interstellar absorption observed in the Galaxy.

- After applying the corrections for absorption and duration, each supernova yields an estimate for the luminosity distance to its redshift. Together, the supernovae in the observed sample constrain the evolution of the luminosity distance with redshift, which is then fit varying the cosmological parameters except for H_0, i.e. typically Ω_{m0} and $\Omega_{\Lambda 0}$. This yields an "allowed" region in the Ω_{m0}-$\Omega_{\Lambda 0}$ plane compatible with the measurements which is degenerate in the direction calculated before.

- More information or further assumptions are necessary to break the degeneracy. The most common assumption, justified by the CMB measurements, is that the Universe is spatially flat. Based upon it, the SNLS data yield a matter density parameter of

$$\Omega_{m0} = 0.263 \pm 0.037 \ . \qquad (12.16)$$

This is a remarkable result. First of all, it confirms the other independent measurements we have already discussed, which were based on kinematics, cluster evolution and the CMB. Second, it shows that, in the assumed spatially flat universe, the dominant contribution to the total energy density must come from something else than matter, possibly the cosmological constant.

- It is important for the later discussion to realise in what way the parameter constraints from supernovae differ from those from the CMB. The fluctuations in the latter show that the Universe is at least nearly spatially flat, and the density parameters in dark and baryonic matter are near 0.25 and 0.045, respectively. The rest must be the cosmological constant, or the dark energy. Arising *early* in the cosmic history, the CMB itself is almost insensitive to the cosmological constant, and thus it can only constrain it indirectly.

- Type-Ia supernovae, however, measure the angular-diameter distance during the *late* cosmic evolution, when the cosmological constant is much more important. As (12.14) shows, the luminosity distance constrains the *difference* between the two parameters,

$$\Omega_{\Lambda 0} = 1.17\,\Omega_{\mathrm{m}0} + P\,, \tag{12.17}$$

where the degenerate parameter P is determined by the measurement. Assuming $\Omega_{\Lambda 0} = 1 - \Omega_{\mathrm{m}0}$ as in a spatially-flat universe yields

$$P = 1 - 2.17\,\Omega_{\mathrm{m}0} \approx 0.43 \tag{12.18}$$

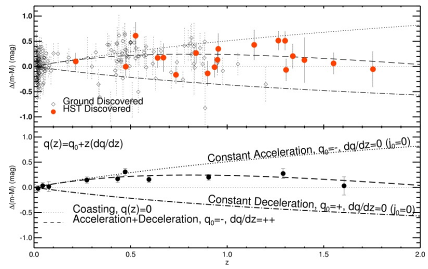
from the SNLS first-year result (12.16), illustrating that the survey has constrained the density parameters to follow the *relation*

$$\Omega_{\Lambda 0} \approx 1.17\,\Omega_{\mathrm{m}0} + 0.43\,. \tag{12.19}$$

- The relative *acceleration* of the universe, $\ddot{a}/a$, is given by the equation

$$\frac{\ddot{a}}{a} = H_0^2\left(\Omega_{\Lambda 0} - \frac{\Omega_{\mathrm{m}0}}{2a^3}\right) \tag{12.20}$$

if matter is pressure-less, which follows directly from Einstein's field equations. Thus, the expansion of the universe accelerates today ($a = 1$) if $\ddot{a} = H_0^2(\Omega_{\Lambda 0} - \Omega_{\mathrm{m}0}/2) > 0$, or $\Omega_{\Lambda 0} > \Omega_{\mathrm{m}0}/2$. Given the measurement (12.19), the conclusion seems inevitable that the Universe's expansion does indeed accelerate today.



Above redshift $z \approx 1$, the cosmic acceleration seems to turn into deceleration.

- If the Universe is indeed spatially flat, then the transition between decelerated and accelerated expansion happened at

$$1 - 0.263 \approx \frac{0.263}{2a^3} \quad \Rightarrow \quad a = 0.56\,, \tag{12.21}$$

or at redshift $z \approx 0.78$. Luminosity distances to supernovae at larger redshifts should show this transition, and in fact they do.

12.2.3 Potential problems

- The main observational issues is dust in the SN host galaxy. If one is careful enough (an observational issue), then one can estimate the reddening of the supernova accurately. Yet, there is a terrible conceptual issue that is currently one of the major stumbling-blocks of the field - we do not know with certainty what the relationship between reddening and the extinction (the amount of light lost - this affects the inferred absolute magnitude).

- There are two ways around this issue - trying to better estimate the attenuation curve shape (the relationship between reddening and extinction, this is not straightforward and requires excellent photometry over a long wavelength range; see Wood-Vasey et al.

2007), and going to systems that one *a priori* expects to be free of dust (elliptical galaxies; this involves a considerable reduction in sample size; see, e.g., Sullivan et al. 2003, MNRAS, 340, 1057). Both approaches have been taken.

- The problem with possible grey dust has already been mentioned: While the typical colours of type-Ia supernovae allow the detection and correction of the reddening coming with typical interstellar absorption, grey dust would leave no trace in the coulours and remain undetectable. However, grey dust would re-emit the absorbed radiation in the infrared and add to the infrared background, which is quite well constrained. It thus seems that grey dust is not an important contaminant, if it exists.

- Gravitational lensing is inevitable for distant supernovae. Depending on the line-of-sight, they are either magnified or demagnified. Since, due to nonlinear structures, high magnifications can occasionally happen, the magnification distribution must be skewed towards demagnification to keep the mean of zero magnification. Thus, the *most probable* magnification experienced by supernova is *below unity*. In other words, lensing may lead to a slight demagnification if lines-of-sight towards type-Ia supernovae are random with respect to the matter distribution. In any case, the *rms* cosmic magnification adds to the intrinsic scatter of the supernova luminosities. It may become significant for redshifts $z \gtrsim 1$.

- It is a difficult and debated question whether supernovae at high redshifts are intrinsically the same as at low redshifts where they are calibrated. Should there be undetected systematic differences, cosmological inferences could be wrong. In particular, it may be natural to assume that metallicties at high redshifts are lower than at low redshifts. Since supernovae last longer if their atmospheres are more opaque, lower metallicity may imply shorter supernova events, leading to underestimated luminosities and overestimated distances. Simulations of type-Ia supernovae, however, seem to show that such an effect is probably not significant.

- It was also speculated that distant supernovae may be intrinsically bluer than nearby ones due to their lower metallicity. Should this be so, the extinction correction, which is derived from reddening, would be underestimated, causing intrinsic luminosities to be under- and luminosity distances to be overestimated. Thus, this effect would lead to an *underestimate* of the expansion rate and counteract the cosmological constant. There is currently no indication of such a colour effect.

- Supernovae of types Ib/c may be mistaken for those of type Ia if the identification of the characteristic silicon lines fails for some reason. Since they are typically fainter than type-Ia supernovae, they would contaminate the sample and bias results towards higher luminosity distances, and thus towards a higher cosmological constant. It seems, however, that the possible contamination by non-type-Ia supernovae is so small that it has no noticeable effect.

- Several more potential problems exist. It has been argued for a while that, if the evidence for a cosmological constant was based exclusively on type-Ia supernovae, it would probably not be considered entirely convincing. However, since the supernova observations come to conclusions compatible with virtually all independent cosmological measurements, they add substantially to the persuasiveness of the cosmological standard model. Moreover, recent supernova simulations reveal good physical reasons why they should in fact be reliable, standardisable candles.

Chapter 5

Big-Bang Nucleosynthesis

5.1 Key concepts

- Impossible to produce observed Helium through stellar nucleosynthesis: need primordial generation
- Critical step in nucleosynthesis chain is $p+n \rightarrow d+\gamma$; other reactions in the chain are fast, converting almost all d into Helium-4.
- The Gamow criterion quantifies the 'sweet spot' that one needs to hit to generate Helium: not too little (because no d to generate Helium). The criterion goes $n_B\langle\sigma v\rangle t \sim 1$; given n_B estimates and known velocity-averaged cross section one can estimate t and therefore T, and predict the temperature of the CMB (the answer comes out at ~5K).
- The abundances of d, ^{3}He, ^{4}He and ^{7}Li can be estimated (from very precise spectral measurements of stars and high-redshift absorption-line systems) and yield important constraints on the baryon to photon ratio; because the photon density of the CMB is known to exquisite accuracy such an exercise yields an excellent estimate of baryon density (where most of the power comes from d).
- useful website for more is : http://www.astro.ucla.edu/~wright/BBNS.html

5.2 The origin of Helium-4 and the other light elements

5.2.1 The riddle of Helium

- since conversions between temperatures and energies will occur frequently in this chapter, recall that a thermal energy of 1 eV corresponds to a temperature of 1.16×10^4 K;

- stellar spectra show that the abundance of Helium-4 in stellar atmospheres is of order $Y = 0.25$ by mass, i.e. about a quarter of the baryonic mass in the Universe is composed of Helium-4;

- Helium-4 is produced in stars in the course of hydrogen burning; per ^{4}He nucleus, the amount of energy released corresponds to 0.7% of the masses involved, or

$$\begin{aligned} \Delta E &= \Delta mc^2 = 0.007\,(2m_\mathrm{p} + 2m_\mathrm{n})c^2 \approx 0.028\,m_\mathrm{p}c^2 \\ &\approx 26\,\mathrm{MeV} \approx 4.2 \times 10^{-5}\mathrm{erg}\ ; \end{aligned} \tag{5.1}$$

- suppose a galaxy such as ours, the Milky Way, shines with a luminosity of $L \approx 10^{10}\,L_\odot \approx 3.8 \times 10^{43}\,\mathrm{erg\,s^{-1}}$ for a good fraction of the age of the Universe, say for $\tau = 10^{10}\,\mathrm{yr} \approx 3 \times 10^{17}$ s; then, it releases a total energy of

$$E_\mathrm{tot} \approx L\tau \approx 1.1 \times 10^{61}\,\mathrm{erg}\ ; \tag{5.2}$$

- the number of ^{4}He nuclei required to produce this energy is

$$\Delta N = \frac{E}{\Delta E} \approx \frac{1.1 \times 10^{61}}{4.2 \times 10^{-5}} \approx 2.8 \times 10^{65}\ , \tag{5.3}$$

  which amounts to a Helium-4 mass of

$$M_\mathrm{He} \approx 4m_\mathrm{p}\Delta N \approx 1.9 \times 10^{42}\,\mathrm{g}\ ; \tag{5.4}$$

- assume further that the galaxy's stars were all composed of pure hydrogen initially, and that they are all more or less similar to the Sun; then, the mass in hydrogen was $M_\mathrm{H} \approx 10^{10}\,M_\odot \approx 2 \times 10^{43}$ g initially, and the final Helium-4 abundance by mass expected from the energy production amounts to

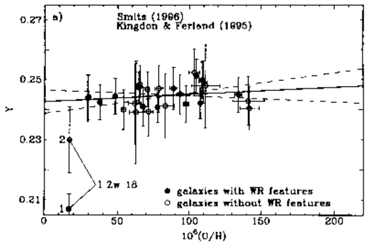
$$Y_* \approx \frac{1.9 \times 10^{42}}{2 \times 10^{43}} \approx 10\%\ , \tag{5.5}$$

  which is much less than the Helium-4 abundance actually observed;

- this discrepancy is exacerbated by the fact that ^{4}He is destroyed in later stages of the evolution of massive stars, and that most of this Helium should be locked up in the centres of stellar remnants (mostly white dwarfs).

- Another key argument against stellar nucleosynthesis as the *main* production route of Helium is that it is observed that Helium-4 abundance is a very weak function of metallicity (i.e., Helium-4 does not increase in lockstep with the other elements produced by stellar nucleosynthesis).

- we thus see that the amount of ^{4}He observed in stars can by no means have been produced by these stars themselves under reasonable assumptions during the lifetime of the galaxies; we must therefore consider that most of the ^{4}He which is now observed must have existed already before the galaxies formed;



From Izotov et al. 1997; the key point is that Helium is a *very weak* function of metallicity, arguing powerfully against stellar nucleosynthesis as its main production route.

5.2.2 Elementary considerations

- nuclear fusion of ^{4}He and similar light nuclei in the early Universe is possible only if the Universe was hot enough for a sufficiently long period during its early evolution; the nuclear binding energies of order $\sim$ MeV imply that at least temperatures of $T \sim 10^6 \times 1.16 \times 10^4\,\mathrm{K} \approx 1.2 \times 10^{10}\,\mathrm{K}$ must have been reached; since the temperature of the (photon background in the) Universe is now $T_0 \sim 3\,\mathrm{K}$ as we shall see later, this corresponds to times when the scale factor of the Universe was

$$a_{\mathrm{nuc}} \sim \frac{3}{1.2 \times 10^{10}} \approx 2.5 \times 10^{-10} \; ; \tag{5.6}$$

- at times so early, the actual mass density and a possible cosmological constant are entirely irrelevant for the expansion of the Universe, which is only driven by the radiation density; thus, the expansion function can be simplified to read $E(a) = \Omega_{\mathrm{r},0}^{1/2} a^{-2}$, and we find for the cosmic time according to (2.11)

$$t(a) = \frac{1}{\Omega_{\mathrm{r},0}^{1/2} H_0} \int_0^a a' \mathrm{d}a' = \frac{a^2}{2\Omega_{\mathrm{r},0}^{1/2} H_0} \approx 4.3 \times 10^{19} a^2\,\mathrm{s} \; , \tag{5.7}$$

  where we have inserted the Hubble constant from (4.18) and the radiation-density parameter today $\Omega_{\mathrm{r},0} \approx 2.5 \times 10^{-5}$, which will be justified later;

- inserting a_{nuc} from (5.6) into (5.7) yields a time scale for nucleosynthesis of order a few seconds; we shall argue later that it is in fact delayed until a few minutes after the Big Bang;

- it is instructive for later purposes to establish a relation between time and temperature based on (5.7); using $T = T_0/a$, we substitute $a = T_0/T$ to obtain

$$t = 4.3 \times 10^{19} \left(\frac{T_0}{T}\right)^2 \text{ s} \approx 1.6 \left(\frac{T}{\text{MeV}}\right)^{-2} \text{ s} ; \qquad (5.8)$$

5.2.3 The Gamow criterion

- a crucially important step in the fusion of ^{4}He is the fusion of deuterium ^{2}H or d,

$$p + n \to d + \gamma \qquad (5.9)$$

  because the direct fusion of ^{4}He from two neutrons and two protons is extremely unlikely;

- If too little deuterium is produced, no ^{4}He is produced because deuterium forms a necessary intermediate step; realising this, Gamow suggested that the amount of deuterium produced has to be "just right", which he translated into the intuitive criterion

$$n_{\text{B}} \langle \sigma v \rangle t \approx 1 , \qquad (5.10)$$

  where n_{p} is the baryon number density, $\langle \sigma v \rangle$ is the velocity-averaged cross section for the reaction (5.9), and t is the available time for the fusion, which we have seen in (5.8) to be set by the present temperature of the cosmic radiation background, T_0, and the temperature T required for deuterium fusion;

- thus, from an estimate of the baryon density n_{B} in the Universe, from the known velocity-averaged cross section $\langle \sigma v \rangle$, and from the known temperature required for deuterium fusion, Gamow's criterion allows us to estimate the present temperature T_0 of the cosmic radiation background; already in the 1940's, Gamow was able to predict $T_0 \approx 5\,\text{K}$!

- summarising, we have arrived at two remarkable arguments so far; first, the observation that the ^{4}He abundance is $Y \approx 25\%$ by mass shows that stars alone are insufficient for the production of light nuclei in the Universe, so we are guided to suggest that the early Universe must have been hot enough for nuclear fusion processes to be efficient; in other words, the observed abundance of ^{4}He indicates that there should have been a hot Big Bang; second, the crucially important intermediate step of deuterium fusion allows an estimate of the present temperature of the cosmic radiation background which lead Gamow already in 1942 to predict that it should be of order a few Kelvin;

- after these remarkably simple and far-reaching conclusions, we shall now study primordial nucleosynthesis and consequences thereof in more detail;

5.2.4 Elements produced

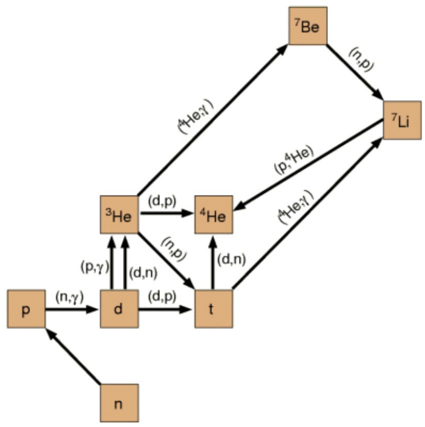
- the fusion of deuterium (5.9) is the crucial first step; since the photodissociation cross section of d is large, destruction of d is very likely because of the intense photon background until the temperature has dropped way below the binding energy of d, which is only 2.2 MeV, corresponding to 2.6×10^{10} K; in fact, substantial d fusion is delayed until the temperature falls to $T = 9 \times 10^{8}$ K or $kT \approx 78$ keV! as (5.8) shows, this happens $t \approx 270$ s after the Big Bang;

- from there, Helium-3 and tritium (^{3}H or t) can be built, which can both be converted to ^{4}He; these reactions are now fast, immediately converting the newly formed d; in detail, these reactions are

$$\begin{aligned} d + p &\rightarrow {}^3\mathrm{He} + \gamma\,, \\ d + d &\rightarrow {}^3\mathrm{He} + n\,, \\ d + d &\rightarrow t + p\,, \quad \text{and} \\ {}^3\mathrm{He} + n &\rightarrow t + p\,, \end{aligned} \tag{5.11}$$

  followed by

$$\begin{aligned} {}^3\mathrm{He} + d &\rightarrow {}^4\mathrm{He} + p \quad \text{and} \\ t + d &\rightarrow {}^4\mathrm{He} + n\,; \end{aligned} \tag{5.12}$$



Nuclear fusion reactions responsible for primordial nucleosynthesis

- fusion reactions with neutrons are irrelevant because free neutrons are immediately locked up in deuterons once deuterium fusion begins, and passed on to t, ^{3}He and ^{4}He in the further fusion steps;

- since there are no stable elements with atomic weight $A = 5$, addition of protons to ^{4}He is unimportant; fusion with d is unimportant because its abundance is very low due to the efficient follow-up reactions; we can therefore proceed only by fusing ^{4}He with t and ^{3}He to build up elements with $A = 7$,

$$\begin{aligned} t + {}^4\mathrm{He} &\rightarrow {}^7\mathrm{Li} + \gamma\,, \\ {}^3\mathrm{He} + {}^4\mathrm{He} &\rightarrow {}^7\mathrm{Be} + \gamma\,, \quad \text{followed by} \\ {}^7\mathrm{Be} + e^- &\rightarrow {}^7\mathrm{Li} + \nu_e\,; \end{aligned} \tag{5.13}$$

  some ^{7}Li is destroyed by

$${}^7\mathrm{Li} + p \rightarrow 2\,{}^4\mathrm{He}\,; \tag{5.14}$$

  the fusion of two ^{4}He nuclei leads to ^{8}Be, which is unstable; further fusion of ^{8}Be in the reaction

$${}^8\mathrm{Be} + {}^4\mathrm{He} \rightarrow {}^{12}\mathrm{C} + \gamma \tag{5.15}$$

is virtually impossible because the low density of the reaction partners essentially excludes that a ^{8}Be nucleus meets a ^{4}He nucleus during its lifetime;

- thus, while the reaction (5.15) is possible and extremely important in stars, it is suppressed below any importance in the early Universe; this shows that the absence of stable elements with $A = 8$ prohibits any primordial element fusion beyond ^{7}Li;

5.2.5 Predicted approximate Helium abundance

- once stable hadrons can form from the quark-gluon plasma in the very early universe, neutrons and protons are kept in thermal equilibrium by the weak interactions

$$\begin{aligned} p + e^- &\leftrightarrow n + \nu_e\ , \\ n + e^+ &\leftrightarrow p + \bar{\nu}_e \end{aligned} \tag{5.16}$$

until the interaction rate falls below the expansion rate of the Universe;

- while equilibrium is maintained, the abundances n_n and n_p are controlled by the Boltzmann factor

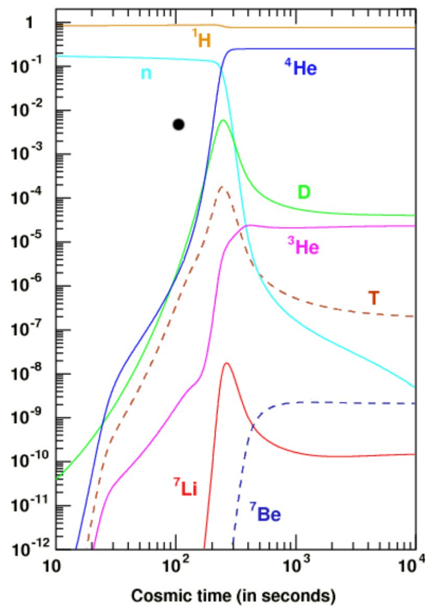
$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{3/2} \exp\left(-\frac{Q}{kT}\right) \approx \exp\left(-\frac{Q}{kT}\right)\ , \tag{5.17}$$

where $Q = 1.3\,\mathrm{MeV}$ is the energy equivalent of the mass difference between the neutron and the proton;

- the weak interaction freezes out when $T \approx 10^{10}\,\mathrm{K}$ or $kT \approx 0.87\,\mathrm{MeV}$, which is reached $t \approx 2\,\mathrm{s}$ after the Big Bang; at this time, the n abundance by mass is

$$X_n(0) \equiv \frac{n_n m_n}{n_n m_n + n_p m_p} \approx \frac{n_n}{n_n + n_p} = \left[1 + \exp\left(\frac{Q}{kT_n}\right)\right]^{-1} \approx 0.17\ ; \tag{5.18}$$

detailed calculations show that this value is kept until $t_n \approx 20\,\mathrm{s}$ after the Big Bang, when $T_n \approx 3.3 \times 10^9\,\mathrm{K}$;



Light-element abundances as a function of cosmic time during primordial nucleosynthesis

- afterwards, the free neutrons undergo β decay with a half life of $\tau_n = 886.7 \pm 1.9\,\mathrm{s}$, thus

$$X_n = X_n(0) \exp\left(-\frac{t - t_n}{\tau_n}\right) \approx X_n(0)\mathrm{e}^{-t/\tau_n}\ ; \tag{5.19}$$

when d fusion finally sets in at $t_d \approx 270\,\mathrm{s}$ after the Big Bang, the neutron abundance has dropped to

$$X_n(t_d) \approx X_n(0)\mathrm{e}^{-t_d/\tau_n} \approx 0.125\ ; \tag{5.20}$$

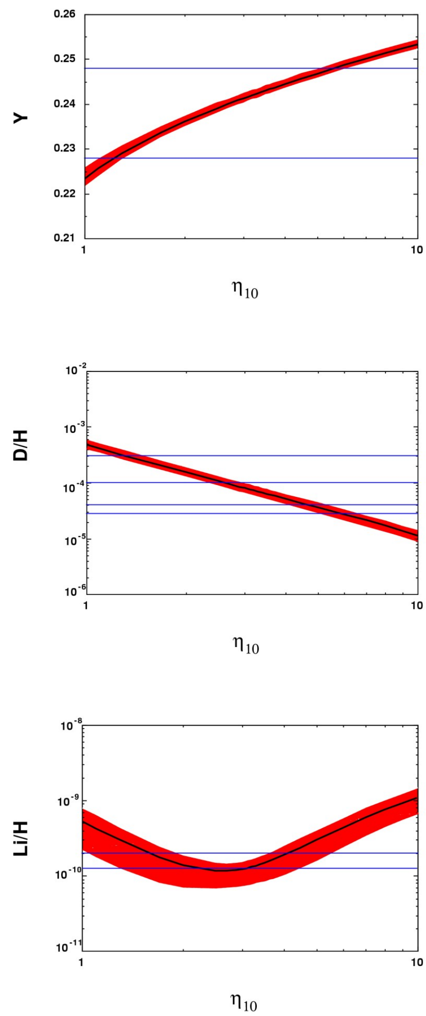
now, essentially all these neutrons are collected into ^{4}He because the abundances of the other elements can be neglected to first order; this yields a ^{4}He abundance by mass of

$$Y \approx 2X_n(t_d) = 0.25 \tag{5.21}$$

because the neutrons are locked up in pairs to form ^{4}He nuclei;

- the Big-Bang model thus allows the prediction that ^{4}He must have been produced such that its abundance is approximately 25% by mass, which is in remarkable agreement with the observed abundance and thus a strong confirmation of the Big-Bang model;

5.2.6 Expected abundances and abundance trends



Dependence of the ^{4}He, d, and ^{7}Li abundances on the parameter η

- the detailed abundances of the light elements as produced by the primordial fusion must be calculated solving rate equations based on the respective fusion cross sections; uncertainties involved concern the exact values of the cross sections and their energy dependence, and the precise life time of the free neutrons;

- since primordial nucleosynthesis happens during the radiation era (which we shall confirm later on), the expansion rate is exclusively set by the radiation density; then, the only other parameter controlling the primordial fusion processes is the baryon density;

- in fact, the only relevant parameter defining the primordial abundances is the ratio between the number densities of baryons and photons; since both densities scale like a^{-3} or, equivalently, like T^3, their *ratio* η is constant; anticipating the photon number density to be determined from the temperature of the CMB,

$$\eta = \frac{n_{\mathrm{B}}}{n_\gamma} = 10^{10}\eta_{10}\ , \quad \eta_{10} \equiv 273\Omega_{\mathrm{B}}h^2\ ; \tag{5.22}$$

  thus, once we know the photon number density, and once we can determine the parameter η from the primodial element abundances, we can infer the baryon number density;

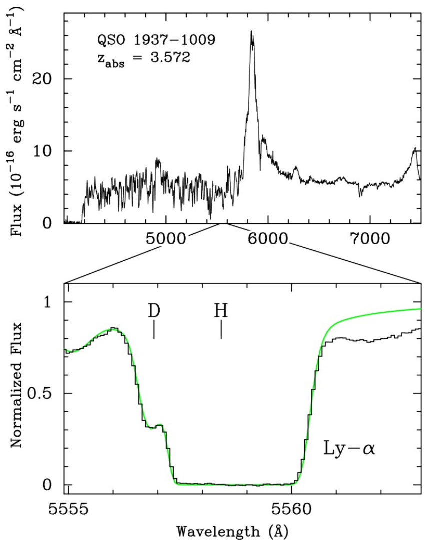
- typical 2-σ uncertainties from cross sections and neutron half-life are, at a fiducial η parameter of $\eta_{10} = 5$, 0.4% for ^{4}He, 15% for d and ^{3}He, and 42% for ^{7}Li;

- the ^{4}He abundance depends only very weakly on the η because the largest fraction of free neutrons is swept up into ^{4}He without strong sensitivity to the detailed conditions;

- the principal effects determining the abundances of d, ^{3}He and ^{7}Li are the following: with increasing η, they can more easily be burned to ^{4}He, and so their abundances drop as η increases; at low η, an increase in the proton density causes ^{7}Li to be destroyed by the reaction (5.14), while the precursor nucleus ^{7}Be is more easily produced if the baryon density increases further; this creates a characteristic "valley" of the predicted ^{7}Li abundance near $\eta \approx (2\ldots3)\times10^{-10}$;

5.3 Observed element abundances

5.3.1 Principles



Deuterium signature in the wing of a damped (saturated) hydrogen absorption line in a QSO spectrum

- of course, the main problem with any comparison between light-element abundances predicted by primordial nucleosynthesis and their observed values is that much time has passed since the primordial fusion ceased, and further fusion processes have happened since;

- seeking to determine the primordial abundances, observers must therefore either select objects in which little or no contamination by later nucleosynthesis can reasonably be expected, in which the primordial element abundance may have been locked up and separated from the surroundings, or whose observed element abundances can be corrected for their enrichment during cosmic history in some way;

- deuterium can be observed in cool, neutral hydrogen gas (HI regions) via resonant UV absorption from the ground state, or in radio wavebands via the hyperfine spin-flip transition, or in the sub-millimetre regime via DH molecule lines; these methods all employ the fact that the heavier d nucleus causes small changes in the energy levels of electrons bound to it;

- Helium-3 is observed through the hyperfine transition in its ion ^{3}He$^+$ in radio wavebands, or through its emission and absorption lines in HII regions;

- Helium-4 is of course most abundant in stars, but the fusion of ^{4}He in stars is virtually impossible to correct precisely; rather, ^{4}He is probed via the emission from optical recombination lines in HII regions;

- measurements of Lithium-7 must be performed in old, local stellar populations; this restricts observations to cool, low-mass stars because of their long lifetime, and to stars in the Galactic halo to allow precise spectra to be taken despite the low ^{7}Li abundance;

5.3.2 Evolutionary corrections

- stars brooded heavy elements as early as $z \sim 6$ or even higher; any attempts at measuring primordial element abundances must therefore concentrate on gas with as low a metal abundance as possible; the dependence of the element abundances on metallicity allows extrapolations to zero enrichment;
- such evolutionary corrections are low for deuterium because it is observed in the Lyman-α forest lines, which arise from absorption in low-density, cool gas clouds at high redshift; likewise, they are low for the measurements of Helium-4 because it is observed in low-metallicity, extragalactic HII regions;
- probably, little or no correction is required for the Lithium-7 abundances determined from the spectra of very metal-poor halo stars (there is quite a debate about this point);
- inferences from Helium-3 are different because ^{3}He is produced from deuterium in stars during the pre-main sequence evolution; it is burnt to ^{4}He during the later phases of stellar evolution in stellar cores, but conserved in stellar exteriours; observations indicate that a net destruction of ^{3}He must happen, possibly due to extra mixing in stellar interiours; for these uncertainties, ^{3}He commonly excluded from primordial abundance measurements;

5.3.3 Specific results

- due to the absence of strong evolutionary effects and its steep monotonic abundance decrease with increasing η, deuterium is the ideal baryometer; since it is produced in the early Universe and destroyed by later fusion in stars, all d abundance determinations are lower bounds to its primordial abundance;
- measurements of the deuterium abundance at high redshift are possible through absorption lines in QSO spectra, which are likely to probe gas with primordial element composition or close to it;
- such measurements are challenging in detail because the tiny isotope shift in the d lines needs to be distinguished from velocity-shifted hydrogen lines, H abundances from saturated H lines need to be corrected by comparison with higher-order lines, and high-resolution spectroscopy is required for accurate continuum subtraction;
- at high redshift, a deuterium abundance of

$$\frac{n_\mathrm{D}}{n_\mathrm{H}} = 3.4 \times 10^{-5} \tag{5.23}$$

relative to hydrogen is consistent with all relevant QSO spectra at 95% confidence level; a substantial depletion from the primordial value is unlikely because any depletion should be caused by *d* fusion and thus be accompanied by an increase in metal abundances, which should be measurable;

- some spectra which were interpreted as having $\lesssim 10$ times the *d* abundance from (5.24) may be due to lack of spectral resolution; the *d* abundance in the local interstellar medium is typically lower

$$\frac{n_\mathrm{d}}{n_\mathrm{H}} \sim (1\ldots 1.5)\times 10^{-5} \tag{5.24}$$

  which is consistent with *d* consumption due to fusion processes; conversely, the *d* abundance in the Solar System, is higher because *d* is locked up in the ice on the giant planets;

- in low-metallicity systems, ^{4}He should be near its primordial abundance, and a metallicity correction can be applied; possible systematic uncertainties are due to modifications by underlying stellar absorption, collisional excitation of observed recombination lines, and the exact regression towards zero metallicity;

- a conservative range is $0.228 \le Y_\mathrm{p} \le 0.248$, and a high value is likely, $Y_\mathrm{p} = 0.2452 \pm 0.0015$;

- observations of the Lithium-7 abundance aim at stars in the stellar halo with very low metallicity; they should have locked up very nearly primordial gas, but may have processed it;

- cool stellar atmospheres are difficult to model, and ^{7}Li may have been produced by cosmic-ray spallation on the interstellar medium;

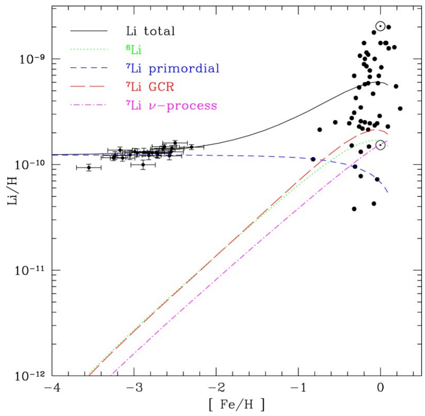
- in the limit of low stellar metalicity, the observed ^{7}Li abundance turns towards the Spite plateau, which is asymptotically independent of metalicity,

$$A(^7\mathrm{Li}) = 12 + \log(n_\mathrm{Li}/n_\mathrm{H}) = 2.2 \pm 0.1\ , \tag{5.25}$$

  and shows very little dispersion; stellar rotation is important because it increases mixing in stellar interiors;

- the Spite plateau is unlikely to reflect the primordial ^{7}Li abundance, but a corrections are probably moderate; a possible increase of ^{7}Li with the iron abundance indicates low production of ^{7}Li, but the probable net effect is a depletion with respect to the primordial abundance by no more than ~ 0.2 dex; a conservative estimate yields

$$2.1 \le A(^7\mathrm{Li}) \le 2.3\ ; \tag{5.26}$$



The Spite plateau in the ^{7}Li abundance as a function of the metalicity

- in absence of depletion, this value falls into the valley expected in the primordial ^{7}Li at the boundary between destruction by protons and production from ^{8}Be; however, if ^{7}Li was in fact depleted, its primordial abundance was higher than the value (5.26), and then two values for η_{10} are possible;

5.3.4 Summary of results

- through the relation $\eta_{10} = 273\,\Omega_\mathrm{B}h^2$, the density of *visible* baryons alone implies $\eta_{10} \geq 1.5$;

- the deuterium abundance derived from absorption systems in the spectra of high-redshift QSOs indicates $\eta_{10} = 4.2\ldots6.3$;

- the ^{7}Li abundance *predicted* from this value of η is then $A(^7\mathrm{Li})_\mathrm{p} = 2.1\ldots2.8$ which is fully consistent with the observed value $A(^7\mathrm{Li}) = 2.1 - 2.3$, even if a depletion by 0.2 dex due to stellar destruction is allowed;

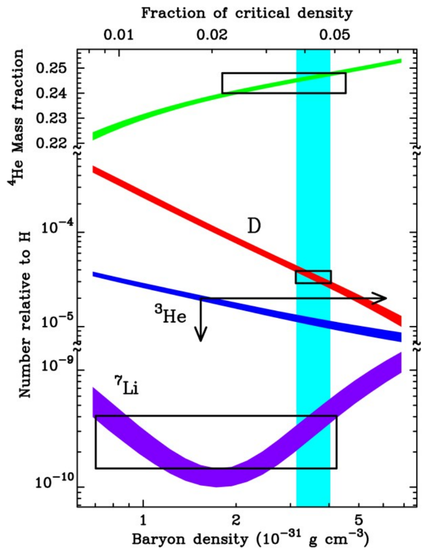
- the *predicted* primordial abundance of helium-4 is then $Y_\mathrm{p} = 0.244\ldots0.250$, which overlaps with the measured value $Y_\mathrm{P} = 0.228\ldots0.248$; thus, the light-element abundances draw a consistent picture for low deuterium abundance; however, this is also true for high deuterium abundance: if $\eta_{10} = 1.2\ldots2.8$, the lithium-7 and helium-4 abundances are $A(^7\mathrm{Li}) = 1.9\ldots2.7$ and $Y_\mathrm{P} = 0.225\ldots0.241$, which are also compatible with the observations;

- we thus find that Big-Bang nucleosynthesis *alone* implies

$$\Omega_\mathrm{B}h^2 = 0.019 \pm 0.0024 \quad \text{or} \quad \Omega_\mathrm{B} = 0.037 \pm 0.009 \tag{5.27}$$

  at 95% confidence level if conclusions are predominantly based on the deuterium abundance in high-redshift absorption systems; we shall later see that this result is in fantastic agreement with independent estimates of the baryon density obtained from the analysis of structures in the CMB;

- a historically very important application of Big-Bang nucleosynthesis begins with the realisation that, at fixed baryon density, the light-element abundances are set by the cosmic expansion rate while the Universe was hot enough to allow nuclear fusion, and that the expansion rate in turn depends on the density of relativistic particle species; a larger number of relativistic species, as could be provided by a number of lepton flavours larger than three, gave rise to a faster expansion, which allowed fewer neutrons to decay until the Universe became too cool for fusion, and thus implied a higher number of neutrons per proton, leading to



Predicted primordial element abundances as a function of η, overlaid with the measurements (boxes). The η parameter compatible with all measurements is marked by the vertical bar.

a higher abundance of ^{4}He; in this way, the ^{4}He abundance was found to limit the number of lepton families to three;

Chapter 6

The Matter Density in the Universe

We saw in the last section that $\Omega_b \sim 0.04$, i.e., 4% of the closure density is in baryons; we will see later that $\Omega_m \sim 0.25 - 0.3$. In this section, we will try to understand where these baryons are at the present day and try to estimate the present-day dark matter density.

6.1 Key concepts

- Constructing a satisfying census of cosmic mass is an incredibly difficult task, and at the end of the lecture no-one will be happy; a key reference for anyone who is interested is Fukugita, Hogan and Peebles 1998, ApJ, 503, 518, and can be downloaded at http://www.mpia-hd.mpg.de/homes/bell/teaching/fhp.pdf
- Almost all baryonic mass is in plasma at the present day.
- Only $\sim$ 10% of baryons are in stars or cold (HI or H_2) gas.
- The other 90% is in warm/hot plasma, either in hot galactic halos, intergalactic space, or in galaxy clusters. While in this state, it is actually pretty hard to observe, so we have only a *very* rough census of these baryons.
- We will see later that there are a variety of ways to estimate the present-day dark matter density; local galaxy motions provide a rough method for their measurement, and galaxy clusters allow two quasi-independent estimates to be made, both suggesting that $\Omega_m \sim 0.3$ (dark and baryonic mass density).

6.2 Baryonic mass in galaxies

6.2.1 Stars

- Given the luminosity of a stellar population, what is its mass? The astronomical community has more-or-less settled on an approach to attack this problem, but there are a number of subtleties/debates in the literature about this issue.

- A key concept is that of the stellar "initial mass function", which describes the distribution of stellar masses of a newly-formed stellar population. A convenient form is the Kroupa (2001) IMF:

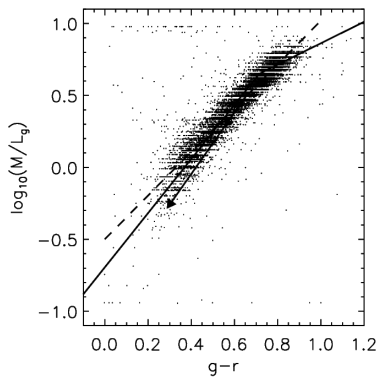
$$\frac{\mathrm{d}N}{\mathrm{d}\ln M} \propto M^{-0.3} \quad 0.1 < M/M_\odot < 0.5 \tag{6.1}$$

$$\frac{\mathrm{d}N}{\mathrm{d}\ln M} \propto M^{-1.3} \quad 0.5 < M/M_\odot < 120. \tag{6.2}$$

  The limits are somewhat arbitrary: the lower mass limit for a star is $0.08M_\odot$ because nuclear hydrogen burning cannot take place below that mass. The upper mass limit of $120M_\odot$ is a number that is widely debated, but for our purposes is not important so long as the upper mass limit is $\gg 10M_\odot$.

- Recall from earlier that $L \propto M^3$ on the main sequence and $L \propto T^6$; thus one can see that M/L is a strong function of temperature (i.e., color);

- Thus, if one *assumes* a universally-applicable form of the stellar IMF (this is a strong assumption), one can use the colors or spectral line indices (which are sensitive to temperature) to estimate stellar M/L



The relationship between $g - r$ color and stellar mass, assuming a universally-applicable stellar IMF. The arrow shows the effect of a large amount of obscuring dust; the solid line shows schematically the result of a different method using spectral line indices.

- For example, Bell & de Jong (2001); Bell et al. (2003) find:

$$\log_{10} M/L \sim -0.4 + 1.1(g - r) \tag{6.3}$$

  where the M/Ls are in solar units and g and r are magnitudes from the SDSS k-corrected to $z = 0$, and the g and r absolute magnitude of the Sun is 5.15 and 4.67 respectively.

- galaxy luminosities and stellar masses are observed to be distributed approximately according to the Schechter function,

$$\frac{\mathrm{d}N}{\mathrm{d}L} = \frac{\Phi_*}{L_*}\left(\frac{L}{L_*}\right)^{-\alpha} \exp\left(-\frac{L}{L_*}\right), \tag{6.4}$$

  where the normalising factor is $\Phi_* \approx 3 \times 10^{-3}\,\mathrm{Mpc}^{-3}$, the scale mass is $M_* \approx 10^{11}\,M_\odot$ and the power-law exponent is $\alpha \approx 1.1$;

- irrespective of what physical processes this distribution originates from, it turns out to characterise mixed galaxy populations very well, even in galaxy clusters;

- the stellar mass density in galaxies is easily found to be

$$\begin{aligned} \mathcal{M}_{\rm g} &= \int_0^\infty M \frac{\mathrm{d}N}{\mathrm{d}M}\,\mathrm{d}M = \Phi_* M_* \int_0^\infty m^{1-\alpha} \mathrm{e}^{-m}\,\mathrm{d}m \\ &= \Gamma(2-\alpha)\Phi_* M_* \approx \Phi_* M_* \approx 3\times 10^8\,\frac{M_\odot}{\mathrm{Mpc}^3}\;; \qquad (6.5) \end{aligned}$$

  or in terms of $\Omega = \rho_*/\rho_{\rm cr}$ with $\rho_{\rm cr} = 9.65\times 10^{30}\,\mathrm{g\,cm^{-3}}$, one obtains $\Omega_* \sim 0.002$.

6.2.2 Cold gas

- Galaxies also contain cold gas (especially spiral galaxies), primarily in the form of HI and H_2 (of course with Helium and metals).

- Direct blind HI surveys can yield an estimate of HI mass density (with e.g., Parkes or Aricebo).

- Molecular hydrogen is a much stickier problem — there are no observable transitions of the *cool* H_2 gas that dominates the gas mass (there are some IR transitions of warm H_2 gas with $T > 100K$, but these are pretty useless as such a small fraction of H_2 is at such warm temperatures). So instead, historically, one has used CO masses (which one can measure) plus a CO-to-H_2 ratio which has been calibrated in the Milky Way (in molecular clouds by comparing the CO mass and a virial mass $M \propto \sigma^2 r$). One then measures the CO fluxes from galaxies (this is hard, the CO line is faint and telescopes that work at this wavelength are not that big, yet) and one uses this (hopefully) representative sample of galaxies to estimate the cosmic H_2 mass. Obviously, this is not a particularly well-posed problem.

- Using up-to-date surveys, incorporated with statistical methods, I estimated (in 2003) a 'cold gas in galaxies' density of $\Omega_{\rm coldgas} \sim 0.0004$.

- Thus, cold gas and stars have $\Omega_{*,\rm gas} \sim 0.0024$, or $< 10\%$ of the available $\Omega_b \sim 0.04$.

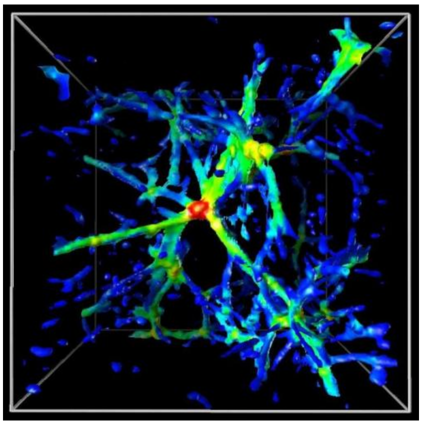
6.2.3 Warm/hot gas in between galaxies

- We have accounted for $< 10\%$ of the baryons expected at $z = 0$ (and even this involved using some dodgy assumptions).

- Where is the rest? Our current idea is that it is in warm/hot intergalactic medium. This is diffuse, ionised filamentary gas that fills out the spaces between galaxies. In clusters of galaxies, the temperatures and densities are hot enough that it is possible to detect via its X-ray emission (see next section). In filaments, the gas is neither hot nor dense enough to emit much in X-rays, and instead must be constrained by detection of absorption line systems in the far-UV or X-ray (very highly ionised oxygen or nitrogen).

- A huge breakthrough in recent times has been the detection of 6-times-ionised oxygen and nitrogen from filaments of the IGM (see attached article by Nicastro 2004) from which $> 1/2$ of the baryonic density of the Universe has been inferred. These lines are *so* faint that one has to wait until a bright flare from a blazar happens to take the spectra (otherwise one needs to integrate on a 'normal' bright X-ray source for $\gg$months).

- I want to give an idea of how extreme this extrapolation is. At typical column densities for detection of $\sim 10^{15} cm^{-2}$, and for an $\sim$AU/pc-sized source, one estimates around 10^{42} or 10^{52} ions were along the line of sight that were detected, corresponding to $\sim$ the mass of an asteroid / the mass of Jupiter. From this small amount of (more-or-less) detected matter, one has extrapolated more than 1/2 of the baryonic density of the Universe!



A schematic diagram of the warm-hot intergalactic medium; the bulk of the gas is in filaments which connect galaxies

6.3 Total mass in galaxies

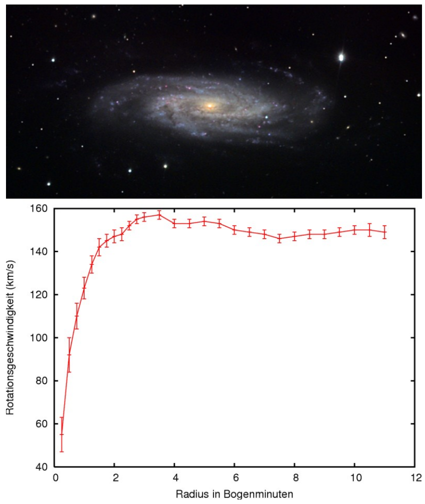
6.3.1 Galaxies

- the rotation velocities of stars orbiting in spiral galaxies are observed to rise quickly with radius and then to remain roughly constant; if measurements are continued with neutral hydrogen beyond the radii out to which stars can be seen, these *rotation curves* are observed to continue at an approximately constant level;

- in a spherically-symmetric mass distribution, test particles on circular orbits have orbital velocities of

$$v_{\rm rot}^2(r) = \frac{GM(r)}{r} \; ; \tag{6.6}$$

flat rotation curves thus imply that $M(r) \propto r$; based on the continuity equation $dM = 4\pi r^2 \rho dr$, this requires that the density falls off as $\rho(r) \propto r^{-2}$ (theory predicts a r^{-3} fall-off at large radii); this is *much* flatter than the light distribution, which shows that spiral galaxies are characterised by an increasing amount of dark matter as the radius increases;



After a quick rise, stellar velocities in spiral galaxies remain approximately constant with radius. (The galaxy shown is NGC 3198.)