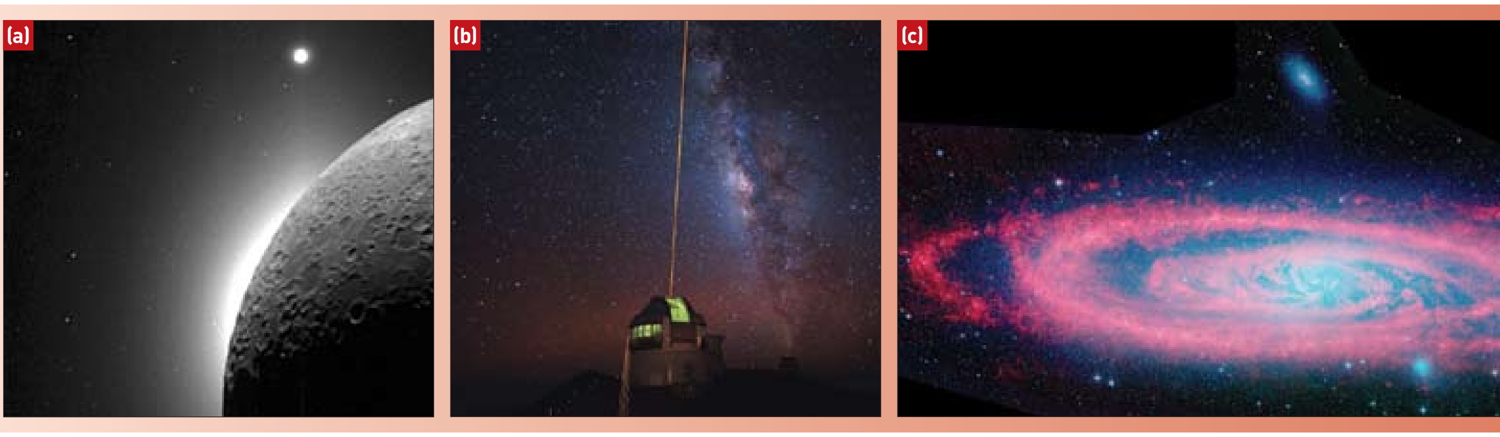


Climbing the cosmolo

In his Presidential Address for 2008, Michael Rowan-Robinson describes the steps taken to extend our knowledge of cosmological distance – towards redshift 1000!

Abstract

Humankind's efforts to measure the distances of the planets, stars and galaxies are closely bound up with the evolution of our ideas about the universe we find ourselves in. This link stretches from classical times to today, with the very latest analysis of the fluctuations in the cosmic microwave background.

Aristotle (384–322 BC) was the first to estimate the size of the Earth, using the angle of the shadow of a pole at noon at a location 100 miles south of the equator. Eratosthenes and Poseidonius later used a similar method. All these estimates are within about 10% of the modern value. In the 2nd century BC Hipparcos used an eclipse method to estimate the distance of the Moon and deduced a value 59 R_E, compared to the modern value of 60.3 R_E. Aristarcos tried to estimate the distance of the Sun using an eclipse method, but was out by a factor of 20. The Greeks also gave us Euclidean geometry (Euclid 300 BC), the idea of absolute, uniform time (Aristotle), and the idea of an infinite physical frame (the atomists, Epicurus). Interestingly, and contrary to the picture held by medieval thinkers, Aristotle believed that the stars were at a range of distances.

A discovery of Copernicus (1473–1543) that is less well-known than his heliocentric system is that he worked out, for the first time, the correct relative distances of the Sun and planets. His values were within 5% of the modern values. The absolute scale of the solar system was not determined accurately till the 19th century.

The Copernican picture also immediately implied a much greater distance for the "immovable" stars. Newton tried, unsuccessfully, to estimate the distances of stars through their brightness, but the first step on the distance ladder outside the solar system was taken by Bessel in 1838 when he measured the parallax of 61 Cyg, its change in apparent direction on the sky due to the Earth's orbit round the Sun. This was the final proof of the Copernican system. Bradley had discovered aberration, the elliptical motion of all stars on the sky due to the Earth's motion, a century earlier.

Cepheids, M31 and the Hubble Law

The next crucial step on the distance ladder, still of prime importance today, was the discovery by Henrietta Leavitt in 1912, working at the Harvard Observatory, that the periods of Cepheid variable stars in the Small Magellanic Cloud are related to their luminosity: the period–luminosity relation. In 1924 Edwin Hubble used Leavitt's discovery to estimate the distance of M31, the Andromeda Nebula. It clearly lay far outside our Milky Way Galaxy, thus resolving the long-standing controversy about the spiral nebulae and opening up the universe of galaxies. Three years later Hubble announced, based on the distances of 18 galaxies, that the more distant a galaxy, the faster it is moving away from us (the Hubble Law):

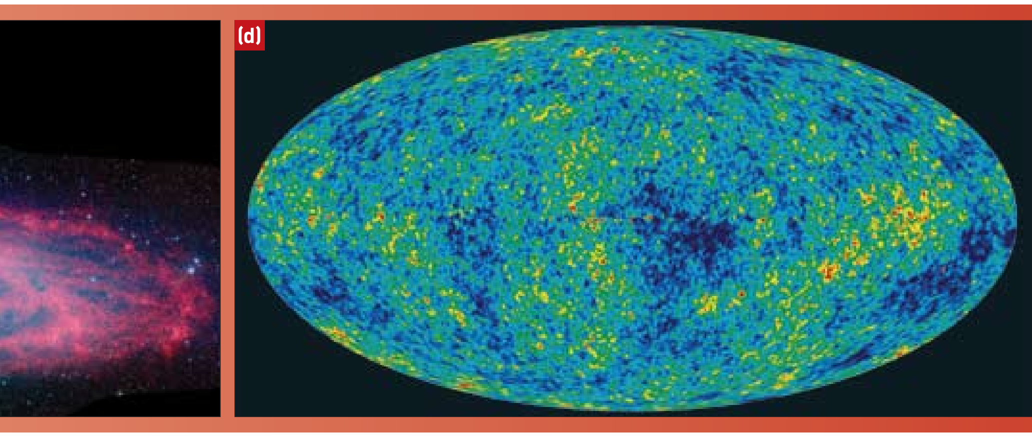
$$\text{velocity/distance} = \text{constant}, H_0 \qquad (1)$$

This is just what would be expected in an expanding universe. Aleksandr Friedmann had shown in 1922 that expanding universe models are what would be expected according to Einstein's General Theory of Relativity, if the universe is (a) homogeneous (everyone sees the same picture) and (b) isotropic (the universe looks the same in every direction).

This unlikely assumption, the cosmological principle, had been introduced by Einstein in 1917 when he derived a static model of the universe in which gravity is balanced by a new force, the cosmological repulsion. Einstein's inspired guess that the universe must be very simple (homogeneous and isotropic) is confirmed to very high accuracy today.

Controversy over H_0

Hubbles's estimate of H_0 was 500 km/s/Mpc. Now H_0 has the dimensions of time^{-1} and so $1/H_0$ is the expansion age of the universe, the age the universe would have if no forces were acting. Hubble's value for H_0 implied an age of the universe of 2 billion years and it was soon realized this was shorter than the age of the Earth as derived from radioactive isotopes. From 1927 to 2001 the value of the Hubble constant was a matter of fierce controversy. Baade pointed out in 1952 that there were two different types of Cepheid, so Hubble's calibration had been incorrect. This reduced H_0 to 200 km/s/Mpc. In 1958 Sandage recognized that objects that Hubble had thought were the brightest stars in some of his galaxies were in fact HII regions and arrived at the first recognizably modern value of H_0 of 75 km/s/Mpc. During the 1970s there was an acute disagreement between Sandage and Tammann, on the one hand, favouring $H_0 = 50$, and de Vaucouleurs, on the other, favouring 100 km/s/Mpc. This disagreement stimulated me to write my monograph *The Cosmological Distance Ladder* (1985), in which I set out to review all aspects of the distance ladder and to reconcile the systematic differences in distance estimates from different methods. With an objective weighting scheme based on quoted errors, and with higher weight for purely



1: Looking out from Earth into the universe. (a) The solar system: a Clementine image of Venus and the solar corona behind the Moon lit by earthshine (NASA/JPL/USGS). (b) The galaxy: the Milky Way seen above the dome of the Gemini North telescope on Hawaii, taken during commissioning of the laser guide star system (P Michaud/Gemini Observatory). (c) Other galaxies: M31, the Andromeda Galaxy in infrared, from the Spitzer Space Telescope (NASA/JPL-Caltech/P Barmby, Harvard-Smithsonian CfA). (d) The universe as a whole: fluctuations in the cosmic microwave background measured by the Wilkinson Microwave Anisotropy Probe satellite. (WMAP)

gical distance ladder

geometric distance methods (or those based on theoretical arguments), I concluded that there were systematic errors in the Type Ia supernova method (too high distances) and in the Tully-Fisher and HII region methods (too low) and that the best overall value was

$$H_0 = 67 \pm 12\,\text{km/s/Mpc}.$$

$H_0 = 67$ would give an expansion age for the universe of 15.3 billion years (Gyr). In the simplest, Einstein de Sitter ($\Omega_m = 1$, $\Lambda = 0$) model, with only gravity acting to slow the expansion, the age of the universe would be 10.2 billion years. This could be compared with ages of the oldest stars in globular clusters, between 10 and 15 Gyr. Chaboyer *et al.* (1998) estimated 12.6 ± 1.1 Gyr, and the age of the galaxy derived from radioactive isotope abundances was also 10–15 Gyr. Was this already a headache for the Einstein de Sitter model?

HST Key Project

Following the launch of the Hubble Space Telescope (HST) in 1990, and the subsequent repair mission, substantial amounts of HST time were dedicated to measuring Cepheids in galaxies out to distances of 20 Mpc, to try to measure the Hubble constant accurately and to give the different distance methods a secure and consistent calibration. The HST Key Project soon split into two teams, one led by Wendy Freedman, Jeremy Mould and Rob Kennicutt (Kennicutt *et al.* 1995), and the other by Allan Sandage and Gustav Tammann. In 2001 Freedman *et al.* announced their final result

$$H_0 = 72 \pm 8\,\text{km/s/Mpc} \qquad (2)$$

This, as we shall see, agreed extremely well with the first results from the WMAP CMB mission (72 ± 5 km/s/Mpc, Spergel *et al.* 2003). It gave an age of the universe for an Einstein de Sitter model of 9.1 Gyr, which meant that a positive cosmological constant would be required for constancy with the age of the oldest stars. As we shall see later, evidence from Type Ia supernovae presented in 1998 was already supporting the idea of a positive cosmological constant. But it was still of interest to see whether there were any possible doubts about this HST Key Project value for H_0. The uncertainties in this value are (a) the distance of the Large Magellanic Cloud, which remains uncertain by 10%, (b) the adopted Cepheid calibration, based on OGLE Cepheids, (c) corrections for the effects of dust extinction, (d) corrections for differences in metallicity between the LMC and the Cepheid host galaxies, (e) corrections for the local peculiar velocity flow. Using the Freedman *et al.* data, my own best estimates for these corrections and the weighting scheme of *The Cosmological Distance Ladder* (1985), I concluded (Rowan-Robinson 2000)

$$H_0 = 63 \pm 6\,\text{km/s/Mpc}.$$

Type Ia supernovae

In 1998 two teams announced that using Type Ia supernovae as standard candles out to significant redshifts (~0.5) implied that the cosmological constant Λ had to be greater than 0 (Riess *et al.* 1998, Perlmutter *et al.* 1999). There were issues with (a) the treatment of extinction by dust, and (b) the consistency of the assumed correlation of the luminosity at maximum light with the exponential decline rate after maximum raised by Liebundgut (2001), Rowan-Robinson (2002). I also raised two other issues: (c) inconsistencies with earlier Type Ia supernova data, (d) inappropriate use of supernovae not observed before maximum light. A group which combined members of the high redshift supernova team and the HST Key Project team announced a value for H_0 from Type Ia supernovae of 68 ± 5 km/s/Mpc (Gibson *et al.* 2000).

The supernova data is clearly excellent and the latest results (Astier *et al.* 2006, Riess *et al.* 2007), reaching out to redshift 1.5, are extremely impressive. A problem with the Gibson *et al.* (2000) analysis was that it used photographic magnitudes for some of the older supernovae. Riess *et al.* (2005) used new HST–ACS observations of Cepheids in galaxies with well-observed recent Type Ia supernovae and concluded that $H_0 = 73 \pm 6$ km/s/Mpc. This analysis also demonstrated that some of the inconsistencies with earlier Type Ia supernovae can be attributed to systematic errors in the photographic magnitudes. The issue of the luminosity–decline rate relation has been addressed by Jha *et al.* (2007) and by Nobili *et al.* (2005) and Wang *et al.* (2006). There are still some unresolved inconsistencies in the derivation of extinction, which can only be resolved with the use of more photometric bands in future supernova studies.

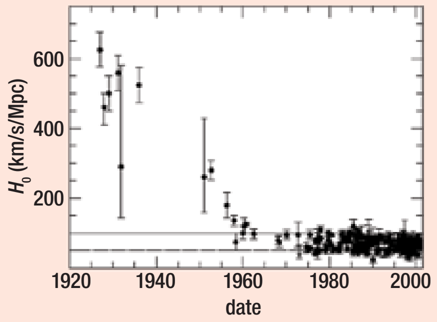
A consensus?

With the WMAP three-year results yielding $H_0 = 73 \pm 3$ km/s/Mpc (Spergel *et al.* 2007), it looks as though we have a consensus around $H_0 = 73$ km/s/Mpc, $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, and an age of the universe 13.7 Gyr. However, in 2006 Sandage *et al.* announced the results of their HST programme, with

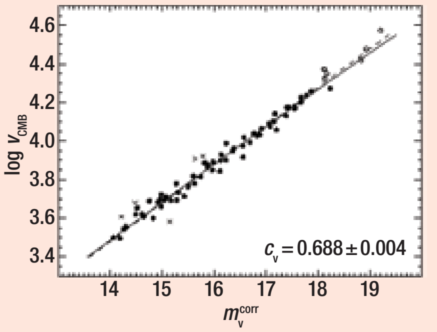
$$H_0 = 62 \pm 5\,\text{km/s/Mpc} \qquad (3)$$

This was based on a new extensive study of the Cepheid period–luminosity relation (Tammann *et al.* 2003), and recognition that there is a difference between the P–L relation in the galaxy and the LMC (Sandage *et al.* 2004). They used a new Cepheid calibration based on the Baade–Wesselink expanding photosphere method, so do not incur the uncertainty in the LMC distance. And they give a new discussion of extinction in supernovae. In my view this is an analysis that has to be taken very seriously. I will discuss below whether this is inconsistent with the WMAP CMB estimate.

Mike Feast has presented a recent review of work on H_0 (Feast 2007). New HST Cepheid



2: Estimated values of H_0 from 1927–2001. (Copyright SAO)



3: Hubble diagram for Type Ia supernovae. (From Sandage *et al.* 2006)

Table 1: CMB fluctuation results for H_0

Data set	assumptions/other data	H_0	reference
Boomerang, Maxima	flat universe	75±10	Jaffe *et al.* 2001
WMP first year		72±5	Spergel *et al.* 2003
WMP first year	+SLOAN LSS data	68±10	Tegmark *et al.* 2004
WMP first year	+BAO data	65±4.5	Eisenstein *et al.* 2005
WMAP three-year data		73±3	Spergel *et al.* 2007
WMAP three-year data	+LSS, BAO	69–72	Spergel *et al.* 2007
WMAP five-year data	+LSS, BAO	70.1±1.3	Komatsu *et al.* 2008

distances by Benedict *et al.* (2007) and revised Hipparcos parallaxes result in a revision of Sandage *et al.*'s H_0 value from 62 to 69.6 km/s/Mpc (van Leeuwen *et al.* 2007). The Freedman *et al.* value is also increased. Macri *et al.* (2006) have shown that the Cepheid distance to NGC 4258 is consistent with a geometrical estimate derived from maser emission. The latest H_0 estimates from the gravitational lens time-delay method are 68 ± 10 (Oguri 2007), 72 ± 10 (Saha *et al.* 2006), and from the Sunyaev–Zeldovich method for clusters of galaxies are 66 ± 14 (Jones *et al.* 2005) and 76 ± 10 km/s/Mpc (Bonamente *et al.* 2006).

The gravitational lens time-delay and Sunyaev–Zeldovich methods offered the prospect of completely independent geometrical methods which could be applied at high redshift, thereby overcoming any uncertainty due to the peculiar velocities of local galaxies. The gravitational lens time-delay method uses double images of quasars caused by gravitational lensing by an intervening galaxy. If the background quasar varies its light output, the two images will be seen to vary out of phase because of the different time it takes the light to arrive via the two different routes. The time delay can then be used to estimate the distance of the quasar. The Sunyaev–Zeldovich method is based on the fact that very hot X-ray emitting gas in rich clusters of galaxies interacts with the photons of the cosmic microwave background to produce either a brightening or dimming at microwave wavelengths. A combination of the microwave and X-ray data allows the distance to be estimated if the gas cloud is assumed to be spherical and smooth.

Unfortunately both methods appear to have irreducible systematic uncertainties. In the case of gravitationally lensed systems we have to know the exact distribution of matter, including dark matter, in the foreground lensing galaxy. In the case of the S–Z method we can not be sure that the gas clouds are spherical and there is a strong possibility that the gas is clumpy.

CMB first Doppler peak and BAOs

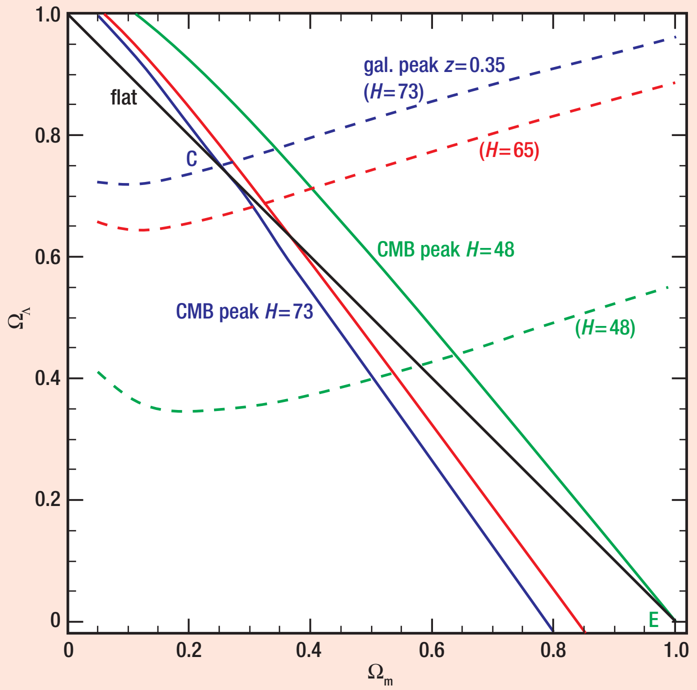
Finally I want to discuss the distance method that takes us to a redshift of 1100, the angular scale of the first Doppler peak in the CMB fluctuations. If we think we know the physics of the universe at the time of decoupling of matter from radiation (epoch of "recombination"), then we know the sound speed in the universe at that time and hence the linear scale of the acoustic horizon. This will translate to the angular scale of the largest structures in the CMB fluctuations, the first Doppler peak, depending on H_0 and the cosmological model. Analysis of the CMB fluctuations usually proceeds by fitting the whole CMB fluctuation spectrum, with some assumptions about the primordial density fluctuation spectrum (usually that it is a power-law, sometimes with the further restriction that the power-law index has the Harrison–Zeldovich scale-free value $n = -1$), the spatial curvature (often taken to be zero) and requiring consistency with other astrophysical data (Type Ia supernovae, large-scale structure). Results of some of these CMB analyses are summarized in table 1.

Spergel *et al.* (2003) show that with the assumption of a power-law spectrum, but no restriction to flat models, consistency with the WMAP (year 1) fluctuation spectrum can be achieved with a wide range of cosmological models and values for H_0. Priors on H_0 or the assumption of flatness then force us to the $\Omega_\Lambda = 0.75$ consensus model. However, if we also drop the assumption of a power-law primordial density fluctuation spectrum, which is not necessarily expected in a universe that has been through a series of phase transitions, the possibilities are opened up even further. Blanchard *et al.* (2003) showed that if we relax the assumption of a power-law to the simplest alternative, a broken power-law, then we can fit the CMB fluctuation spectrum just as well as the consensus model with an $\Omega_m = 1$, $\Lambda = 0$ (Einstein de Sitter) model, provided H_0=46. We can also get consistency with galaxy large-scale structure data provided there are one or more neutrinos with a mass of a few keV, such that $\Omega_\nu \sim 0.2$ (a mixed dark matter model). However, this model is inconsistent with the Type Ia supernova data and H_0=46 is 3σ from the direct HST Key Project estimates. Recently Shafieloo and Souradeep (2007) have confirmed that the low-H_0, Einstein de Sitter model, is as good a fit to the WMAP CMB fluctuation spectrum as the consensus model if the primordial density fluctuation spectrum is allowed to have a free form.

So the CMB fluctuations do not on their own determine H_0. An important advance is the discovery of the baryon acoustic oscillation (BAO) feature in the power spectrum of galaxy density fluctuations (Eisenstein *et al.* 2005, Cole *et al.* 2005). This feature is essentially the same acoustic horizon scale seen in the CMB fluctuations, but now seen in galaxy redshift surveys at $z \sim 0.35$. At this epoch it has a linear scale of about 150 Mpc. Blanchard *et al.* (2006) admit that this feature, if confirmed (it is between about 2 and 3σ significance at the moment), would be fatal for their low-H_0, Einstein de Sitter model. The combination of the CMB first Doppler peak and the baryon acoustic oscillation peak is the ultimate geometric measurement of H_0. Using the WMAP five-year data combined with baryon acoustic oscillation and Type Ia supernova data, Komatsu *et al.* (2008) conclude that $H_0 = 70.1 \pm 1.3$ km/s/Mpc.

Figure 4 shows the result of applying these two tests as a pure diameter–distance test. The black line shows the locus of a zero-curvature universe. The solid curves show the loci which give the same observed angular scale for the first Doppler peak, for H_0 = 73, 65 and 48 km/s/Mpc. We see that the H_0 = 48 curve intersects the zero-curvature line at the $\Omega_m = 1$, $\Lambda = 0$ Einstein de Sitter model, consistent with the Blanchard *et al.* (2003) and Shafieloo and Souradeep (2007) claims.

The broken curves show loci for the same three values of H_0 for models that give the same observed angular diameter for the baryon acoustic oscillation peak at $z = 0.35$. The H_0 = 65



4: Models with the same angular size CMB Doppler peak (solid blue, red and green, for H_0=73, 65, 48), and the BAO feature (broken curves). C marks the "consensus" model, E the Einstein de Sitter model. CMB data alone cannot separate C and E, without prior knowledge of *H* or the shape of the spectrum of initial density fluctuations. The BAO feature can break this degeneracy, but models with H_0=65–73 are permitted.

Some formulae

radius of particle horizon at decoupling

$r_{ph} = R(t_{dec})\chi_{ph}$

$\chi_{ph} = A^{1/2}\int_0^{1/Z_{dec}} \{\Omega_0 x + \Omega_r + \Omega_\Lambda x^{4-3(1+w)} + (1-\Omega_0-\Omega_r-\Omega_\Lambda)x^2\}^{-1/2} dx$

$A = |(1-\Omega_0-\Omega_r-\Omega_\Lambda)|$ if $k=+1, -1$,

$=1$ if $k=0$

$z_{dec} \sim 1100$, $w=-1$

radius of acoustic horizon

$r_{acoust} = r_{ph}/\{3(1+3\rho_b/4\rho_\gamma)\}^{1/2} = r_{ph}/\{3(1+1.25(\Omega_b h^2)/(\Omega_0 h^2)\}^{1/2}$

$\Omega_b h^2 \sim 0.022$ (Doppler peak ratios + nucleosynthesis)

angular radius of first Doppler peak

$\theta_{Doppler} = r_{acoust}/D_{diam}(z_{dec})$

angular radius of baryon acoustic peak

$\theta_{BAO} \sim 150(\Omega_0 h^2/0.25\times0.73^2)^{-0.0853}\,\mathrm{Mpc}/D_{diam}(z)$

diameter distance

$D_{diam}(z) = ct_0 r_o(z)/\{A^{1/2}(1+z)\}$

$ct_0 = 9.8h^{-1}$ Gyr

$r_o(z) = \sin\chi(z)$ for $k=+1$,

$=\chi(z)$ for $k=0$,

$=\sinh\chi(z)$ for $k=-1$

$\chi(z) = A^{1/2}\int_1^{1/(1+z)} \{\Omega_0 x + \Omega_r + \Omega_\Lambda x^{4-3(1+w)} + (1-\Omega_0-\Omega_r-\Omega_\Lambda)x^2\}^{-1/2} dx$

locus passes close where the corresponding first Doppler peak locus intersects the zero-curvature line. However, the H_0=48 locus lies nowhere near the Einstein de Sitter model. The conclusion is that H_0=65 is consistent with this combined test, but H_0=48 is ruled out.

Conclusions

- Local direct estimates of H_0 are in the range 62–72 km/s/Mpc, with an uncertainty of 10%, and this is a big advance on the range 50–100 possible in the 1970s. My 1985 value of 67 still looks quite plausible.
- The CMB fluctuation estimates of H_0 lie in the range 65–73 km/s/Mpc, depending on assumptions made, with an uncertainty of 2%.
- The angular scale of the baryonic acoustic oscillation peak, combined with the CMB first Doppler peak, is the ultimate geometrical measurement of H_0.
- It is still worthwhile to improve the direct local estimates of H_0. If an accuracy of, say, 1%, could be achieved then there is the prospect of learning about new physics beyond the standard model of particle physics, since tension between direct and CMB estimates would show that the underlying assumptions being used in CMB physics were incorrect. Such an accuracy could be achieved through:

(a) Direct parallax measurements of the distance to the LMC, from the GAIA mission. This is the main uncertainty in most current local estimates, accounting for 10% uncertainty in H_0.
(b) Use of the Baade–Wesselink expanding atmosphere method for Cepheids and supernovae. For Cepheids this would reduce the uncertainty in the absolute calibration. To generate accurate atmospheric models for Type Ia supernovae is challenging, but would improve confidence in the method enormously and would eliminate the need for *ad hoc* corrections for the luminosity dependence on decline rate.
(c) Use of multiwavelength photometry, especially in the infrared, to control extinction and metallicity. Further work on estimating distances via Cepheids and supernovae really does not seem worthwhile without this development.
(d) Much better mapping of the local density field would be needed to reduce the uncertainty in the peculiar velocities of galaxies.

In conclusion, progress in understanding the universe has always been strongly connected with our ability to measure distances. Today we have distance measurements to redshift 1100, the epoch when matter and radiation finally decoupled at the end of the hot Big Bang phase. Apparently we have reached a precision of 2% in our measurements of the Hubble constant and a consensus model of the universe with a dominant role for dark energy. Our inability to provide a motivation for this dark energy remains troubling and we should remain open to the possibility of new physics beyond the standard model that might change our whole picture of the universe. ●

Michael Rowan-Robinson, President of the RAS 2006–2008; Imperial College, London.

References

Astier P *et al.* 2006 *AA* **447** 31.
Benedict G F *et al.* 2007 *AJ* **133** 1810.
Blanchard A *et al.* 2003 *AA* **412** 35.
Blanchard A *et al.* 2006 *AA* **449** 925.
Bonamente M *et al.* 2006 *ApJ* **647** 25.
Chaboyer B *et al.* 1998 *ApJ* **494** 96.
Cole S *et al.* 2005 *MNRAS* **362** 505.
Eisenstein D J *et al.* 2005 *ApJ* **633** 560.
Feast M 2007 *From IRAS to Herschel/Planck: Cosmology with Infrared Surveys*.
Freedman W L *et al.* 2001 *ApJ* **553** 47.
Gibson B K *et al.* 2000 *ApJ* **529** 723.
Jaffe A H *et al.* 2001 *Phys.Rev.Lett.* **86** 3475.
Jha S *et al.* 2007 *ApJ* **659** 98.
Jones M E *et al.* 2005 *MNRAS* **357** 518.
Kennicutt R C *et al.* 1995 *AJ* **110** 1476.
Komatsu E *et al.* 2008 *ApJS* astro-ph/0803.0547.
Liebundgut B 2001 *ARAA* **39** 67.
Macri L M *et al.* 2006 *ApJ* **652** 1133.
Nobili S *et al.* 2005 *AA* **437** 789.
Oguri M 2007 *ApJ* **660** 1.
Perlmutter S *et al.* 1999 *ApJ* **517** 565.
Riess A G *et al.* 1998 *AJ* **116** 1009.
Riess A G *et al.* 2005 *ApJ* **627** 579 .
Riess A G *et al.* 2007 *ApJ* **659** 98.
Rowan-Robinson M 1985 *The Cosmological Distance Ladder* (W H Freeman, New York).
Rowan-Robinson M 2000 astro-ph/001206.
Rowan-Robinson M 2002 *MNRAS* **332** 532 .
Saha A *et al.* 2006 *ApJS* **165** 108.
Sandage A *et al.* 2004 *AA* **424** 43.
Sandage A *et al.* 2006 *ApJ* **653** 843.
Shafieloo A and Souradeep T 2007 astro-ph/0709.1944.
Spergel D N *et al.* 2003 *ApJS* **148** 175.
Spergel D N *et al.* 2007 *ApJS* **170** 377.
Tammann G A *et al.* 2003 *AA* **404** 423.
Tegmark M *et al.* 2004 *ApJ* **606** 702.
van Leeuwen F *et al.* 2007 *MNRAS* in press astro-ph/0705.1592.
Wang L *et al.* 2006 *ApJ* **641** 50.