The cosmic distance scale and the Hubble constant

M. Fukugita, C. J. Hogan & P. J. E. Peebles

Although astronomers can reliably measure the relative distances of distant galaxies, attempts to calibrate the absolute cosmic distance scale remain controversial. Nevertheless, the sources of possible error are now clearly defined and a convincing result seems to be within reach.

In a perfectly uniform and isotropic expanding universe, particles move according to the Hubble law, $v = H_0 r$, where v and r are the relative velocity and separation of any two particles. The value of the Hubble constant, H_0, fixes the fundamental scale of length and time of the cosmic expansion. Precise knowledge of H_0 is needed for a quantitative test of virtually any aspect of cosmological theory: it bears on the predicted age of the Universe, the problem of the cosmological constant, comparisons of Big Bang nucleosynthesis with observations of matter density, predictions for dark-matter mass density and the confrontation of structure formation theories with anisotropies in the thermal cosmic background radiation.

The global value of H_0 has long been uncertain by about a factor of two[1,2], largely because absolute distances are easily measured only on small scales, where inhomogeneous gravitational acceleration causes motions to depart significantly from the simple expansion described by the Hubble law. The uncertainty arises from the disagreement of different techniques used to connect the local distances to the smooth large-scale Hubble flow.

On local scales, there is a consensus regarding the distance to the nearest galaxy outside our own, the Large Magellanic Cloud (LMC), at $r = 50 \pm 3$ kpc. At least four independent measures agree; the expansion of the photosphere of supernova 1987A (ref. 3), the parallax of its light echo[4], and the two types of well characterized variable stars whose absolute luminosity can be inferred from their period, Cepheids and RR Lyrae stars[5]. The LMC is so close that it simply orbits our Galaxy; its radial velocity has nothing to do with the Hubble expansion. Nevertheless, the LMC measurements provide a validation of the Cepheids, which then give reliable distances to other nearby 'calibrator' galaxies. The most distant reliable calibrator is M81, with a distance (now accurately measured by the Hubble Space Telescope[6]) of 3.63 ± 0.34 Mpc.

The convergence to uniform Hubble flow on the very largest scales is most clearly demonstrated using the (remarkably

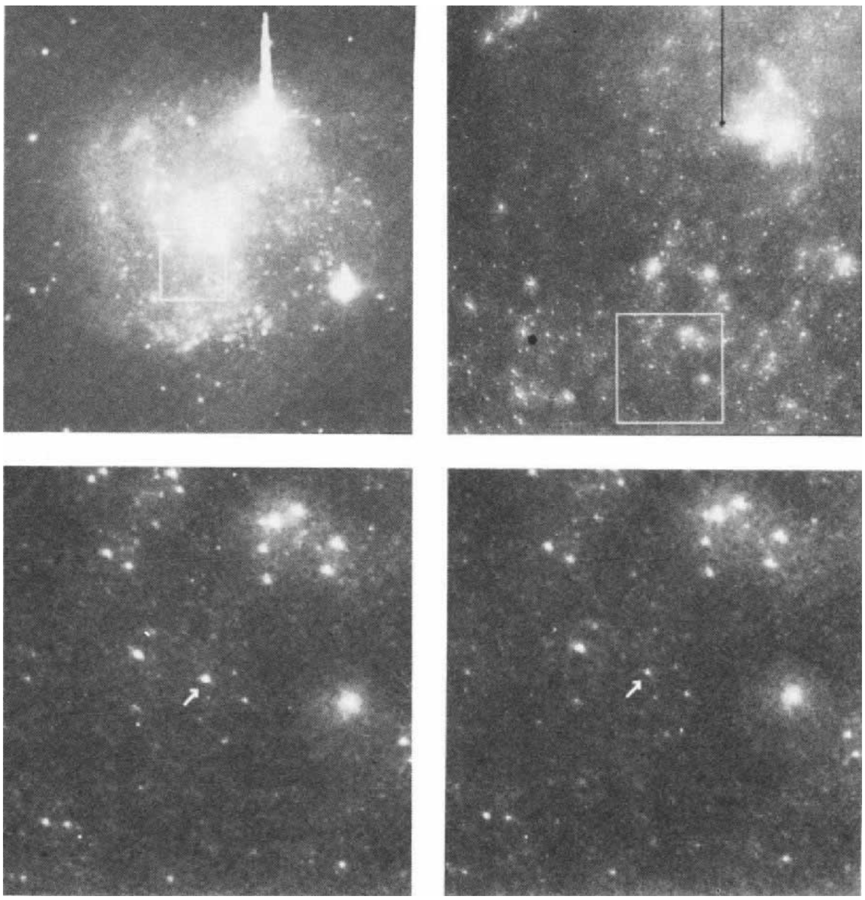

FIG. 1 Images of the small local galaxy IC4182. The two upper frames show the galaxy. The box in the left frame is enlarged to give the right frame; the box in the right frame shows the area enlarged in the two lower frames. The arrows in the lower frames show a Cepheid variable star in this galaxy at a bright and a faint stage in its luminosity variation. For such stars, the period of the light variation gives an accurate estimate of its intrinsic brightness and therefore of its distance. This galaxy is of special importance because it was the site of a type Ia supernova, 1937C. Measurements last year of 27 Cepheid variables in this galaxy by the Hubble Space Telescope[13] (HST) gave the first reliable distance estimate for IC4182, and hence the first reliable direct determination of intrinsic brightness for a type Ia supernova. This calibration gives the distance to other supernovae farther away, allowing a direct estimate of the Hubble constant. The HST has now measured Cepheids in other galaxies, including another supernova host galaxy (NGC5253) and the important local calibrator M81. A key project is underway to measure direct Cepheid distances to galaxies in the Virgo galaxy cluster, eliminating some of the most controversial intermediate stages in establishing the cosmic distance scale. (Images courtesy of Hubble Space Telescope News, Space Telescope Science Institute.)

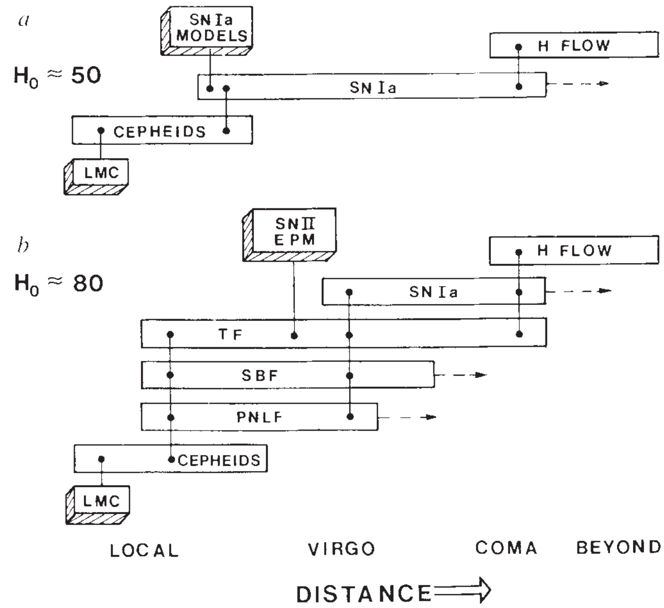

FIG. 2 *a*, *b*, Two cosmological distance ladders. Horizontal bars indicate the distance range for each method as discussed in the text, vertical bars indicate direct links between methods. Absolute calibration is established through Cepheid variable-star distances, by models of the type Ia supernova luminosity (SNIa), or model fits to the expanding photospheres of type II supernovae (SNII). Cepheid calibrations are currently available only for galaxies in or near the Local Group of galaxies. Precise measurement of H_0 however requires connecting these to the large-scale smooth Hubble flow. (LMC, Large Magellanic Cloud; TF, Tully–Fisher; SBF, surface brightness fluctuations; PNLF, planetary nebula in a galaxy; EPM, expanding photosphere method; H FLOW, Hubble flow.)

uniform) brightest galaxies in clusters[7,8], which show that fractional departures from the Hubble law $\delta v/v \lesssim 10\%$ for $v \gtrsim 10{,}000$ km s^{-1}. Distance ratios between clusters obtained from the Tully–Fisher relation[9] (see below) and other techniques[10] tie individual galaxy clusters to this cosmic reference frame and reliably establish their true 'Hubble velocity', separated from peculiar non-Hubble motions. For example, the Hubble velocity at the distance of the Coma cluster of galaxies is determined to be $v \approx 7{,}100$ km s^{-1}. If we only knew the distance to the Coma cluster, the value of H_0 would be accurately determined.

The nearest cluster tied into this system is the Virgo cluster of galaxies, whose distance ratio to Coma (~0.18) is known to ~10% accuracy. Again, if we knew the absolute distance to Virgo, we would know the global value of H_0 without problems connected with peculiar velocities.

But the distance to Virgo is uncertain by a large factor: different arguments give a distance to the cluster core of between 14 and 22 Mpc. The current debate, and our discussion, thus focus on the absolute calibration of a few of the most reliable distance measures which bridge the gap at intermediate distances between a few megaparsecs and a few tens of megaparsecs—that is, between the local reliable calibrations of distance (such as M81) and the distant reliable calibrators of Hubble velocity (such as the Virgo cluster). Although this gap is closing, current techniques appear to give dichotomous results, even allowing for their internal error estimates. We will discuss in turn the case for a 'low' H_0, of 50 ± 10 km s^{-1} Mpc^{-1}, and a 'high' H_0, of 80 ± 10 km s^{-1} Mpc^{-1}, with the critical links for each shown schematically in Fig. 2.

The longer distance scale: $H_0 = 50 \pm 10$

The most reliable link leading to a low Hubble constant is provided by the supernovae classified as type Ia (on the basis of their characteristic spectra and, more often, the time history of the luminosity). The Hubble diagram of recession velocity against relative distance inferred from observed maximum supernova brightness has a fairly narrow dispersion at large distances, showing that the standard deviation in supernova luminosity is <40%, and possibly better with suitable selection[11,12]; these supernovae are therefore reasonably good 'standard candles'.

In an important recent development[13], the Hubble Space Telescope measured Cepheid variable stars in the host galaxy, IC4182, of the third supernova discovered in 1937; SN1937C, a prototypical type I supernova (Fig. 1). If SN1937C has a standard type Ia luminosity, the distance to IC4182 (4.8 Mpc) fixes the distances of remote type Ia supernovae, giving a low value of H_0. Moreover, the intrinsic luminosity of SN1937C implied by this distance is consistent with theoretical predictions from the standard model for type Ia supernovae[14,15]. (In this model a white dwarf star made of carbon or oxygen explosively burns to the iron group when it reaches the Chandrasekhar mass of ~1.4 solar masses by accretion from a nearby star. Once one accepts this idea, the maximum total luminosity follows from the mass of radioactive nickel produced.) Thus two ways to establish the absolute luminosity of type Ia supernovae give the same answer.

It follows that if the longer distance scale were wrong, then two independent ways to calibrate type Ia supernovae (Fig. 2*a*) would have to be wrong yet fortuitously agree. For example, if $H_0 = 80$ km s^{-1} Mpc^{-1}, then SN1937C would have to be the most (intrinsically) luminous type Ia supernova on record—about 2.5 times as bright as average, or almost 3 standard deviations. This is not to be excluded: for example, if SN1937C, instead of being typical, were as intrinsically luminous as the brightest of the dozen or so well studied distant type Ia supernovae, we would get $H_0 = 69 \pm 8$ km s^{-1} Mpc^{-1}. Second, if $H_0 = 80$ km s^{-1} Mpc^{-1}, then the typical type Ia supernova would have to be less than half as bright as predicted by the standard Ia supernova model (the agreement with SN1937C would still hold, as the distance to IC4182 is fixed). This is also conceivable; a number of recent modifications[16,17] to the standard model, including improved models of the opacity, radiative transfer, and the explosive deflagration/detonation process, partly motivated by observed light curves, ejecta velocity and nucleosynthesis products, tend towards a higher value of H_0.

The shorter distance scale: $H_0 = 80 \pm 10$

A sensible case can also be made for a high value of H_0[18,19]. As indicated in Fig. 2*b*, there are at least four independent methods for measuring relative distances to galaxies, and two independent empirical methods of attaching an absolute distance to these distance ratios which consistently give $H_0 = 80 \pm 10$ km s^{-1} Mpc^{-1}.

One way to establish distance ratios uses the Tully–Fisher (TF) relation between the luminosity of a spiral galaxy and its rotation speed[20]; more rapidly rotating galaxies tend to be more massive, and hence more luminous. Like its somewhat less precise sister relation for elliptical galaxies (the Faber–Jackson or

D_n–σ relation relating size D_n and velocity dispersion σ), TF reflects poorly understood but apparently precise regularities in the ways galaxies happen to have formed: over a wide range of masses, galaxies exhibit narrowly defined correlations between mass-to-light ratio and surface brightness. Because rotation velocities are determined from the Doppler shift in galaxy spectra (usually 21-cm linewidth), which can be measured at large distances, the TF relation can be used to predict relative galaxy luminosities and hence distance ratios. The Tully–Fisher relation gives a reliable distance ratio between two galaxies with an empirical accuracy of ~15% per galaxy[18], and establishes even more accurate distance ratios between galaxy clusters[9]. A large dispersion in the TF distances to spiral galaxies in the direction of the Virgo cluster, which had cast doubt on the viability of this technique, is now known to arise from the true depth of this cluster in space[18,21]. The true scatter of TF is remarkably narrow, so that the notorious 'Malmquist bias' which formerly plagued this method is small, and is now statistically controlled[22]. (Malmquist bias occurs if a sample systematically favours intrinsically brighter galaxies at greater distance.)

A check on the reliability of TF comes from two relatively new techniques, which have matured in the past few years to equal or even surpass the precision of TF. One technique uses surface brightness fluctuations (SBF)[23] in galaxies, in which the statistical variance in surface brightness is adopted as a measure of the distribution of star luminosities in a galaxy. This can be used at large distances, even when the individual stars cannot be spatially resolved because of atmospheric and optical blurring. The method has been applied to several hundred galaxies, and, with corrections for variations in stellar populations, yields relative distances accurate to ~5% per galaxy. Figure 3*a* shows the remarkable consistency of distance ratios for galaxies where both the TF and SBF methods have been applied.

The second new method uses the shape of the luminosity function for the planetary nebulae in a galaxy (PNLF)[24], which is found to display a sharp upper cutoff in luminosity. Observational evidence[24] shows that the cutoff is surprisingly universal. Again, PNLF distance ratios are in excellent agreement with SBF and TF, and shown in Fig. 3*b*. For all these techniques, TF, SBF and PNLF, the empirical precision is remarkably good, even better than one might have anticipated from first principles. Their precision is even confirmed by comparison with distance ratios from type Ia supernovae, although they clearly disagree with the absolute brightness calibration using the methods described above.

The value $H_0 = 80 \pm 10$ km s^{-1} Mpc^{-1} then follows if one uses Cepheid distances to a half-dozen nearby calibrator galaxies[25] for which one or more of these methods can be applied. Alternatively, instead of relying on a few local galaxies calibrated by Cepheids, we can anchor TF distances with a direct physical measure of distance called the expanding photosphere method (EPM)[3]. This method uses type II supernovae, produced by core collapse during the death throes of massive stars (SN1987A being the best known example). These are not standard candles like type Ia supernovae, but their distances can be estimated using models that fit measured spectral properties of their ejecta. Like the older Baade–Wesselink method applied to variable stars, EPM yields an absolute distance by comparing the physical velocity of the ejecta (measured from the Doppler shift of absorption lines in the gas) to rate of change in angular size of the material (derived from the temperature and brightness, and measured at several epochs). The EPM requires a model atmosphere to enable the calculation of the emergent flux, for the photosphere does not radiate like an ideal black body. It also requires a tricky estimate of temperature from observed colours. Although the correction is substantial and the colour–temperature identification contains some uncertainties, the calculation yields an accurate distance to SN1987A in the LMC[3], and the stability of the estimated distances as the supernovae expand and cool is impressively good. In at least seven of the ten most carefully studied type II supernovae, EPM distances agree well with TF distances to the host galaxy calibrated with Cepheids[26], independently supporting a high value of H_0. (TF is needed here because the host galaxies of these type II supernovae happen to lie in a local region of slower than average expansion, so that an application of EPM on its own underestimates H_0.)

Thus, if $H_0 = 50 \pm 10$ km s^{-1} Mpc^{-1}, all of the following are required: the TF distance for the six nearby calibrators needs to be anomalous, PN and SBF techniques need to be anomalous for two of these (M31, M81) and the EPM technique must be systematically wrong in the seven best studied cases, except in the independently validated cases, SN1987A and SN1970g, whose EPM distances are confirmed by Cepheid distances[3]. Any of these is a possibility, but the required anomalies are much larger than expected from empirically measured dispersions.

The age of the Universe

The value of H_0 has an important effect on cosmological theory, in particular on the relativistic, homogeneous expanding universe models which relate expansion rate to cosmic age. For

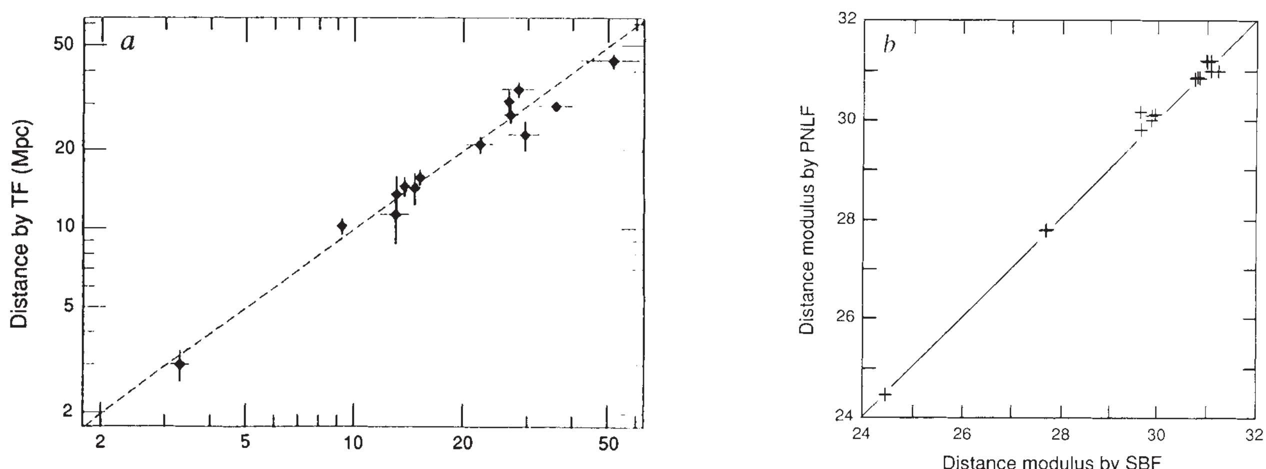

FIG. 3 *a*, Comparison of galaxy distances determined by the Tully–Fisher (TF) and surface brightness fluctuation (SBF) methods. Part of Fig. 29 from Jacoby *et al.*[18] by permission. *b*, Comparison of galaxy distances determined by the surface brightness fluctuation (SBF) and planetary nebula luminosity functions (PNLF) methods, for 16 galaxies. The logarithmic scale here is the distance modulus 5 log (d/10 pc); for instance 25 stands for 1 Mpc, 32 for 25 Mpc. (Data from G. Jacoby, R. Ciardullo and J. Tonry, personal communication).

$H_0 = 50$ km s^{-1} Mpc^{-1} the Hubble time, given by H_0^{-1}, is 20 Gyr. This is the age of the expanding universe if the expansion 'coasts', as in a cosmological model in the limit of very low mass density. The real Universe of course contains matter, which tends to decelerate the expansion by its gravity, and may also contain a gravitating physical vacuum (or 'cosmological constant' Λ), which accelerates the expansion. On grounds of elegance, most physicists would prefer the Einstein–de Sitter model, in which $\Lambda = 0$ and the mass density is such that the effective gravitational potential energy per unit mass is equal to the kinetic energy. (This means that the Universe expands at precisely its escape velocity, and, in relativistic cosmology, that space curvature is negligibly small.) In the Einstein–de Sitter model, the age of the Universe from high density to now is two-thirds of the Hubble time, or about 13 Gyr if $H_0 = 50$ km s^{-1} Mpc^{-1}. This is comparable to the ages of the oldest stars, as derived from models for stellar evolution. On the other hand, the higher value $H_0 =$ 80 km s^{-1} Mpc^{-1} would make the Hubble time $H_0^{-1} = 12$ Gyr and the Einstein–de Sitter age only 8 Gyr, too small by any accepted measure: one would have to learn to live in a low-density Universe, with an age of ~11 Gyr if $\Lambda = 0$ and the mass density is 10% of the Einstein–de Sitter value (the minimum implied by dynamical estimates[36]). This would still not be welcomed by those who study stellar evolution, who consider it difficult to push the maximum ages below 13 Gyr. It would be abhorred by those who study the inflation picture as a possible explanation for the large-scale isotropy of the visible Universe, for that requires space curvature be negligibly small, as in the Einstein–de Sitter model.

A low-density Universe with a non-zero cosmological constant can have negligible space curvature (as demanded by inflation) and an expansion timescale longer than the Hubble time; however, in such models the predicted frequency of gravitational lensing of distant quasars by galaxies positioned close to the line of sight is considerably larger than is observed[27]. One way out of this constraint is that at the large distances typical for lensing events, galaxies are often observed to have considerably larger star formation rates than in their nearby older counterparts, and might also contain larger quantities of dust; lensing events could therefore be obscured. If so, the reduced rate of observed lensing events might be consistent with the low-density cosmologically flat Universe which is the second choice for those who study inflation, and which may thus offer the easiest way to reconcile the shorter distance scale with the ages of stars and the elements. Confirmation of non-zero cosmological constant—a gravitating physical vacuum—would have important ramifications for fundamental physics.

An end to the controversy?

We have focused so far on the currently most precise and best statistically controlled methods. There are, however, other methods of fixing the absolute distance scale[18,19,28], of which we list a few particularly active cases. Recently, high-resolution ground-based imaging has revealed variable stars in some Virgo cluster galaxies[29]. Some of these are long-period variables, and some may be Cepheids; the current data are still too sparse to be certain, but the fact that they are bright enough to be seen at all tends to support a shorter distance scale, or high H_0. The same argument applies to the fact that we can resolve the brightest stars in galaxies in the direction of the Virgo cluster[30,31].

Other promising techniques may eventually provide a way of bypassing entirely the laborious steps of the traditional distance ladder. Time delays (δt) between images of gravitationally lensed quasars[32,33] give a direct physical estimate of the scale of the Universe: $\delta t \approx \theta^2 H_0^{-1}$, where θ is the angle by which the lensed images are separated. But current estimates from the best studied example, Q0957+561, are close to both of the distance ladder estimates, giving a range of values consistent with high or low H_0 depending on model parameters[34]. Observations of comptonization of background radiation by X-ray emitting gas in one specific galaxy cluster[35] currently suggest low H_0, based on a simple smooth spherical model for the cluster gas. A reliable independent estimate of H_0 from this technique will require a large statistical sample to control effects of cluster elongation and selection effects in cluster samples.

We have raised several issues that seem to be particularly critical. Is SN1937C a fair representative of the class of type Ia supernovae? If so, are the nearby calibrator galaxies for TF, SBF and PNLF indeed unrepresentative of more distant galaxies, by an offset that happens to be the same for three independent kinds of observation? Must we modify the standard model for type Ia supernovae, or the apparently successful model for the physics of the expanding photospheres of type II supernovae? The issues are sharply defined, and they are likely to be resolved by observational programs that are either in progress or at least seem feasible. An important test, now underway using the Hubble Space Telescope, is the measurement of additional Cepheid distances to obtain type Ia, PNLF, SBF, TF and EPM calibrations in other, more distant galaxies; in addition, a valuable check on EPM systematic errors will be provided by the fortuitous appearance this year of a new type II supernova, 1993J, in M81, which now also has a firm Cepheid distance. □

M. Fukugita is at the Yukawa Institute of Kyoto University, Kyoto 606, Japan and at the Institute of Advanced Study, Princeton, New Jersey 08540, USA. C. J. Hogan is at the University of Washington, Seattle, Washington 98195, USA. P. J. E. Peebles is at Princeton University, Princeton, New Jersey 08544, USA.

1. Sandage, A. *Astrophys. J.* **402,** 3–14 (1993).
2. de Vaucouleurs, G. in *10th Texas Symp. on Relativistic Astrophysics* (eds Ramaty, R. & Jones, F. C.) *Ann. N.Y. Acad. Sci.* **375,** 90–122 (1981).
3. Schmidt, B. P., Kirshner, R. P. & Eastman, R. G. *Astrophys. J.* **395,** 366–386 (1992).
4. Panagia, N. *et al. Astrophys. J.* **380,** L23–L26 (1991).
5. Freedman, W. & Madore, B. in *New Perspectives on Stellar Pulsation and Pulsating Stars, IAU Colloq. No. 139* (eds Nemec, J. & Matthews, J.) (Cambridge Univ. Press, 1993).
6. Freedman, W. *et al. Astrophys. J.* (submitted).
7. Kristian, J., Sandage, A. & Westphal, J. *Astrophys. J.* **221,** 383–394 (1978).
8. Lauer, T. & Postman, M. *Astrophys. J.* **400,** L47–L50 (1992).
9. Mould, J. R. *et al. Astrophys. J.* **409,** 14–27 (1993).
10. Sandage, A. & Tammann, G. A. *Astrophys. J.* **365,** 1–12 (1990).
11. Leibundgut, B. & Pinto, P. A. *Astrophys. J.* **401,** 49–59 (1992).
12. Branch, D. & Miller, D. L. *Astrophys. J.* **405,** L5–L8 (1993).
13. Sandage, A., Saha, A., Tammann, G. A., Panagia, N. & Macchetto, D. *Astrophys. J.* **401,** L7–L10 (1992).
14. Arnett, W. D., Branch, D. & Wheeler, J. C. *Nature* **314,** 337–338 (1985).
15. Branch, D. *Astrophys. J.* **392,** 35–40 (1992).
16. Khokhlov, A., Muller, E. & Höflich, P. *Astr. Astrophys.* **270,** 223–248 (1993).
17. Yamaoka, H., Shigeyama, T. & Nomoto, K. Univ. of Tokyo preprint (1993).
18. Jacoby, G. H. *et al. Publs astr. Soc. Pacif.* **104,** 599–662 (1992).
19. van den Bergh, S. *Publs astr. Soc. Pacif.* **104,** 861–883 (1992).
20. Tully, R. B. & Fisher, J. R. *Astr. Astrophys.* **54,** 661–673 (1977).
21. Fukugita, M., Okamura, S. & Yasuda, N. *Astrophys. J.* **412,** L13–L16 (1993).
22. Ichikawa, T. & Fukugita, M. *Astrophys. J.* **394,** 61–80 (1992).
23. Tonry, J. *Astrophys. J.* **373,** L1–L4 (1991).
24. Jacoby, G. H., Ciardullo, R. & Ford, H. C. *Astrophys. J.* **356,** 332–349 (1990).
25. Madore, B. F. & Freedman, W. L. *Publs astr. Soc. Pacif.* **103,** 933–957 (1991).
26. Pierce, M. *Astrophys. J.* (in the press); National Optical Astronomy Observatories preprint No. 502.
27. Maoz, D. *et al. Astrophys. J.* **402,** 69–75 (1993).
28. Huchra, J. P. *Science* **256,** 321–325 (1992).
29. Pierce, M. *et al.* in *New Perspectives on Stellar Pulsation and Pulsating Stars, IAU Colloq. No. 139* (eds Nemec, J. & Matthews, J.) (Cambridge Univ. Press, 1993). **(See ref. 5)**
30. Shanks, T. *et al. Mon. Not. R. astr. Soc.* **256,** 29P–32P (1992).
31. Pierce, M. J., McClure, R. D. & Racine, R. *Astrophys. J.* **393,** 523–529 (1992).
32. Press, W. H., Rybicki, G. B. & Hewitt, J. N. *Astrophys. J.* **385,** 416–420 (1992).
33. Lehár, J., Hewitt, J. N., Roberts, D. H. & Burke, B. F. *Astrophys. J.* **384,** 453–466 (1992).
34. Bernstein, G. M., Tyson, J. A. & Kochanek, C. S. *Astro. J.* **105,** 816–830 (1993).
35. Birkinshaw, M., Hughes, J. P. & Arnaud, K. A. *Astrophys. J.* **379,** 466–481 (1992).
36. Peebles, P. J. E. *Principle of Physical Cosmology* (Princeton Univ. Press, 1993).

ACKNOWLEDGEMENTS. This work was inspired by the January 1993 Aspen Winter Conference on Astrophysics; we thank the participants for their guidance. We are grateful to D. Branch, G. Jacoby, R. Kirshner, J. Peacock, M. Pierce, S. van den Bergh, J. Tonry and C. Wheeler who helped improve (while not necessarily approving) a draft version of the report. This work was supported (in part) at Kyoto by the Japan–US Cooperative Science Programme, by JSPS and the US NSF; at the University of Washington by NASA and the US NSF; and at Princeton by the US NSF.