The world according to the Hubble Space Telescope

Mario Livio*

Space Telescope Science Institute, Baltimore, MD 21218

(Dated: March 24, 2003)

Abstract

The Hubble Space Telescope (HST), in its thirteen years of operation, has allowed us to observe properties of the universe humans have been able, until very recently, to probe only with their thoughts. This review presents a brief summary of a few of the highlights of HST discoveries, discusses their physical implications, and identifies unsolved problems. A broad range of topics is covered, from our own solar system to cosmology. The topics fall into the general categories of: planets (including both in the solar system and extrasolar), stellar evolution, black holes (including both of stellar-mass and supermassive), galaxy formation and evolution, the determination of cosmological parameters, and the nature of the recently discovered "dark energy."

*mlivio@stsci.edu; http://www-int.stsci.edu/~mlivio/

Contents

I. INTRODUCTION

When Galileo Galilei peered through his small telescope 400 years ago, it resulted in an unprecedented period of astronomical discovery. There is no doubt that the Hubble Space Telescope (HST; Fig. 1) in its first 13 years of operation has had a similarly profound impact on astronomical research. But the Hubble Space Telescope did much more that that. It literally brought a glimpse of the wonders of the universe into millions of homes worldwide, thereby inspiring an unprecedented public curiosity and interest in science.

The Hubble Space Telescope has seen farther and sharper than any optical/UV/IR telescope before it. Unlike astronomical experiments that were dedicated to a single, very specific goal (like the Cosmic Background Explorer or the Wilkinson Microwave Anisotropy Probe), the Hubble Space Telescope's achievements are generally not of the type of singular discoveries. More often, the Hubble Space Telescope has taken what were existing hints and suspicions from ground-based observations and has turned them into certainty. In other cases, the level of detail that HST has provided forced theorists to re-think previous broad-brush models and to construct new ones that would be consistent with the superior emerging data. In a few instances, the availability of HST's razor-sharp resoution at critical events provided unique insights into inividual phenomena. In total, by observing tens of thousands of astronomical targets, the Hubble Space Telescope has contributed significantly to essentially all the topics of current astronomical resarch, covering objects from our own solar system to the most distant galaxies.

The 12.5-ton orbiting observatory was launched into orbit on April 24, 1990. It circles the Earth every 90 minutes in a 330 nmi (607 km) orbit and operates around the clock above all but the thinnest remnants of the Earth's atmosphere. The telescope provides information to many individual astronomers and teams of scientists worldwide, and it is engaged in the study of virtually all the astronomical constituents of the universe.

Crucial to fulfilling the objective of a 20-year mission is a series of on-orbit manned servicing missions. The first servicing mission (SM1) took place in 1993, the second (SM2) was flown in 1997, and the third was separated into two parts, one (SM3A) in 1999 and the second (SM3B) in 2002. A fourth mission is currently planned for the end of 2004 (pending the investigation into the shuttle Columbia disaster). During these missions, shuttle astronauts upgrade the the observatory's capabilities (by installing new instruments) and perform planned maintenance activities and necessary repairs. To facilitate this process, the telescope was designed so that its science instruments and several vital engineering subsystems have been configured as modular packages with standardized fittings accessible to astronauts in pressurized suits.

Hubble was designed to provide three basic capabilities, all of which are essential for innovative astronomical research:

1. High angular resolution—the ability to image fine details. The spatial resolution of HST in the optical regime is about 0.05 arcseconds.

2. Ultraviolet performance—using the fact that the observatory is above the Earth's atmosphere to produce ultraviolet images and spectra. The Hubble Space Telescope is sensitive down to a wavelength of 1150 Å (below the Lyman Alpha line of hydrogen).

3. High sensitivity—the ability to detect and take spectra of even very faint objects. The Hubble Space Telescope is about a hundred to a thousand times more sensitive in the ultraviolet than the previous space observatory (the International Ultraviolet Explorer).

Astronomers and astrophysicists using HST data have published over 3000 scientific papers to date. Hubble's findings are thus far too numerous to be described even briefly in one article. The following sections therefore simply highlight a few of the astronomical discoveries which, in my clearly biased view, have significantly advanced our understanding of the cosmos. In the spirit of progressing by "powers of ten" in cosmic distance, I shall start with the solar system and finish with cosmological distances. I apologize in advance to the many astronomers whose important work will not be presented in this article. I also regretfully acknowledge the incompleteness of the list of references. My list should be regarded as representative rather than comprehensive. Finally, the different sections are not intended to be exhaustive reviews—they are concise summaries of HST contributions to major astrophysical poblems.

II. OUR OWN BACKYARD—THE SOLAR SYSTEM

The Hubble Space Telescope has proved well suited for the study of objects in our own solar system. In fact, it has produced images of the outer planets that approach the clarity of those obtained from Voyager flybys. The Hubble Space Telescope's images of Mars have also been surpassed only by close-up shots taken by visiting space probes. The Hubble Space Telescope's images of the auroras of Jupiter and Saturn, of the disintegration of comet LINEAR, of the binary Kuiper belt object KBO 1998 WW31, and of a global dust storm on Mars have been truly spectacular and informative. However, no observation could compete with the drama provided by Comet Shoemaker-Levy 9 (SL-9) slamming into Jupiter.

A. The collision of Comet SL-9 with Jupiter

From a cosmic perspective, the impact of the ninth periodic comet discovered by the team of Carolyn and Gene Shoemaker and David Levy on Jupiter was unremarkable. The cratered surfaces of many planetary bodies and satellites and the prehistoric mass extinctions on Earth are a testament to the fact that such impacts are ubiquitous in the history of the solar system. Furthermore, even the "seeds" of life on Earth may have been planted by such a bombardment.

From a human perspective, however, this collision marked a "once in a lifetime" event, which caused the mobilization of all the astronomical resources for an unprecedented observational campaign in July 1994.

The Hubble Space Telescope played a crucial role in the SL-9 campaign, both in the initial characterization of the comet fragments and in producing unique images and spectra of the impact events themselves and later of the impact sites.

The multiple comet SL-9 was discovered in late-March 1993 (Shoemaker *et al.*, 1993). Subsequent images taken with HST in July of 1993 (Weaver *et al.*, 1994) revealed that the comet had spectacularly split into 20-odd fragments, probably because of tidal disruption (during a previous close approach to Jupiter), producing a train of fragments fittingly dubbed a "string of pearls" (Fig. 2).

The conditions for tidal break-up of a self-gravitating incompressible cometary sphere can generally be written in the form (e.g., Dobrovolskis, 1990; Sekanina *et al.*, 1994)

$$\frac{\rho_c}{\rho_p}\left(\frac{d}{R_p}\right)^3 \lesssim k\frac{P_c}{U} \ , \tag{1}$$

where ρ_c, ρ_p are the densities of the comet and planet, respectively, R_p is the planet's radius, d is the distance of the comet from the planet, P_c is the comet's central pressure, U is a measure of the comet's tensile strength, and k is a constant of order unity. The facts that: (i) secondary fragmentation was observed by HST, and (ii) the gradual disappearance of condensations was also well documented by HST (Weaver *et al.*, 1995), argued strongly for tidal splitting of a discrete nucleus as a result of gradual fissure propogation. Computations of the orbit suggested that the comet would collide with Jupiter in July 1994 (Nakano, 1993; Yeomans and Chodas, 1993).

The impact of the first fragment (fragment "A") occurred on July 16, 1994, followed by all the other fragments smashing into Jupiter's atmosphere over the following week. The Hubble Space Telescope's images taken during the impact events (Hammel *et al.*, 1995) showed spectacular plumes resembling nuclear "mushrooms" above the limb of Jupiter (Fig. 3). Typically, the plumes attained altitudes of about 3000 km within 6–8 minutes of impact, and were seen falling and spreading within 10 minutes of impact. The manner in which the height of the plume scales with SL-9 parameters and the ambient pressure can be obtained from the following simple considerations (e.g., Chyba *et al.*, 1993; Field and Ferrara, 1995; Zahnle and MacLow, 1994). Once aerodynamic forces overcome material strength, the incoming fragment deforms to form a flattened "pancake" due to the difference in ram pressure acting on the front and on the sides. Due to the increase in the cross-section (and concomitantly in the drag), the fragment is brought to an abrupt halt, accompanied by an explosive energy release. If we denote the energy being released per steradian by $E_i/4\pi$, then the obtained ejection velocity is given by

$$V_{ej} \simeq \sqrt{\frac{E_i}{4\pi\rho_a H^3}} = \sqrt{\frac{E_i g}{4\pi P_a H^2}} \ , \tag{2}$$

where ρ_a, P_a are the ambient density and pressure, respectively, H is the pressure scale height in the atmosphere and g is the gravitational acceleration. The plume height, therefore, scales like

$$h_{\text{plume}} \simeq \frac{V_{ej}^2}{2g} \sim \frac{E_i}{P_a H^2}. \tag{3}$$

Detailed numerical simulations have confirmed the basic physical picture described above (e.g., MacLow, 1996).

A combination of the observations from the Keck Telescope, HST, Palomar, and the Galileo spacecraft, eventually produced a comprehensive explanation for the observed 2.3μ lightcurve (shown in Fig. 4 for fragment R; Graham *et al.*, 1995). The sequence of events can be described phenomenologically by the cartoon in Fig. 5 (Zahnle, 1996).

The first precursor marked the meteor trail (with a brightening timescale of order $H/v \sim 1s$). As explained above, most of the fragment's kinetic energy was released at the last second of its existence, resulting in an explosive fireball. The second precursor marks the rise of the fireball above Jupiter's limb, having already cooled to $\sim$ 500–700 K (Carlson *et al.*, 1995). As the fireball expanded, silicates condensed, followed by carbonaceous matter. The sunlit parts of these condensates were the tracers that marked the plume seen by HST. The main peak in the light curve was caused by the ejecta plume falling back onto the atmosphere (Nicholson *et al.*, 1995). The rise and fallback occurred on a timescale of $\sqrt{8h_{\rm plume}/g} \sim 10^3$ s (for a plume height of $\sim$ 3000 km). The final, smaller peak, was probably caused by material bouncing off the atmosphere and falling back.

Another interesting phenomenon revealed by the HST images of the impact *sites* was the propagation of waves (Fig. 6). The radius of the prominent dark ring in Fig. 6 is 3700 km, and a smaller ring, of radius 1750 km is also visible (Hammel *et al.*, 1995).

Measurements of the radii of the circular rings as a function of time showed that the "fast" wave was seen to spread at a constant speed of 450 m s^{-1}, while the speed of the "slow" wave (corresponding to the fainter ring) was somewhat less constrained. Ingersoll and Kanamori (1996) examined many possibilities for the identification of the fast wave, including f-modes, p-modes, r-modes and g-modes. These waves are classified by the different restoring forces that cause them to propagate. In acoustic waves (p-modes) the restoring force is compressibility; in surface waves (like those in the ocean; f-modes) it is gravity; in inertial oscillations (r-modes) it is the Coriolis force, and in internal gravity waves (g-modes) it is gravity. Ingersoll and Kanamori showed that the only waves that match the observed speed of 450 m s^{-1} are g-modes propagating in a water cloud which has an O/H ratio that is ten times higher than that ratio in the Sun, $e_{\rm H_2O} = 10$ (the waves were too slow to be acoustic waves and moved at too constant a speed to be rings of debris). The propagation speed of these modes varies as the square root of the water abundance. The requirement of a high O/H ratio was found to be in conflict with data from the Galileo probe, which suggested that from above the cloud tops down to the 20 bar level the water was *less* abundant than in the Sun. The cause for this discrepancy is still unknown, although it has been suggested that the probe may have entered a clearing in the clouds (a "5-micron hot spot") associated with a dry downdraft. This uncertainty demonstrates that even the highest quality data do not provide all the answers. Nevertheless, there is no question that the impact of comet

SL-9 on Jupiter provided a unique opportunity to study such events.

III. EXTRASOLAR PLANETS

The solar system was not the only place to provide for planetary drama. During the past eight years we have witnessed the number of known extrasolar planets orbiting Sun-like stars going from *zero* to about one hundred at the time of this writing! The human fascination with the possibility of discovering extraterrestrial life has helped turn planet detection into one of the major frontiers of today's astronomy.

The Hubble Space Telescope's contributions to this exciting field have been related mostly to the discovery of circumstellar disks from which planets form, and these will be described in Section IV. In two cases, however, HST has produced three truly unique sets of observations, one in a globular star cluster, and two following the light curve and spectrum of a transiting planet.

A. Where have all the planets gone?

The discovery of a planet orbiting the star 51 Peg (Mayor and Queloz, 1995) in a remarkably tight orbit of 4.231 days challenged prevailing theoretical views and marked the start of a "Golden Age" in planet discoveries by radial velocity surveys (see, for example, Marcy *et al.*, 2000). The existence of gas giant planets in very close orbits (at 0.04–0.05 AU) enables even *photometric* searches for these planets ("hot Jupiters"). At such small separations, the probability of the planet transiting ("eclipsing") the parent star, given random orbital inclinations, is about 10%. More precisely, from Kepler's laws, the *duration* of a transit in hours (reduced by $\pi/4$ for average chord length) is given by

$$\tau_{\rm tran} \simeq 1.4 M_*^{-1/3} R_* P_{\rm orb}^{1/3}\ {\rm hr}\ , \tag{4}$$

where M_*, R_* are the stellar mass and radius in solar units and $P_{\rm orb}$ is the orbital period in days. The *probability* of a transit per system is given by

$$Pr_{\rm tran} \simeq 23.8\% M_*^{-1/3} R_* P_{\rm orb}^{-2/3}\ . \tag{5}$$

Transits are very important because of the following reasons: (a) With transits one can immediately obtain the *size* of the planet [the transit depth is equal to $(R_p/R_*)^2$]. (b) In the case of transits, measurements of radial velocities immediately provide the planet's *mass* (and not just $m_p \sin i$, where i is the otherwise unknown orbital inclination). (c) A knowledge of the mass and radius can be used to test theoretical mass-radius relations for close-orbiting planets.

In 1998, a team of astronomers led by Ron Gilliland of the Space Telescope Science Institute embarked on the ambitious program of searching for transits in some 40,000 stars in the core of the globular cluster 47 Tucanae (Fig. 7). This cluster is a particularly interesting target for such a search for several reasons. (i) For a typical star in the cluster, $M_* = 0.81$, $R_* = 0.92$, and a typical "hot Jupiter" orbital period (e.g., $P_{\rm orb} = 3.8$ days), the probability of transit is 9.6% and the duration is 2.2 hours. For a planet of radius $R_p = 1.3 R_J$ (where R_J is Jupiter's radius, as observed for the transiting system HD 209458; Charbonneau *et al.*, 2000; Henry *et al.*, 2000), the expected transit depth is about 2%, easily detectable by HST. (ii) Given that the frequency of "hot Jupiters" with orbital periods shorter than 5 days (ten are known at the time of writing) in the solar neighborhood is about 0.8%–1%, with a 10% chance of transit one expects roughly one in every 1000 surveyed stars in 47 Tuc to show a transit (and therefore a few tens of detections when monitoring $\sim 40,000$ stars). (iii) The globular cluster 47 Tuc is about 11 billion years old and it has a total metallicity (abundance of elements heavier than helium) of about one-third that of the Sun (Salaris and Weiss, 1998). The frequency of planets in such an environment can therefore have important implications for models of planet formation.

The Hubble Space Telescope observed 34,091 stars in 47 Tuc, obtaining time series photometry over a period of 8.3 days (continuously, except for Earth occultations and passages through the South Atlantic Anomaly). A more detailed calculation of the expected number of transits, taking into account the actual distribution of V magnitudes of the stars in the sample (Gilliland *et al.*, 2000) and time series noise, showed that *17 transits should have been seen*, assuming the same frequency of hot Jupiters as in the solar neighborhood. None was detected! Undoubtedly, this was not a problem with the detection method. As many as 75 variable stars have actually been detected. Furthermore, an eclipsing system with a period of 1.34 days and an eclipse depth of $\sim 3\%$ was actually discovered. However, the presence of a secondary eclipse in the system showed that this was a grazing eclipse by a low-mass (K dwarf) star and not a planet.

The direct implication is therefore that with a very high degree of confidence the frequency of hot Jupiters in 47 Tuc is *at most 1/10* that in the solar neighborhood.

The main question is then: why are there much fewer orbiting planets in 47 Tuc? There are, in principle, two possible explanations for the paucity of planets in orbits around globular cluster stars: (i) Planets do not form at all (or form much more rarely) in globular clusters. (ii) Planets are torn from their parent stars.

Planet *formation* may be suppressed in old stellar clusters because of two main reasons: low metallicity, or photoevaporation of protoplanetary disks.

Conventional wisdom of giant planet formation suggests that this is a two-step process. In the first, the collisional accumulation of rocky and big planetesimals leads to a runaway growth of a solid core (e.g., Lissauer, 1987) of about ten Earth masses (10 $M_\oplus$). In the second, the core acquires a gaseous atmosphere, which eventually collapses, leading to the rapid accretion of hydrogen and helium from the protoplanetary disk. One might suspect, therefore, that in low-metallicity environments, the nucleation of dust grains would be greatly suppressed, and concomitantly planets would form more rarely. Furthermore, it has generally been noted that stars which bear planets are, on average, enriched in metals in comparison to the Sun (e.g., Butler *et al.*, 2000; Gonzalez, 1997; Laughlin, 2000; Reid, 2002), and this trend is even more pronounced for the short-period planets. Thus, the paucity of hot Jupiters in 47 Tuc may simply reflect the cluster's low metallicity.

There is another reason why giant planet *formation* may be suppressed in dense stellar clusters. The disk around any star in a cluster is exposed to ultraviolet radiation from massive cluster stars. This radiation heats the disk surface, raises the sound speed, c_s, and causes gas beyond a radius of

$$R_g \simeq 0.5 \frac{GM_*}{c_s^2} \tag{6}$$

to become unbound and flow away as a thermal wind. The mass-loss rate in such a photoevaporation flow scales as (e.g., Bertoldi and McKee, 1990; Johnstone *et al.*, 1998)

$$\dot{M}_{\rm outflow} \sim \Phi^{1/2} d^{-1} \ , \tag{7}$$

where Φ is the flux of ionizing photons and d is the disk's distance from the emitting source. In a star forming region like Orion, the ionizing flux is of order $\Phi \sim 10^{49}$ s^{-1}, of the same order as the flux expected in a cluster with $n_{\rm cluster} \sim 10^3$ members. The mass loss rate from one of the disks in Orion, HST 182$-$413, has been estimated from HST observations to be $\dot{M}_{\rm outflow} \sim 4\times10^{-7}$ M$_\odot$ yr^{-1}. Since the mass of that disk is estimated to be (Johnstone *et al.*,

1998) $M_{\rm disk} \sim 0.04$ M$_\odot$, the disk lifetime is expected to be $\tau_{\rm disk} \sim M_{\rm disk}/\dot{M}_{\rm outflow} \sim 10^5$ yr. For somewhat smaller disk sizes (e.g., like that in the solar system ~ 30 AU) the disk lifetime can be somewhat extended (the EUV-induced flow scales as $r_{\rm disk}^{3/2}$). Nevertheless, the disk lifetime may be significantly shorter than the planet formation timescale, estimated to be of order 10^6–10^7 yrs (e.g., Pollack *et al.*, 1996; Shu *et al.*, 1993). Armitage (2000) calculated the expected disk lifetime as a function of the cluster richness. He found that giant planet formation is strongly suppressed in custers with $\sim 10^5$ stars out to around 1 pc or more.

There are two ways in which, in principle at least, the conclusion of a low probability for planet formation in clusters, due to disk destruction, may be avoided: (i) If giant planet formation occurs not via a core accretion mechanism, but rather through direct and rapid (within $\sim 10^3$ yr) fragmentation (due to a gravational instability) into self-gravitating clumps of gas (e.g., Boss, 2000; Mayer *et al.*, 2002). (ii) If the formation of high-mass stars (which produce the photoevaporating radiation) is significantly delayed (by $\sim 10^6$ yr) compared to the formation of low mass stars. At present both of these processes are sufficiently uncertain that no definitive conclusion about their viability can be drawn.

The second possible explanation for the absence of "hot Jupiters" in 47 Tuc is that even if giant planets do form, they are torn away from their parent stars because of encounters in the crowded cluster environment, and thus cannot exhibit transits. This possibility has been investigated in some detail by Davies and Sigurdsson (2001). The cross-section for two stars with a relative velocity at infinity V_∞ to pass within $R_{\rm min}$ from each other is given by

$$\sigma = \pi R_{\rm min}^2 \left(1 + \frac{V^2}{V_\infty^2}\right) \ , \tag{8}$$

where V is the relative velocity ($V^2 \propto (M_1 + M_2)/R_{\rm min}$) at closest approach. The second term expresses the effect of gravitational focusing. When $V \gg V_\infty$, as is expected in globular clusters, $\sigma \propto R_{\rm min}$. The timescale for an encounter between a star-planet system and a star is of order ($\tau_{\rm enc} = 1/n\sigma V$)

$$\tau_{\rm enc} \simeq 3 \times 10^8 \ {\rm yr} \left(\frac{n}{10^5 \ {\rm pc}^{-3}}\right)^{-1} \left(\frac{V_\infty}{10 \ {\rm km \ s}^{-1}}\right) \left(\frac{d}{1 \ {\rm AU}}\right)^{-1} \left(\frac{M}{{\rm M}_\odot}\right)^{-1} \ , \tag{9}$$

where n is the number density of stars of mass M and d is the semimajor axis of the planetary orbit. Encounters therefore occur on a timescale that is much shorter than a Hubble time in a cluster like 47 Tuc, which has a density of about 1.5×10^5 pc^{-3} at its center and about an order of magnitude lower at its half-mass radius (e.g., Howell *et al.*, 2000). Generally, binaries (with masses M_1 and M_2 for the primary and secondary, respectively) in which the binding energy is lower than the kinetic energy of the colliding star (of mass M_3) are disrupted. The condition for breakup is

$$V_\infty^2 \gtrsim \frac{GM_1M_2(M_1+M_2+M_3)}{M_3(M_1+M_2)d} \ . \tag{10}$$

Since $M_2 \ll M_1, M_3$, in the case of a star-planet system, breakup of the system is expected to occur with a high probability when the third star passes within a distance d. Davies and Sigurdsson (2001) performed a large number of simulations for various values of $V_\infty/V_{\rm orb}$ and considered both encounters (of star-planet systems) with single stars and with binaries. They found that wide ($d \gtrsim 0.3$ AU) planetary systems are likely to be broken up within the half-mass radius of 47 Tuc, but tighter systems ($d \lesssim 0.1$ AU) or systems in less dense regions may survive. The results of Davies and Sigurdsson still do not provide a complete explanation for the absence of transits in 47 Tuc. Future observations, perhaps of less dense, somewhat higher metallicity clusters (like NGC 6352) will be required to explain which factor is responsible for the dearth of planetary transits in 47 Tuc. An attempt to detect transits associated with Galactic bulge and disk stars (that span 1.5 dex in metallicity, and are in an environment that is orders of magnitude less dense than the center of a globular cluster) would also be very valuable. An understanding of the environments that are either conducive to or prohibitive for the existence of planetary transits may constitute an important step in the study of planet formation.

B. The transiting planet HD 209458

The low-mass companion to the star HD 209458 was the first extrasolar planet found to transit the optical disk of its parent star (Charbonneau *et al.*, 2000; Henry *et al.*, 2000; Mazeh *et al.*, 2000). As I explained in the previous section, transits offer a unique opportunity to determine the properties of the orbiting planet. The Hubble Space Telescope followed four transits (Brown *et al.*, 2001) and obtained an exquisitely detailed light curve (Fig. 8; note that the depth of the eclipse is only 1.7%!). The orbital period was determined to be 3.5247 days, the lower limit on the mass of the planet $M_p \sin i = 0.69 \pm 0.05 M_J$ (with the inclination angle $i = 86\overset{\circ}{.}6 \pm 9\overset{\circ}{.}14$), and the radius of the planet $R_p = 1.347 \pm 0.060\ R_J$. Most impressively, the precision of the HST light curve allowed even for searches for *rings* around the planet, for constraints on planetary *satellites*, and for probing the planet's *atmosphere*.

A ring system with significant opacity around the planet would cause distortions of the light curve relative to that of a spherical body. In particular, one would expect small dips in the light curve before the first and after the fourth contact. The observations were consistent with *no rings* (maximum ring radius consistent with observations was 1.8 R_p).

A satellite orbiting the planet might be detectable either from its photometric signature (the satellite would block light in addition to that obstructed by the planet), or from its influence on the orbital motion of the planet. Brown *et al.* (2001) showed that satellites larger than 1.2 $R_\oplus$ (where $R_\oplus$ is the Earth radius), or with masses larger than 3 $M_\oplus$ are excluded by the data.

Finally, precision spectrophotometric observations in the region of the sodium resonance doublet at 589.3 nm revealed that the photometric dimming during transit was deeper by $(2.32 \pm 0.57) \times 10^{-4}$ relative to simultaneous observations in adjacent bands (Charbonneau *et al.*, 2002). This additional dimming has been interpreted as absorption by sodium in the planet's atmosphere. In fact, the existence of a detectable sodium feature in the spectrum has been predicted by theoretical models (e.g., Seager and Sasselov, 2000). Furthermore, observations of brown dwarfs (stellar objects below the hydrogen burning limit) with similar surface temperatures show strong absorption in alkali metal lines (e.g., Burrows *et al.*, 2000). HD 209458 therefore represents the first *detection of an atmosphere* of an extrasolar planet. While not a shocking result in itself, since giant planets are clearly expected to have atmospheres, the actual detection marks the beginning of a new era in extrasolar planets research.

Even more interestingly perhaps, HST observations of HD 209458 during three transits revealed atomic hydrogen absorption of $\sim$15% over the stellar Lyman α line (Vidal-Madjar *et al.*, 2003). A comparison of the observations with models showed that the observations could be explained in terms of hydrogen atoms evaporating and escaping the planet in an asymmetric cometary-like tail. To account for the observed absorption depth, the simulations implied a minimum escape flow rate of $\sim 10^{10}$ g s^{-1}.

In the future, other spectral features (e.g., water vapor; with implications for the searches for extraterrestrial life) can be searched for. Furthermore, in principle, HST can detect even the secondary eclipse (when the planet is eclipsed by the star), expected to be characterized by a $\sim 10^{-5}$ decrease in the luminosity. This will allow for a determination of the planet's albedo, and for a comparison with models of planetary atmospheres. All of these results, whether already obtained, or expected, are somewhat of a surprise, since HST has been heralded primarily as a "cosmology machine," not as an exquisite tool in exo-planetary research.

Moving now from planets to stars, the Hubble Space Telescope has been particularly instrumental in revealing new details of stellar births and deaths.

IV. STARS AT BIRTH AND DEATH

Using HST's high resolution on one hand and its infrared capabilities on the other, astronomers have been able to probe the dusty environments of young stellar objects (YSOs) and of stars in their late evolutionary stages with unprecedented detail. While these observations have not revolutionized the field, they have, on one hand, led to a much deeper understanding of the processes involved, and on the other, opened an entirely new set of questions, while producing some of the most spectacular images.

A. Outflows and jets from young stellar objects

Molecular clouds are the reservoirs of mass and angular momentum from which stars are born. In the initial phase, dynamical infall occurs. The finite angular momentum of infalling cloud material leads to the formation of an accretion disk. Angular momentum is transported outward in this disk resulting in the accretion of mass (and some angular momentum) onto the central object (e.g., Camenzind, 1990; Najita, 2000; Shu *et al.*, 1994).

For a long time it has been known that outflows and collimated jets are signposts of stellar birth (see, for example, Königl, 1989; Reipurth and Heathcote, 1993). In particular, many of the radiative shocks known as Herbig-Haro (HH) objects were found to be associated with highly collimated jets emanating from the vicinity of YSOs (e.g., Reipurth, 1999). In some cases, these jets were found to be several parsecs in length (Reipurth *et al.*, 1998). Figure 9 shows a few remarkable HST images of jets from YSOs.

In spite of the ubiquity of astrophysical jets, we still lack a comprehensive theory of their acceleration and collimation (see e.g., review by Livio, 2000). The most promising mechanism relies on an accretion disk (around a central compact object like a YSO or a black hole) threaded by a large-scale poloidal magnetic field. Some magnetic flux is assumed to be in open field lines, inclined by an angle of more than 30° to the vertical to the disk's surface. Ionized gas is forced to flow along field lines. Since the foot points of these lines are anchored in the disk and rotate with it, material is accelerated centrifugally like beads on rotating wires (Blandford and Payne, 1982; Ogilvie and Livio, 2001). In this picture, the acceleration basically stops at the Alfvén surface, where the kinetic energy density in the jet becomes comparable to the magnetic energy density. Collimation, however, occurs primarily outside the Alfvén surface. There, the field gets wound up by the rotation, generating toroidal loops. The curvature force exerted by the toroidal field acts in the direction of the rotation axis, thus achieving collimation by "hoop stresses." Alternatively, in a vertical field of the form $B_z \sim (r^2/R_{\rm in} + 1)^{-1/2}$ (where $R_{\rm in}$ is the radius at the accretion disk's inner edge), in which the flux is largest at the outer disk, the field lines have a naturally collimating shape (e.g., Spruit, 1996). Physically, poloidal collimation is achieved in this case as the material encounters the high flux in the outer disk (especially if $R_{\rm disk} \sim R_{\rm Alfvén}$).

While this theoretical picture of jets being accelerated and collimated by accretion disks has been largely developed prior to any HST observations, HST has provided the first *direct* evidence for the fact that jets indeed originate at the centers of disks (at least in the case of YSOs). Observations of the objects HH 30 (Burrows *et al.*, 1996, Fig. 10), DG Tau (Fig. 9), and a few other YSOs clearly show the jet emanating from the disk center, and they reveal even the illuminated upper and lower disk surfaces.

Furthermore, the high resolution of HST has enabled the determination of proper motions in YSO jets, from images obtained over a time interval of a few years (e.g., Bally, 2003). The highest velocities have been observed along the jet axis. Typically, the velocities are of order 200 to 400 km s^{-1}. Since these are precisely of the order of the Keplerian velocity in the inner disk around a YSO, the proper motion observations provide additional support for the jet formation scenario described above. Incidentally, similar observations of the optical jet in the active galaxy M87 also showed proper motion. In this case at the apparent "superluminal" speed was of 4c–6c (Biretta, Sparks, and Macchetto, 1999), again corresponding to the fact that one expects relativistic speeds from the vicinity of a central black hole.

Another related phenomenon that only HST could discover is the existence of externally irradiated small jets ("microjets"; Bally, O'Dell and McCaughrean, 2000). Many of these microjets (Fig. 11) are only about 0.1″ wide, and are therefore usually undetectable against the nebular background in ground-based observations. For irradiated jets, observations of the Hα surface brightness, the emission measure ($EM = \int n_e^2 dx$; where n_e is the electron density and x is the linear size of the emitting region), and the jet width, allow for a determination of the electron density. Together with the jet speed one can therefore obtain the rate of mass loss in the jet, $\dot{M}_j$. Typically, the irradiated small jets in the Orion nebula are characterized by $\dot{M}_j \sim 10^{-9}$ M$_\odot$ yr^{-1}, at least an order of magnitude lower than the rates observed in the long jets associated with Herbig-Haro objects. Generally, the rate of mass loss in jets from YSOs is found to be about 1–10% of the rate at which mass is accreted through the disk onto the young stellar object.

In addition to the important disk-jet connection, HST observations of YSOs have provided another interesting new element, this time related to planet formation.

B. Protoplanetary disks

Contrary perhaps to the expectation that protoplanetary disks would be deeply embedded within the clouds from which they form, and they would therefore be inaccessible to optical observations, HST revealed many dozens of protoplanetary disks ("proplyds"; e.g., Bally, O'Dell and McCaughrean, 2000; O'Dell, 2001; O'Dell *et al.*, 1993; O'Dell and Wong, 1996, following the initial correct identification by Churchwell *et al.* 1987 and Meaburn 1988). Many of these disks are seen silhouetted against the background nebular light (when they are shielded from photoionization), with some possessing ionized skins and tails (e.g., Bally, O'Dell and McCaughrean, 2000; Henney and O'Dell, 1999, Fig. 12).

The ubiquity of the protoplanetary dust disks (they are seen in 55%–97% of stars; Hillebrand *et al.*, 1998; Lada *et al.*, 2000) demonstrates that at least the raw materials for planet formation are in place around many young stars. Indeed, in a few cases, like the dust ring and disk in HR 4796A and the nearly edge-on disk surrounding Beta Pictoris, the detailed HST images reveal gaps and warping (respectively) that could represent the effects of orbiting planets (Kalas *et al.*, 2000; Schneider *et al.*, 1999).

Another aspect of the protoplanetary disks that is significant for planet formation is the discovery of evaporating disks in the Orion Nebula. As was noted in Section IIIA, some of the Orion proplyds were shown to be evaporating (due to photo-ablation by UV radiation from young, nearby stars) at rates of $\dot{M} \sim 10^{-7}$ to 10^{-6} M$_\odot$ yr^{-1} (e.g., Henney and O'Dell, 1999). Given that the masses of these disks are typically of order 10^{-2} $M_\odot$ (if normal interstellar grains are assumed, so that the observed dust emission can be scaled to the total mass), this implies lifetimes for these disks of 10^5 years or less. There exists, however, some evidence that the grain sizes in Orion's disks may, in fact, be relatively large—perhaps of the order of millimeters (Throop, 2000). The latter conclusion is based on the fact that the outer portions of the disks appear to be gray (they do not redden background light), and on the failure to detect the disks at radio wavelengths in spite of the implied large extinction in the infrared (hiding the central star in some cases). The observations are thus consistent with grain sizes in excess of the radio wavelength used, of 1.3 mm. When we think about the potential implications of these two findings (about disk lifetimes and grain sizes), we realize that they may have interesting consequences for the *demographics* of planets in Orion. The relatively short disk lifetimes but relative large grain sizes may mean that while rocky (terrestrial) planets can form in these strongly irradiated environments, giant planets (that require the accretion of hydrogen and helium from the protoplanetary disk) cannot (unless their formation process is extremely fast; Boss, 2000; Mayer *et al.*, 2002). It is nevertheless clear from the many observations of "hot Jupiters" (giant planets with orbital radii $\lesssim 0.05$ AU) that less extreme environments do exist, in which giant planets not only form, but also have sufficient time to gravitationally interact with their parent disk and migrate inward, to produce the distribution in orbital separations we observe today (see, for example, Armitage *et al.*, 2002; Lin, Bodenheimer, and Richardson, 1996).

While disks around young stars produce jets and form planets, similar structures around old stars help perhaps to shape incredible "sculptures" around dying stars.

C. Morphology of stellar deaths

The evolution of stars is determined primarily by their mass, because evolutionary timescales are set by the rate of consumption of the nuclear fuel. Figure 13 shows the evolutionary tracks of 1 $M_\odot$, 5 $M_\odot$, and 25 $M_\odot$ stellar models in the luminosity-effective temperature plane. It is clear that certain phases of stellar evolution are characterized by considerable mass ejection, either via relatively gentle processes such as radiation pressure on atomic lines or on dust, or through stellar explosions. A process of the former type produces the objects known as *planetary nebulae*—shells of ejected matter from giant stars that are ionized by the radiation emitted by the hot ($\gtrsim$30,000 K) central core. The explosions known as Type II supernovae are the result of the dynamical collapse of the iron cores of massive stars, and the ejected nebulae form the objects known as *supernova remnants*.

One of the most striking phenomena revealed by HST observations of stellar deaths and of stars in their late stages of evolution is the fact that *axisymmetric* nebulae are extremely common. This observation remains true even after recognizing that the morphologies are blurred by projection effects, and by the lines used to obtain the image (i.e., the [O III] image may look different than the [N II] image). Furthermore, axisymmetry is found not only in planetary nebulae (PNe), that are formed by intermediate-mass stars (1–8 $M_\odot$), but also around Luminous Blue Variables (LBVs), that are extremely massive stars ($\sim$ 100 $M_\odot$), and supernovae (SNe), that represent the death of massive stars (8–30 $M_\odot$).

Eta Carinae (η Car), for example, belongs to the small class of very massive stars known as Luminous Blue Variables. These stars are believed to represent a rapid evolutionary phase, in which the stars experience severe mass loss (losing many solar masses in $\sim 10^4$ years), sometimes via giant outbursts (e.g., Davidson and Humphries, 1997). In the 1840s η Car suffered such an outburst, increasing in brightness by several magnitudes and ejecting a large amount of material. The star stabilized around 1870 (except for a minor eruption between 1887 and 1895). The precise cause for this giant outburst is still unknown, but it has produced a spectacular, bipolar nebula commonly referred to as the Homunculus (Fig. 14). In addition to the Homunculus, the HST images revealed for the first time the presence of a ragged ejecta disk around the "waist" of the hourglass structure, composed, at least partly, of radial streaks. The total mass in the equatorial disk has been estimated (based on conventional gas to dust ratios) to be at least 0.1–0.2 $M_\odot$, but this could underestimate the mass significantly.

Recent spectroscopic observations of η Car by the Space Telescope Imaging Spectrograph (STIS) on board HST showed that the spectrum could be fitted well using a model with a mass-loss rate of $\sim 10^{-3}$ $M_\odot$ yr^{-1}. The minimum mass of the system is currently estimated to be 120 $M_\odot$ (Hillier *et al.*, 2001).

On the basis of photometric and radial velocity variations (in particular the disappearance of high-excitation lines like He I, Fe III) and x-ray observations, it appears that η Car is a binary system, with an orbital period of 5.52 years (Damineli, Conti, and Lopes, 1997; Damineli *et al.*, 2000; Ishibashi *et al.*, 1999). The mass of the secondary is not known, but it may be less the 30 $M_\odot$ (Hillier *et al.*, 2001).

Eta Carinae is an extremely enigmatic object on many fronts. However, the morphology of its ejected nebula becomes particularly intriguing when we realize that HST observations reveal almost identical nebular morphologies in objects of very different mass and evolutionary history. One of the best known of these is the supernova 1987A (SN 1987A) in the Large Magellanic Clouds (see, for example, McCray, 2003, for a review). The supernova was first observed in February 1987 (and hence 3 years prior to the launch of HST) and was immediately classified as a Type II Supernova (SN II; representing the collapse of the iron core of a massive star) by virtue of its strong hydrogen lines (coming from the hydrogen-rich envelope). For the first time, the detection of neutrino events (formed copiously as matter rapidly neutronizes) directly confirmed the association between SNe II and core collapses of massive stars. The exploding star itself, SK $-69°202$, had actually been observed prior to the explosion to be a B3 blue supergiant, with a luminosity of $L \simeq 1.1 \times 10^5$ $L_\odot$. Since the HST launch, SN 1987A has become a prime target for the telescope, being the nearest supernova in modern times. The HST observations have revealed a remarkable system of circumstellar rings surrounding the bright center (Fig. 15; Burrows *et al.*, 1995; Pun, 1997). These rings reflect the morphology of material ejected by the supernova progenitor a couple of tens of thousands of years before the explosion (Burrows *et al.*, 1995). This can be inferred from the fact that the central ring is expanding at about 10 km s^{-1} (Crotts and Heathcote, 1991) and it currently has a radius of about 6.3×10^{17} cm. While a full explanation for the formation of the rings is still lacking, there is very little doubt that what we are observing is a bipolar structure, in which the inner ring marks the narrow "waist," while the larger rings are somehow "painted" on the bipolar lobes, or possibly mark their edges. The entire structure is thus very similar to the one observed in η Car.

Bipolar structures have not been restricted only to massive stars, however. Some planetary nebulae and symbiotic nebulae exhibit morphologies that are almost identical to those of η Car and SN 1987A. Some of the best examples are probably My Cn18 (Fig. 16; the "hourglass" nebula, usually classified as a planetary nebula), and the "Southern Crab" (Fig. 17), now recognized as a symbiotic nebula (Corradi *et al.*, 2001).

Planetary nebulae represent the late stages in the lives of stars of about 1–8 $M_\odot$. At that phase, the stars eject their outer envelopes exposing the hot cores ($T \gtrsim 30,000$ K), which, in turn, ionize the nebulae causing them to fluoresce. Symbiotic nebulae have at their centers symbiotic binary systems, consisting typically of a red supergiant and a white dwarf that provides the ionizing radiation.

The main question that arises, therefore, is: What is (are) the mechanism(s) that is (are) capable of producing such bipolar morphologies in stars of different masses, and different ages and evolutionary histories? The two main mechanisms that have been proposed are: (1) *Interacting winds* in the presence of an equatorial to polar *density contrast*, and (2) *Magnetic tension* of a toroidal field (see Balick and Frank, 2002, for a more extensive discussion).

Let us first examine the interacting winds model. The original "interacting winds" model for planetary nebulae (Kahn, 1982; Kwok, 1982) suggested that old, intermediate-mass stars, in the phase of evolution known as the asymptotic giant branch (AGB), first emit a slow (~ 20 km s^{-1}; of the order of the escape velocity from the AGB star's surface) wind, followed by a fast wind (~ 1000 km s^{-1}), once the hot and compact nucleus (the AGB star's core) is exposed. The fast wind catches up with the slowly moving material and shocks it. Balick (1987) proposed that when the interacting winds are allowed to operate in the presence of a *density contrast* between the equator and the pole, a variety of axially symmetric morphologies can be obtained. The idea is that, for reasons that will be explored below, the slow wind contains a non-spherical density distribution, with material being denser around the equator than in the polar direction. Consequently, the (spherically symmetric) fast wind can penetrate more easily at the poles, forming an axisymmetric nebula. Numerical simulations have shown that when a range of density contrasts (between the equatorial and polar directions) is used, and, in addition, the nebular inclination with respect to the line of sight is taken into consideration, most of the observed morphologies can be reproduced (Dwarkadas *et al.*, 1996; Frank *et al.*, 1993; Icke *et al.*, 1992; Mellema, 1995; Soker and Livio, 1989).

The second class of models involves the action of a magnetic field.

The toroidal field component in an outflow from a star is given by (Parks, 1991)

$$B_\phi = B_S \frac{V_{\rm rot}}{V_W(\theta)} \left(\frac{R_S}{r}\right)^2 \left(\frac{r}{R_S} - 1\right) \sin\theta \quad , \tag{11}$$

where R_S, B_S are the stellar radius and the surface magnetic field, $V_{\rm rot}$ is the equatorial rotational velocity and $V_W(\theta)$ is the wind terminal velocity. Consequently, the ratio of the toroidal to radial component increases like (r/R_S) at large distances, leading potentially to an axisymmetric configuration, by the fact that magnetic stresses can slow down the flow in the equatorial direction (while not interfering with the polar direction). The key physical parameters determining the obtained morphology are the stellar rotation rate, the ionizing radiation, and the stellar magnetic field. These can be expressed by: $\Omega_{\rm rot}/\Omega_{\rm crit}$, F_*, and σ (Chevalier and Luo, 1994; Garcia-Segura *et al.*, 1999). Here $\Omega_{\rm crit}$ is the critical (Keplerian) angular velocity, $\Omega_{\rm crit} = (GM_S/R_S^3)^{1/2}$, F_* is the average flux of Lyα photons (of order 10^{45}–10^{47} s^{-1} for typical PNe nuclei), and σ is the ratio of the magnetic energy density to the kinetic energy density in the wind (e.g., Begelman and Li, 1992)

$$\sigma = \frac{B^2}{4\pi\rho V_W^2} = \frac{B_S^2 R_S^2}{\dot{M} V_W}\left(\frac{V_{\rm rot}}{V_W}\right)^2 \quad . \tag{12}$$

Let us now examine the effects of each one of these parameters.

For the rotation to have a significant effect and produce a bipolar morphology, the star needs to rotate at a significant fraction of its breakup speed ($\Omega_{\rm rot}/\Omega_{\rm crit} \gtrsim 0.5$). Under these conditions, conservation of angular momentum results in a significant focusing of the wind towards the equatorial plane, leading to an equatorially compressed outflow (e.g., Bjorkman and Cassinelli, 1993; Livio, 1994; Owocki *et al.*, 1994). The concomitant equator to pole density contrast leads to bipolar morphologies.

Ionization does not, in itself, produce a bipolar morphology. Rather, ionization fronts tend to excite instabilities (similar to the Rayleigh-Taylor instability or the instability discussed by Vishniac, 1983), which, in turn, produce finger-shaped structures and dense "knots." Observations of a number of relatively nearby planetary nebulae, and in particular of the "Helix" nebula, reveal that such dense knots are probably very common (O'Dell *et al.*, 2003; Speck *et al.*, 2002).

The toroidal magnetic field is carried by the fast wind and it can, in principle, produce a bipolar morphology even if the slow wind is spherically symmetric. Basically, as the magnetic energy density of the shocked wind becomes larger than the thermal energy density, the flow becomes (due to the increasing importance of the toroidal component) bipolar, with an increased collimation as the value of σ is increased. Numerical simulations (Garcia-Segura *et al.*, 1999) show that qualitatively, the morphologies obtained for different combinations of $\Omega_{\rm rot}/\Omega_{\rm crit}$ and σ are as shown in Fig. 18. The minimum field required to produce magnetic shaping was found to be of order (Chevalier and Luo, 1994)

$$B_S^{\min} \simeq 11G\left(\frac{\sigma}{10^{-4}}\right)^{1/2}\left(\frac{\dot{M}}{10^{-8}\ \mathrm{M}_\odot\ \mathrm{yr}^{-1}}\right)^{1/2}\left(\frac{V_{\rm FW}}{2000\ \mathrm{km\ s}^{-1}}\right)^{1/2}\left(\frac{R_S}{10^{11}\ \mathrm{cm}}\right)^{-1}\left(\frac{V_{\rm rot}/V_{\rm FW}}{0.1}\right)^{-1}, \tag{13}$$

where $V_{\rm FW}$ is the velocity of the fast wind.

From eq. 13 we see that if the star rotates too slowly, the minimum required field may be unattainable. Similarly, if the star does not rotate at a significant fraction of its breakup speed, an equatorially compressed outflow is not formed. The question of: What is the mechanism that produces highly bipolar outflows? can therefore be reduced to: What causes the star to rotate close to breakup? or: What can produce a strong density contrast between the equatorial and polar directions in the slow wind?

One obvious possibility is: binary companions!

Companions to the central star can act in several ways to aid in the formation of bipolar morphologies: (1) For binaries that were initially relatively close (separation less than $\sim 1000\ \mathrm{R}_\odot$), so that the primary could fill its Roche lobe (the critical potential surface beyond which mass transfer onto the companion occurs) during the asymptotic giant branch (AGB) phase, an unstable mass transfer ensues. As a result, the companion and the AGB star's core start spiralling-in inside a common envelope (see, e.g., a review by Iben and Livio, 1993). This has two effects. First, the envelope of the primary can be spun-up to angular velocities of the order of

$$\frac{\Omega_{\rm rot}}{\Omega_{\rm crit}} \simeq 0.1\left(\frac{M_C}{0.01\ \mathrm{M}_\odot}\right)\left(\frac{M_{\rm env}}{\mathrm{M}_\odot}\right)^{-1}\left(\frac{K_g^2}{0.1}\right)^{-1}\left(\frac{a}{R_S}\right)^{1/2}, \tag{14}$$

where M_C is the companion's mass, $M_{\rm env}$ is the mass of the giant's envelope, $M_S K_g^2 R_S^2$ is the star's moment of inertia and a is the initial separation between the giant and the companion. Equation 14 shows that even brown dwarf (sub-stellar) companions can bring the envelope close to critical rotation. Second, since the envelope mass is ejected (due to orbital energy deposition) primarily close to the orbital plane (because angular momentum is also deposited into the envelope), the common envelope phase can generate an equator-to-pole density contrast. Hydrodynamic simulations of common envelope evolution reveal that during the late stages about 80% of the mass is ejected within 30° of the binary orbital plane (e.g., Rasio and Livio, 1996; Terman, Taam, and Hernquist, 1994, 1995). One can expect that due to cooling, the mass will sink even more toward the orbital plane at later times.

Another possibility for the presence of a higher equatorial density, in principle at least, that does not involve binary companions, is the inner rim of the protostellar disk. If the outer part of the protostellar disk survives till late stages in the stellar life (which may not be difficult in the case of massive, short-lived stars), then the fast wind could interact with the inner rim of this disk (Pringle, 1989). In a few planetary nebulae (for example, the "Red Rectangle" and the "Egg Nebula," Bond *et al.*, 1996; Thompson *et al.*, 1997, respectively), HST observations reveal the presence of relatively large disks, similar to the ones observed in young stellar objects.

In at least some symbiotic nebulae (e.g., M2−9), the bipolar morphology may reflect the action of the white dwarf companion. The white dwarf accretes from the wind of the AGB star, an accretion disk is formed, and the disk powers a mildly collimated fast wind, which in turn produces the bipolar morphology (Livio and Soker, 2001; Soker and Rappaport, 2001).

The conclusion from this discussion is that several mechanisms are capable, in principle at least, to produce the observed bipolar morphologies. Different mechanisms may be operating in different systems. In some cases, we can look forward to the future and expect more definitive answers to emerge. For example, in SN 1987A, the supernova blast wave will eventually hit the entire inner ring, and the luminosity that will be generated by this interaction will illuminate the entire SN vicinity (McCray, 2003). A few brightening spots, where the blast wave has already hit protrusions on the ring, have been observed by HST (Fig. 19; see also Panagia, 2002). Generally, a transmitted shock at normal incidence is expected to propagate into the ring with a speed of $V_{\rm ring} \simeq (n_o/n_{\rm ring})^{1/2} V_{\rm blast}$ (where n_o, $n_{\rm ring}$ are the number densities of the circumstellar matter and the ring, respectively, and $V_{\rm blast}$ is the blast wave velocity). For SN 1987A, $V_{\rm blast} \sim 4000$ km s^{-1}, $n_o \sim 150$ cm^{-3}, $n_{\rm ring} \sim 10^4$ cm^{-3} giving $V_{\rm ring} \sim 500$ km s^{-1}. Eventually, the Hα flux from the ring is expected to be more than 30 times higher than today ($F_{\rm H\alpha} \gtrsim 3 \times 10^{-12}$ erg cm^{-2} s^{-1}), and even brighter in UV lines (Luo, McCray, and Slavin, 1994). This ionizing flux will turn the circumstellar matter into an emission nebula, thus revealing its distribution and velocity field, and hopefully allowing for a reconstruction of the mass-loss history of the system. In all of this, HST will provide a front seat view.

The collapses of very massive stars are believed to produce stellar-mass black holes. In some cases, these collapses produce the most spectacular explosions in the universe since the "big bang"—Gamma Ray Bursts. The Hubble Space Telescope has played an important role in the attempts to understand the nature of these dramatic explosions. However, much more massive black holes are also produced in the universe quite commonly, and the questions of how they form and evolve have intrigued astronomers for decades. The initial answers had to come from large collections of stars like galaxies and stellar clusters.

Given the unprecedented resolution that HST provides, very crowded fields, like the centers of galaxies, were obvious targets. In particular, the early suggestive evidence that galactic centers harbor supermassive black holes (see Kormendy and Richstone, 1995, for a review), virtually ensured that centers of galaxies will be observed extensively with HST. Indeed, these observations were carried out and proved to be extremely fruitful.

V. BLACK HOLES—FROM SUPERMASSIVE TO STELLAR

The idea that active galactic nuclei (AGNs) are powered by accretion at high rates (up to a few solar masses per year) onto supermassive (10^6–10^8 $M_\odot$) central black holes has been around for a long time (e.g., Zeldovich and Novikov, 1964). Furthermore, it has long been realized that collapse into a black hole may be inevitable in the center of many dense stellar environments, once the central potential gets sufficiently deep (for example, Lynden-Bell, 1969; Rees, 1978). Consequently, one might expect supermassive black holes to reside at the center of most galaxies.

In addition, searches for the host galaxies of quasars (thought to represent accreting supermassive black holes) have been going on for a long time, following the pioneering works of Kristian (1973), Wyckoff *et al.* (1980), Hutchings and Campbell (1983), and Hutchings, Janson and Neff (1989). The ground-based studies not only detected host galaxies, but also suggested that some of the hosts have experienced tidal interactions (Boroson, Oke and Green, 1982; Kukula *et al.*, 1996).

These tentative detections and suggestions have turned into reality with the spectacular, high-resolution images of quasars obtained with HST (e.g., Bahcall *et al.*, 1997; Bahcall, Kirhakos, and Schneider, 1994; Kirhakos *et al.*, 1999). The HST images revealed unambiguously that nine radio-loud quasars reside either in bright elliptical hosts or in strongly interacting systems. These findings probably indicate that interactions with neighboring galaxies play an important role in the fueling of the central black hole in active galaxies (see Krolik, 1999a).

A. Search techniques for supermassive black holes

The actual *searches* for supermassive black holes rely primarily on stellar- and gas-dynamical evidence (see, e.g., Kormendy and Richstone, 1995, for a review). In particular, the idea is to unambiguously show that the mass-to-light ratio, M/L, increases toward the galactic center to values that are difficult to accomodate with other types of stellar populations. Ideally, one would want to follow this up with the detection of relativistic speeds, but at present even HST cannot resolve orbits at a few gravitational (Schwarzschild) radii ($R_S = 2GM_{\rm BH}/c^2$; where $M_{\rm BH}$ is the black hole mass, G is the gravitational constant and c is the speed of light).

Generally, the basic principle behind the early stellar-dyamical search techniques can be explained by the following, simplified picture. Taking the first velocity moment of the collisionless Boltzmann equation gives

$$M(r) = \frac{V_{\rm rot}^2 r}{G} + \frac{\sigma_r^2 r}{G}\left[-\frac{dln\rho_t}{dlnr} - \frac{dln\sigma_r^2}{dlnr} - \left(1 - \frac{\sigma_\theta^2}{\sigma_r^2}\right) - \left(1 - \frac{\sigma_\phi^2}{\sigma_r^2}\right)\right] \quad , \tag{15}$$

where $M(r)$ is the mass enclosed within radius r, $V_{\rm rot}$ is the rotational velocity, σ_r, σ_θ, σ_ϕ are the components of the velocity dispersion, and ρ_t is the density of the tracer stars being observed (usually assumed to be proportional to the volume brightness). A direct measurement of $\vec{V}_{\rm rot}$ and $\vec{\sigma}$, therefore can determine the central mass. To actually use eq. 15, however, the ranges of *unprojected* quantities (e.g., $\vec{V}_{\rm rot}$, $\vec{\sigma}$) need to be derived, and various techniques to achieve that have been developed (see for example, Dressler and Richstone, 1988; Gerhard, 1993; Kormendy, 1988; van der Marel, 1994).

The more recent search techniques fit axisymmetric, three-integral dynamical models of the galaxy (using the line-of-sight velocity distribution), to the observed light distribution. Basically, the orbits in the (R, z) plane in realistic galactic potentials are often found to have in addition to the two integrals of motion E (the energy) and L_z (the z-component of the angular momentum, where z is the symmetry axis), a third integral, I_3, that can be associated with the approximately conserved total angular momentum, L.

The dynamical model assumes axisymmetry and an inclination for the galaxy and first determines the optimal density distributrion that is consistent with the surface brightness distribution (making certain assumptions about M/L_V). Then, a central point mass is added, and the potential calculated. Orbits that sample the phase space of the three integrals of motion are then calculated, and the data of the full line-of-sight velocity distribution are fitted to the models.

Since galactic gas is affected also by forces which are non-gravitational (e.g., radiation pressure), stellar kinematics are considered more secure than gas dynamics in black hole searches. Nevertheless, in a few cases, the presence of the black hole may be revealed by gas-dynamical searches. The prototype of this technique is provided by the radio galaxy M87. Early stellar-dynamical observations revealed that the velocity dispersion continues to rise inward, to $r \simeq 1''.5$ (e.g., Lauer *et al.*, 1992; Sargent *et al.*, 1978). However, due to the expected anisotropy in the velocity dispersion, models without a black hole could also fit the observations (e.g., Binney and Mamon, 1982; Dressler and Richstone, 1990; van der Marel, 1994). In particular, the last author found $\sigma \simeq 400$ km s^{-1} at $r \lesssim 0''.5$, which could nevertheless be fitted with anisotropic models that do not include a black hole.

A more definitive answer, however, came in this case from the gas. High-resolution HST observations of the nucleus revealed the presence of a small gas disk (about 20 pc in radius), with a major axis perpendicular to the optical jet (Fig. 20). Spectra taken at two opposite points along the major axis (at a luminosity-weighted mean radius of 16 pc) found emission lines separated by $2V = 916$ km s^{-1} (Ford *et al.*, 1994; Harms *et al.*, 1994). For an inclination angle of the disk of $\sim 42^\circ$ (implied by the observed axis ratio), the observed velocity amplitude (if interpreted as a circular Keplerian motion) corresponds to a dark mass of $M_{\rm BH} \simeq 3 \times 10^9$ M$_\odot$ at the center of M87. These results have been further confirmed by Macchetto *et al.* (1997), making the black hole in M87 the most massive observed so far (with a relatively secure mass determination). I should caution that an airtight case for circular Keplerian motion is still to be made for M87. Incidentally, the best case for a black hole based on gas dynamics is for the modest active galactic nucleus NGC 4258. In that case, Very Long Baseline Array (VLBA) observations of high-velocity masers show a rotation curve that can be fitted remarkably well with $V_{\rm rot}(r) \simeq (832 \pm 2)(r/0.25\ {\rm pc})^{-1/2}$ km s^{-1}, implying a black hole mass (mass interior to 0.18 pc) of $M_{\rm BH} \simeq 4.1 \times 10^7$ M$_\odot$.

B. The $M_{\rm BH} - \sigma$ relationship

Using three-integral models and HST spectroscopy, Gebhardt *et al.* (2000) were able to determine black hole masses in 16 galaxies. Adding to these two galaxies with maser mass determinations (NGC 4258 and NGC 1068), six galaxies with black hole masses determined by gas dynamics, and our own galaxy (Genzel *et al.*, 2000; Ghez *et al.*, 1998) and M31 (Dressler and Richstone, 1988), they were able to discover a tight correlation between the black hole mass and the velocity dispersion in the galactic bulge. Specifically, they found that

$$M_{\rm BH} = 1.2 \times 10^{8}\ {\rm M}_{\odot} \left(\frac{\sigma_e}{200\ {\rm km\ s}^{-1}} \right)^{3.75} \quad , \tag{16}$$

where σ_e is the line of sight aperture dispersion within the half-light radius R_e. The obtained relation is much tighter (Fig. 21, right) than a previously determined relation (e.g., Magorrian *et al.*, 1998) between black hole mass and the bulge luminosity (Fig. 21, left). Based on a smaller sample (12 galaxies), Ferrarese and Merritt (2000) found independently a $M_{\rm BH} \propto \sigma^{\alpha}$ relation, with a somewhat steeper slope of $\alpha \sim 4.8$.

Clearly, the observed tight $M_{\rm BH} - \sigma$ correlation strongly suggests that the formation and evolution of the bulge and of the black hole are causally connected. The precise nature of this connection, however, is still a matter of considerable uncertainty. A relatively simple theoretical model for the relation has been suggested by Adams, Graff and Richstone (2001).

In this model, a slowly rotating isothermal sphere (with a seed central black hole) collapses to form the bulge (the dark matter is assumed to move in tandem with the baryons). The density distribution is assumed to be of the form

$$\rho(r) = \frac{c_s^2}{2\pi G r^2} \quad , \tag{17}$$

where c_s is the speed of sound. Note that for dissipationless collapse the velocity dispersion is roughly given by $\sigma^2 \simeq 2c_s^2$. The region is assumed to rotate rigidly (due to effective tidal torques) at an angular speed Ω, and the specific angular momentum is $j = r_\infty^2 \Omega \sin^2\theta$, where r_∞ is the initial radius and θ is the polar angle. For zero-energy orbits, the pericenter distance in the equatorial plane is therefore

$$p = \frac{j^2}{2GM} = \frac{r_\infty^4 \Omega^2}{2GM} = \frac{(GM)^3 \Omega^2}{32 c_s^8} \quad , \tag{18}$$

where in the last equality we used the fact that for the assumed density distribution $M(r) = 2c_s^2 r/G$. For material to be captured by the black hole we need $p \le 4R_S$, where R_S is the gravitational (Schwarzschild) radius $R_S = 2GM/c^2$. Using eq. 18, the condition $p = 4R_S$ therefore reads (assuming that the capture condition defines the black hole mass)

$$M_{\rm BH} = \frac{16 c_s^4}{Gc\Omega} = \frac{4}{Gc\Omega}\sigma^4 \quad , \tag{19}$$

in good agreement with the observed relation (eq. 16).

Other considerations result in similar expressions. For example, in a protogalaxy modeled as an isothermal sphere of cold dark matter, with $\rho(r) = \sigma^2/2\pi G r^2$, with a fraction $f_{\rm gas}$ in the form of gas, a central, accreting black hole will generate an intense wind outflow. The black hole itself may be assumed to form by coherent collapse before most of the bulge gas turns into stars. If the black hole radiates at the Eddington luminosity, $L_{\rm EDD} = 4\pi c G M_{\rm BH}/\kappa$ (at which gravity is balanced by radiation pressure; where κ is the electron scattering opacity), and a fraction $f_{\rm out}$ is deposited into kinetic energy of the outflow, then a shell of swept-up material will be moving outward at a speed

$$V_{\rm out} = \left(\frac{8\pi^2 G f_{\rm out} L_{\rm EDD}}{f_{\rm gas}\sigma^2}\right)^{1/3} \quad . \tag{20}$$

The condition that the shell would escape, and therefore, that the black hole would unbind the bulge gas, requires $V_{\rm out} > \sigma$. This implies that the black hole mass is limited by (Silk and Rees, 1998)

$$M_{\rm BH} \simeq \frac{1}{32\pi^3}\frac{f_{\rm gas}}{f_{\rm out}}\frac{\kappa}{G^2 c}\sigma^5 \quad , \tag{21}$$

similar to the relation found by Ferrarese and Merritt (2000).

Semi-analytical, hierarchical galaxy formation models (see Section VI), in which galaxies form by merging halos (and merging central black holes), with simple prescriptions for gas cooling, star formation, and feedback from supernovae, also tend to produce scalings of the form $M_{\rm BH} \sim \sigma^4$ (for example, Haehnelt and Kauffmann, 2000). Broadly speaking, this relation can be traced to the facts that: (i) In mergers, the black hole mass scales with the halo mass. (ii) $\sigma \sim \rho^{1/6} M_{\rm halo}^{1/3}$, and (iii) $M_{\rm halo}$ scales like ρ^{-2} in typical cold-dark-matter cosmologies (really like $\rho^{-2/(3+n)}$, where $n \sim -2$ is the slope of the dark matter fluctuations spectrum). Combining (i)–(iii) gives $M_{\rm BH} \sim \sigma^4$.

Somewhat more exotic scenarios, which involve the accretion of collisional dark matter (invoked to make galactic halos less dense; for example, Spergel and Steinhardt, 2000), also produce black hole masses which scale roughly with $\sigma^{4.5}$ (Ostriker, 2000).

C. Intermediate-mass black holes?

An interesting question is related to the *range* of black hole masses over which the observed relation (16) applies. Most recently, high-spatial-resolution spectroscopy with HST, (Gerssen *et al.*, 2002, 2003; van der Marel *et al.*, 2002), of the central part of the globular cluster M15, revealed evidence for the existence of a dense, central concentration of dark mass (see also Baumgardt *et al.*, 2002; Dull *et al.*, 2002). The interpretation of the nature of this dark mass has been somewhat controversial. Dull *et al.* (2002) and Baumgardt *et al.* (2002) have shown that the sharp rise in M/L that was observed toward the center of M15 could be explained by a central concentration of neutron stars and massive white dwarfs. On the other hand, Gerssen *et al.* (2003) argued that if one were to allow for the fact that a large fraction of the neutron stars born within the cluster cannot be retained (due to "kick" velocities of a few hundred km s^{-1} at birth), the observed M/L might require the existence of a central black hole of mass $M_{\rm BH} = 1.7^{+2.7}_{-1.7} \times 10^3$ M$_\odot$. Interestingly, the derived mass fits quite well onto the $M_{\rm BH} - \sigma$ relationship (Gerssen *et al.*, 2002; Tremaine *et al.*, 2002). Similarly, evidence for a central black hole with a mass of $\sim 2 \times 10^4$ M$_\odot$ has been found in the stellar cluster G1 in the Andromeda galaxy M31 (Gebhardt, Rich, and Ho, 2002). The latter black hole also fits the $M_{\rm BH} - \sigma$ relation. In an entirely independent work (D'Amico *et al.*, 2002), pulse timing observations determined positions for five millisecond pulsars in the cluster NGC 6752 with a 20 mas accuracy. Three were found to have line-of-sight accelerations larger than the maximum value predicted on the basis of the central mass density derived from optical observations. The measured accelerations thus provided dynamical evidence for a central mass-to-light ratio of $M/L \gtrsim 10$. All of these findings, *if confirmed*, are extremely exciting, since the implied masses make these black holes *intermediate* between the stellar black holes (with $M_{\rm BH} \sim 10$ M$_\odot$) and the suppermassive ones ($M_{\rm BH} \sim 10^6$–10^9 M$_\odot$).

The potential existence of a class of intermediate-mass black holes has also been suggested following the discovery of ultraluminous x-ray sources (e.g., Colbert and Mushotzky, 1999). The latter are x-ray sources outside the nuclei of external galaxies, with luminosities in excess of 10^{39} erg s^{-1}. These sources were originally discovered by the Einstein satellite (Fabbiano, 1989), but have been found in large numbers by the ROSAT and Chandra observatories (e.g., Colbert and Mushotzky, 1999; Jeltema *et al.*, 2002; Lira, Johnson, and Lawrence, 2002). More recently, however, it has been pointed out that although the ultraluminous x-ray sources may form a heterogeneous class of systems, the most likely explanation for the majority of them is that they constitute the high-luminosity tail of the stellar-mass black-hole binary distribution (e.g., King *et al.*, 2001; Podsiadlowski *et al.*, 2002; Roberts *et al.*, 2001). In particular, Podsiadlowski *et al.* (2002) have shown that in binary systems in which the donor star becomes a giant and the evolution is driven by the nuclear evolution of the hydrogen-burning shell, luminosities that are potentially as high as $\sim 10^{41}$ erg s^{-1} can be obtained. The existence of ultraluminous x-ray sources by itself, therefore, may not imply the existence of intermediate-mass black holes.

The potential existence of intermediate-mass black holes also led to a new scenario for the formation of the supermassive ones (Ebisuzaki *et al.*, 2001). The idea is that first, intermediate-mass black holes form in young clusters, due to runaway mergers of massive stars. The massive stars are assumed to sink to the cluster center due to dynamical friction, on a timescale of (e.g., Binney and Tremaine, 1987)

$$t_{\rm df} = \frac{1.17}{\log\Lambda}\frac{r^2\sigma}{Gm} \simeq 2.7\times 10^7 \left(\frac{r}{1\ \rm pc}\right)^2 \left(\frac{r_h}{10\ \rm pc}\right)^{-1/2} \left(\frac{M}{10^6\ \rm M_\odot}\right)^{1/2} \left(\frac{m}{20\ \rm M_\odot}\right)^{-1} \ \rm yr \quad , \qquad (22)$$

where $\log\Lambda$ is the Coulomb logarithm, σ is the velocity dispersion, r is the distance to the cluster center, r_h is the half-mass radius and M and m are the masses of the cluster and the star, respectively. During the same time that the intermediate-mass black holes are forming, the host clusters themselves sink to the galactic center (again due to dynamical friction), evaporate, and deposit their black holes. The latter can then form black hole binaries which merge due to gravitational radiation, eventually leading to the formation of supermassive black holes. I should note that this is only one path out of the many that have been suggested over the years for the formation of supermassive black holes (Rees, 1984).

The installation of HST's Advanced Camera for Surveys (ACS) in March 2002, with its superior sensitivity (by a factor of about 3–5 over the previous Wide Field Planetary Camera 2) and increased resolution (in the high-resolution channel), combined with X-ray observations by the Chandra and XMM-Newton Observatories, and infrared observations of dust-obscured active galactic nuclei by the Space Infrared Telescope Facility (SIRTF) promise that the study of black holes in clusters (if they exist) and in galactic centers is only beginning. In particular, future observations will help clarify the relation between galaxy evolution in general, and the formation and evolution of the cental black holes.

Observations with HST have provided important insights not only in the study of supermassive black holes, but also in researches related to the probable formation of stellar-mass black holes—the study of gamma-ray bursters.

D. Gamma-ray bursts

Gamma-ray bursts were first detected by the military Vela satellites in 1967. These are short flashes of gamma-rays (typically with a peak energy around 100 keV), lasting between a few milliseconds and tens of minutes. Already in the early 1990s, the BATSE experiment on board the Compton Gamma-Ray Observatory demonstrated that the rate of gamma-ray bursts is about 1–2 per day, and that they are distributed isotropically in the sky. The data from BATSE also showed that gamma-ray bursts (GRBs) come in two distinct classes: (i) short-duration ($\lesssim 2$ sec) with hard spectra, and (ii) long duration with softer spectra (Fishman, 2001; Kouveliotou *et al.*, 1993).

Due to the relatively poor localization capabilities of gamma-ray detectors (which did not allow identification of the sources in other wavebands), two thousand bursts have been detected before it became possible to determine the distance to the bursts. The isotropic sky distribution argued for either a cosmological origin or an extremely local one (e.g., the Galactic halo). This situation changed dramatically since the operation of the Italian-Dutch BeppoSAX satellite in 1997. BeppoSAX was able to determine the positions of bursts to within arcminutes. The discovery of rapidly declining "afterglows" in the X-ray, optical and radio bands (Costa *et al.*, 1997; Frail *et al.*, 1999; Metzger *et al.*, 1997; van Paradijs *et al.*, 1997) allowed for redshift determinations that immediately placed GRBs at cosmoligical distances. I should note that to date, only afterglows of the long-duration bursts have been observed. Thus, it is not even clear if the short bursts produce afterglows (the best limit to date, for the burst GRB 020531, failed to detect an afterglow candidate [down to $V \sim 25$] about 20 hours after the burst; Salamanca *et al.*, 2002). Typically, the afterglows decay with time as $t^{-\alpha}$, with α in the ranges 1–2. This behavior was *predicted* by a model in which a fireball is expanding into a homogeneous external medium (Mészáros and Rees, 1997).

At the time of this writing redshifts have been determined to more than two dozen GRBs, and they usually lie in the $z = 0.5$–1.5 range (although redshifts as high as 4.5 have been recorded). The observed fluxes imply energies of up to 10^{54} ergs for isotropic emission. However, there is increasing evidence, in the form of kinks in the afterglow light curve, and in polarization detected in a few bursts, that gamma-ray bursts are in fact collimated into narrow jets. If the observer's line of sight is within the jet solid angle, Ω_j, then as long as the Lorentz factor γ satisfies $\gamma \gtrsim \Omega_j^{-1/2}$, the light-cone is within the jet boundary. However, as the jet decelerates, γ eventually drops below $\Omega_j^{-1/2}$. Consequently, the (transverse) emitting area starts to grow more slowly (as $r_{||}\Omega_j^{-1/2}$ instead of $(r_{||}/\gamma)^2$), resulting in a break (faster decay) in the light curve (Livio and Waxman, 2000; Mészáros and Rees, 1999; Rhoads, 1997), consistent with observations in GRB 990123 (Fruchter *et al.*, 1999; Kulkarni *et al.*, 1999). When collimation is taken into account, the average total energy of GRBs is estimated to be of the order of 2×10^{51} ergs (Frail *et al.*, 2001; Panaitescu and Kumar, 2001).

Based on the energetics, and the fact that the production of astrophysical jets typically relies on the collimation and acceleration provided by an accretion disk around a compact object (Livio, 2000), the most popular models for GRBs involve the formation of stellar-mass black holes, surrounded by a debris torus from which mass accretion onto the central object occurs (Mészáros, 2002). The two most likely progenitors to produce such a configuration are the core collapse of massive stars (Paczynski, 1998; Woosley, 1993), or the ongoing merger of neutron star-neutron star or black hole-neutron star binaries (e.g., Eichler *et al.*, 1989; Goodman, 1986; Paczynski, 1986). In both cases the main source of energy is gravitational, even though tapping into the (large) spin energy of the black hole is also possible in principle (Blandford and Znajek, 1977; Krolik, 1999b). A schematic describing how a GRB and its afterglow may be generated by internal and external shocks (respectively) in a collimated jet arising from a stellar collapse, is shown in Figure 22.

There are several pieces of evidence suggesting that the long duration GRBs may indeed be associated with stellar collapses. First, in merging neutron stars (or a black hole and a neutron star) the associated timescales (dictated at that point by the emission of gravitational wave radiation) are probably too short to produce multi-second-long bursts. On the other hand, core collapses that lead to a black hole and an accretion disk are naturally associated with longer timescales. Second, there exists quite strong evidence that at least some GRBs are associated with supernova explosions.

In particular, the supernova SN 1998bw (of the relatively rare type Ic, caused by the explosion of a massive star that had rid itself of its hydrogen envelope prior to exploding; Filippenko and Sargent, 1986) was found to be approximately coincident in both position and time with the relatively weak GRB 980425, at redshift $z = 0.0085$ (Galama *et al.*, 1998). Furthermore, the supernova light curve was found to be consistent with the formation of a black hole (Iwamoto *et al.*, 1998). In a few other bursts (e.g., GRB 011121) the optical and near infrared afterglow light curve exhibits a "bump" with a time delay and amplitude that are consistent with resulting from a supernova explosion occuring simultaneously with the GRB (e.g., Greiner *et al.*, 2001).

The contributions of HST to this exciting field have been in a few areas.

First, by resolving the host galaxies and being able to pinpoint the GRB on the host, HST has shown unambiguously that at least some GRBs are not associated with galactic nuclei. In fact, they sometimes occur far from the nucleus or in spiral arms (Andersen *et al.*, 2002; Sahu *et al.*, 1997). Second, the Hubble Space Telescope has shown (Fruchter, 2002; Fruchter *et al.*, 1999) that the colors of the hosts of GRBs corresond to the bluest colors observed for galaxies in the Hubble Deep Fields (see Section VI). This means that GRBs with afterglows occur preferentially in galaxies with high star formation rates—a finding that is consistent with GRBs being associated with collapses of massive stars. Furthermore, the fact that in almost all cases the GRB's position was found to be within the extent of the rest-frame ultraviolet image (where star formation is intense) of the hosts, also argues in favor of the "collapsar" model for the long-duration bursts. Neutron star binaries, on the other hand, are born with "kicks" of as much as a few hundred km/sec, which could drive them out of the star-forming regions.

The Hubble Space Telescope also helped to firm the association of at least some GRBs with supernova explosions. In GRB 011121, for example, HST detected an intermediate-time flux excess ("red bump") that was redder in color relative to the GRB afterglow. This "bump" could be well described by a redshifted Type Ic supernova that exploded approximately at the same time as the GRB (Bloom *et al.*, 2002; Garnavich *et al.*, 2003). Near-infrared and radio observations of the afterglow further provided evidence for extensive mass loss ($\dot{M} \sim 2\times10^{-7}$ $M_\odot$ yr^{-1}) from the massive stellar progenitor of GRB 011121 (Price *et al.*, 2002).

To conclude this topic, it is very likely that *every time a GRB goes off, a black hole is born.* Gamma Ray Bursts therefore offer, in principle at least, a direct measure of the massive-star formation rate in the early universe. Future observations with the HETE-2 spacecraft and with the Swift multi-wavelength GRB afterglow mission (equipped with γ-ray, X-ray and optical detectors), as well as follow-ups with HST and other observatories from the ground and space, will hopefully reveal the true nature of these most dramatic explosions (including the short-duration bursts).

VI. DEEP FIELDS AND A BRIEF COSMIC HISTORY

One of the major goals of observational cosmology is to understand the processes involved in galaxy formation and evolution. A full theoretical treatment of this problem requires an understanding of the primordial density fluctuations, of the formation of dark matter halos, of the dissipation that occurs in the cooling gas, of the processes involved in star formation, of the feedback between stellar explosions and the interstellar medium, of galaxy mergers, and of the interactions of galaxies with intergalactic gas.

Most current models for the formation of structure in the universe assume that dark matter halos build up hierarchically, with the assembly being controled by the cosmological parameters, the power spectrum of the density fluctuations, and the nature of the dark matter itself. The build-up of the stellar mass *within* galaxies is a secondary process, in which gaseous dissipation, the more intricate physics of star formation, and feedback, all play a role.

Since galaxy formation involves non-linear processes on many scales, the problem has been addressed theoretically, mainly using large-scale N-body hydrodynamic simulations (e.g., Pearce *et al.*, 2001). On the observational side, HST, the Chandra Observatory and the Keck and VLT telescopes have allowed for an unprecedented view of the high-redshift universe, probing galaxies from infancy to old age.

Short of the awsome pillars of dust and molecular gas revealed by the HST images of the "Eagle Nebula," the best known images that Hubble has produced are those of the "Hubble Deep Fields" (HDFs). In fact, the first HDF and the follow-up observations of the same field by other observatories probably represent the most concentrated research effort ever in astronomy into what was previously a blank piece of the sky!

Observations taken shortly after the first servicing mission, which restored the HST optics capabilities, demonstrated that HST could resolve distant galaxies spectacularly well. In particular, observations of the relatively distant clusters CL 0939+4713 (Dressler *et al.*, 1994) and the cluster around the radio galaxy 3C324 (Dickinson, 1995) revealed faint galaxies of small angular sizes, the morphologies of which were entirely inaccessible from the ground.

The idea of the Hubble Deep Field North—ten days of one continuous observation of a field 2.6 arcminutes on the side in December 1995—was conceived by the then Director of the Space Telescope Science Institute, Bob Williams. Williams decided to use his Director's

Discretionary Time on HST to produce the deepest image of the universe in optical/UV wavelengths.

The northern field itself (Fig. 23) was carefully selected so as to be in HST's northern continuous viewing zone (CVZ; without interference by Earth occultations), to be free of bright stars, nearby galaxies and radio sources, and to have relatively low Galactic extinction. The southern field (Fig. 24) was selected so as to include a quasi-stellar object (QSO) that could be used to study absorption systems along the line of sight. The HDF-N observations were taken in December 1995 and the HDF-S in October 1998.

The filter selection for the observations was driven partly by the desire for depth and color information (to identify high-redshift galaxies by their Lyman-break), and partly by practical considerations involving scattered light within the telescope. Accordingly, images were taken in four broad-band passes, spanning a wavelength range from about 2500 Å to 9000 Å.

In order to achieve a higher resolution than the detectors' pixel sizes, the pointing of the telescope was dithered (shifted slightly) during the observation, thus ensuring that images were recorded on different pixels. A "drizzling" (subpixel linear reconstruction) technique was developed by Fruchter and Hook (2002). Details on the observational techniques and a review of some of the results can be found in the excellent reviews by Ferguson (1998) and Ferguson, Dickinson, and Williams (2000).

As soon as the HDF-N image was obtained, it was obvious that a new era in astronomy has begun. The image revealed some 3000 galaxies of different colors, shapes, and sizes. A broad summary of key results can be found in Livio, Fall, and Madau (1998). Here I will concentrate only on two main topics, to which HST's contribution has been crucial: (i) galaxy sizes and morphologies (and their implications for galaxy formation and evolution), and (ii) the global, cosmic star-formation history. The dramatic discovery of a particular supernova in the HDF-N will be discussed in Section VIII.

A. The morphology and sizes of high-redshift galaxies

One of the key questions in galaxy formation and evolution is the importance of interactions and mergers. In particular, cold-dark-matter and dark-energy dominated models have as one of their natural consequences a higher merger rate in the past, since in these models, today's Hubble Sequence galaxies have been assembled via a process of hierarchical mergers (e.g., Cole *et al.*, 2000; Kauffmann *et al.*, 1999; Somerville, Primak, and Faber, 2001).

In the local universe, the luminosity density is dominated by ellipticals, lenticulars and spirals, even though dwarf ellipticals and dwarf irregulars (all of low luminosity) dominate by number. The HDF-N allowed for the first time for a morphological classification of galaxies down to a magnitude of $I = 25$ (e.g., Abraham *et al.*, 1996). These studies showed that the fraction of irregular, multiple-component, and peculiar-looking galaxies was indeed considerably higher ($\sim 40\%$) than expected from a direct extrapolation of the numbers at $z \simeq 0$. The trend of a rising fraction of faint systems with irregular morphology was already noted in HST's Medium Deep Survey (Griffiths *et al.*, 1994; Windhorst *et al.*, 1995), although there, the survey reached only down to $I \sim 23$.

One worry one might have when examining the morphologies of galaxies at high redshift is that optical images reflect, in fact, the UV rest frame images of the galaxies. Since the UV light traces, in particular, pockets of intense star formation, one might expect a more irregular appearance for high-redshift galaxies (Giavalisco *et al.*, 1996). A study of the redshift distribution of galaxies with $17 < I < 21.5$ (Im *et al.*, 1999), showed indeed the irregular/peculiar class to consist of both low-redshift dwarfs and intermediate redshift ($0.4 \lesssim z \lesssim 1$) galaxies that are by themselves a mix of starbursts (of relatively low mass) and more massive, interacting galaxies.

Nevertheless, Brinchmann *et al.* (1998) and Abraham *et al.* (1999a) have shown that the redshifting of UV wavelengths into the optical is not sufficient to explain the preponderance of irregular morphologies. Further confirmation that the rapid rise with redshift in the fraction of galaxies with irregular/peculiar morphologies is real, came from the Near Infrared Camera and Multi Object Spectrograph (NICMOS) observations. The NICMOS observations (Dickinson, 2000a,b; Thompson *et al.*, 1999) have demonstrated that with very few (interesting by themselves) exceptions, the irregularities seen in the optical (WFPC2) image persist in the infrared image (Fig. 25), even at redshifts as high as $z \sim 3$. Since the NICMOS observations reflect the rest-frame optical light (up to $z \sim 3$), the absence of any significant changes in the morphology (in the majority of galaxies) proves that the peculiarities are not artifacts of wavelength shifting.

A morphological trend that may be indicated by the HDFs is a significant decline in the fraction of barred spirals with redshift (for $z \gtrsim 0.5$; e.g., Abraham *et al.*, 1999b). However, since bars are more difficult to detect at bluer rest wavelengths, optical surveys of galaxies at high-z may be biased against finding bars (Eskridge *et al.*, 2000; van den Bergh *et al.*, 2002). The precise reason for this trend (if confirmed) is not known, but it may be related either to the state of the disk (the development of a bar instability requires a cold disk), or to the timescale needed to form long-lived bars (perhaps through episodic growth, aided by spiral patterns; Selwood, 2000).

Another extremely interesting result to have come out of the HDFs is related to the *sizes* of galaxies. When examining the HDFs, one is immediately struck by the small angular diameters of the faint galaxies. In particular, No-Evolution models (e.g., Bouwens, Cayon, and Silk, 1997), Pure Luminosity Evolution models (in which galaxies form at some redshift and are characterized by their star-formation history, but no merging occurs; e.g., Metcalfe *et al.*, 1996), and models dominated by low-surface-brightness galaxies (e.g., Ferguson and McGaugh, 1995), all produce half-light radii that are considerably larger than those observed for $24 \lesssim I \lesssim 27.5$. Roche *et al.* (1998), who used profile fitting to determine half-light radii, also find that at $z > 0.35$ galaxies are significantly more compact than predicted by Pure Luminosity Evolution models (although no evolution in the size was found for $z \lesssim 0.35$), and more generally, that galaxies at $z \gtrsim 2$ are more compact than today's L^* galaxies.

There is no question that the most attractive explanation for the higher fraction of morphologically peculiar, and smaller galaxies in the past (beyond $z \simeq 1$ smaller angular sizes correspond to smaller physical sizes, essentially irrespective of the cosmological model), is in terms of hierarchical galaxy formation and interactions. In the context of this model, the highly irregular morphologies are a consequence of interactions (that were more frequent in the denser past), direct collisions, and mergers. In this picture, the smaller size objects are essentially the "building blocks" of today's galaxies.

Some attempts have been made to transform this qualitative statement into a quantitative tool. While the Hubble Sequence classification (Hubble, 1926) has proved extremely useful for analyzing the gross morphological properties of nearby galaxies, it becomes rather ineffective at high redshifts, when mergers and intense star formation become the rule, rather than the exception.

A method that is quite successful in distinguishing between irregular galaxies forming stars stochastically and irregularities induced by mergers is based on color-asymmetry diagrams (e.g., Bershady, Jangren, and Conselice, 2000; Conselice, 1997). In these diagrams, galaxies are placed in the $[(B-V), A]$ plane, where A is an asymmetry parameter, based on comparing the galaxy image with its counterpart obtained through a rotation by 180°. Formally, the asymmetry parameter is defined by

$$A = min\left(\frac{\sum |I_o - I_\phi|}{\sum |I_o|}\right) - min\left(\frac{\sum |B_o - B_\phi|}{\sum |I_o|}\right) \quad , \tag{23}$$

where I_o is the intensity distribution in the image pixels, I_ϕ is the intensity distribution in the rotated image (by angle ϕ), and B is the intensity in background pixels (with the second term correcting for the noise). A calibration using nearby galaxies (Conselice, Bershady, and Jangern, 2000) shows that all the galaxies with $A > 0.35$ represent merging systems. Galaxies with enhanced star formation rates are bluer and may become asymmetric due to pockets of star formation. Nevertheless, their asymmetry parameter does not exceeed 0.35.

An examination of the color-asymmetry diagram of the HDF-N shows that the highest fraction of mergers is found in the highest redshift range, $1.5 < z < 2.5$. Using $A > 0.35$ and $M_B < -18$ as criteria for mergers, gives for the merger fraction as a function of redshift $f \propto (1+z)^{2.1\pm0.5}$ (Conselice, 2001). While considerable uncertainties still exist, this suggests that the evolution of galaxies may indeed be primarily dominated by mergers. Furthermore, the merger fraction appears to be starting to flatten (or possibly even to decline) for $z \gtrsim 2$. As we shall soon see, this may be related to the behavior of the star formation rate with redshift.

Additional information about *hierarchical* vs. *monolithic* formation models comes from studies of elliptical galaxies. In the former scenario, giant ellipticals form via mergers of galaxies of comparable mass that had already (prior to merger) used up at least some of their gas to form stars (e.g., Kauffmann, White, and Guideroni, 1993). In the monolithic scenario, on the other hand (e.g., Eggen, Lynden-Bell, and Sandage, 1962; Tinsley and Gunn, 1976), ellipticals form at high redshifts via a single collapse (and a concomitant starburst), and evolve passively thereafter. The two *pure* models have rather distinct observational predictions. Clearly, in the hierarchical scenario the number density of ellipticals is expected to *decrease* with increasing redshift. This can be contrasted with the predictions of the monolithic scenario: the number density of ellipticals should stay fairly constant with redshift, but the bolometric luminosity is expected in increase (up to the starburst phase). The observations, however, proved to be more ambiguous than one might have hoped. In particular, Treu and Stiavelli (1999) found that while Pure Luminosity Evolution models in which all ellipticals are assumed to form at $z \simeq 5$ over-predict the observed counts, these models tend to under-predict the counts when the formation redshift is assumed to be $z \simeq 2$ (see also Benitez *et al.*, 1999; Franceschini *et al.*, 1998; Zepf, 1997). Thus, the observations with both ground-based surveys and HST suggest a picture which is somewhat intermediate between the *pure* hierarchical and monolithic scenarios. Namely, objects that resemble elliptical galaxies (red, obeying a $R^{1/4}$ law in their luminosity profile) existed already at moderately high ($z \gtrsim 1.5$) redshifts (e.g., McCarthy *et al.*, 2001; Moustakas and Somerville, 2002). These objects continued, however, to experience mergers until $z \sim 1$, from which time on the resulting giant ellipticals evolved mostly passively (see also Giavalisco, 2002).

The Advanced Camera for Surveys, installed on board HST in March 2002, is expected to produce a wealth of high-quality data on galaxy sizes and morphologies. We can therefore expect even more significant constraints on hierarchical formation models to emerge in the coming months.

As I noted above, the evolution of galaxies is also intimately related to the way they assemble their mass and, concomitantly, to their star formation rate. As it turned out, the HDF-N proved to be seminal in the discussion of the cosmic star formation history.

B. The global cosmic star-formation history

The rest-frame UV luminosity of galaxies traces nicely their metal production rate, because both are produced primarily by massive stars. On the other hand, the conversion from a UV luminosity density to an actual cosmic star formation rate is somewhat less straightforward, since it involves a knowledge of the Initial Mass Function (IMF).

Early estimates of the metal production rate as a function of redshift, $\dot{\rho}_Z(z)$, relied on ground-based redshift surveys (Gallego *et al.*, 1995; Lilly *et al.*, 1996) and data from QSO absorption line systems (e.g., Lanzetta *et al.*, 1995; Pei and Fall, 1995). These estimates indicated a monotonic increase in the metal production rate from $z = 0$ to $z \simeq 1$ (with the rate at $z \simeq 1$ being about an order of magnitude higher than at the present).

One of the seminal results to have come out of the HDF-N was the attempt to estimate the metal production rate at high redshifts and thereby to generate a continuous plot of the star formation rate as a function of redshift (Madau *et al.*, 1996). In the original diagram, no corrections were made, for either dust extinction or for surface brightness effects. Consequently, the results were presented as lower limits. Both in the original work and in some of the subsequent work that followed (some of which included the effects of dust extinction), it was found that the star formation rate peaks at a redshift of $z \sim 1$–2 and decreases or stays nearly constant at higher redshifts (e.g., Calzetti and Heckman, 1999; Hopkins, Connolly, and Szalay, 2000; Pei, Fall, and Hauser, 1999; Steidel *et al.*, 1999, Fig. 26). The question of whether there truly is a decrease in the star formation rate for $z \gtrsim 2$ has become a focal point of the discussion of the cosmic star formation history. Some doubts were raised on the basis of dust extinction and selection effects on one hand, and cosmological surface brightness dimming effects on the other. Broadly speaking, the UV luminosity density *could* be underestimated if a significant fraction of the star formation occurred in environments obscured by dust or in very low-surface-brightness galaxies. Several attemps were made to correct for dust attenuation (using the empirical attenuation law of Calzetti, Kinney, and Storchi-Bergmann, 1994), by calibrating the relation between the UV spectral slope and the far IR emission (e.g., Meurer, Heckman, and Calzetti, 1999; Steidel *et al.*, 1999). Others used constraints obtained from extragalactic background radiation and neutral gas (e.g., Calzetti and Heckman, 1999; Pei, Fall, and Hauser, 1999). Most of these investigations concluded that the star formation rate rises from the present to $z \sim 1$–2 and then stays approximately flat to $z \sim 5$.

Another way to address the question of the history of mass assembly in galaxies is to try to measure the actual stellar *masses* of galaxies (as opposed to the *rates* of star formation; ideally one would want to do *both*). To this goal, observations in the near-infrared are typically used, since the near-infrared luminosity traces the stellar mass reasonably well. Dickinson *et al.* (2003) used an infrared-selected sample of galaxies from the HDF-N to determine the global stellar mass density, $\Omega_*(z)$, for $0 < z < 3$. They found that $\Omega_*(z)$ increases with time from $z = 3$ to the present (Fig. 27). Dickinson *et al.* concluded that by $z \sim 1$, about 50–75% of the present-day stellar mass density had already formed, but that the stellar mass density at $z \sim 2.7$ was about 17 times lower than today. These observations appear to be in clear contradiction with scenarios in which most stars in today's spheroids formed at $z \gg 2$, but the observations are in general agreement with a global star formation rate that rises from the present to $z \sim 1$ and then stays fairly flat.

A different spin on the cosmic star formation rate has been put by Lanzetta *et al.* (2002). These authors claimed that by neglecting cosmological surface brightness dimming effects, previous works have missed a significant fraction of the ultraviolet luminosity density at high redshifts. Specifically, since the surface brightness decreases with redshift as $(1+z)^{-3}$ (because of the cosmic expansion), intrinsically faint regions of high-redshift galaxies become undetectable.

Lanzetta *et al.* designated the unobscured star formation rate intensity (i.e., the intensity inferred from the *observed* rest-frame UV light) by x, and used a distribution function $h(x)$ (defined so that $h(x)dx$ is the projected proper area per comoving volume of star formation rate intensity in the interval x to $x+dx$), to estimate the ultraviolet luminosity density at high redshifts (including the surface brightness dimming effects). They found that the star formation rate density, $\dot{\rho}_s$, increases monotonically with redshift to the highest redshifts observed ($z \sim 8$; although in one of the possible corrections for incompleteness $\dot{\rho}_s$ remains fairly flat above $z \sim 2$).

The low fraction of stellar mass formed by $z = 3$ according to the Dickinson *et al.* results appears to contradict evidence for significant star formation occurring at still higher redshifts. One possible way out of this conundrum is that the initial mass function (IMF) at high redshifts is very top-heavy (tilted towards massive stars; Ferguson *et al.*, 2002). Massive stars dominate the UV luminosity in star forming galaxies, but their contribution to the total surviving (at lower redshifts) stellar mass is relatively low.

Most recently, Stanway, Bunker, and McMahon (2003) used HST's Advanced Camera for Surveys, the ground-based Sloan Digital Sky Survey and the Very Large Telescope (VLT) to determine the space density of UV-luminous starburst galaxies at $z \sim 6$. They found a lower bound to the integrated, volume-averaged, global star formation rate at $z \sim 6$, that was about six times less than that at $z \sim 3$–4. The question of the true behavior of $\dot{\rho}_s$ above $z \sim 2$ remains, therefore, presently somewhat unresolved.

Fortunately, a more definitive answer may come in the near future, through a combination of planned observations with HST and with the Space Infrared Telescope Facility (SIRTF; currently scheduled to be launched in April 2003). The Great Observatories Origins Deep Survey (GOODS; Principal Investigator M. Dickinson) will produce a very deep image of two fields (the HDF-N and the southern deep field observed with the Chandra Observatory) with SIRTF at 3.6–24 μm, and will thereby produce a much more complete census of stellar mass at high redshifts. At the same time, observations (of the same fields) with Hubble's Advanced Camera for Surveys (Principal Investigator M. Giavalisco; the observations are being carried out as these lines are being written) will determine the star formation rates, sizes, and morphologies of galaxies. The combined observations will allow for the first time for a determination of the evolving mass assembly distribution $f(M, \dot{M}, t)$ (where M denotes the stellar mass and $\dot{M}$ the star formation rate). Given the fact that under pure luminosity evolution, $M = \int \dot{M} dt$, a comparison between the *observed* evolution of the $(M, \dot{M})$ phase space with time and the evolution obtained from *direct integration* (of $\dot{M}$, to produce M), will allow, in principle, for an identification of the role of mergers and interactions. Even when the expected observational uncertainties are taken into account, there is no doubt that the planned GOODS observations will yield a huge step forward in the understanding of the *assembly* of present-day galaxies, and their *morphological* evolution (the emergence of the "Hubble Sequence"). Furthermore, the planned Hubble Ultra Deep Field with the Advanced Camera for Surveys (currently scheduled for July–August 2003) could extend the redshift coverage unambiguously to $z \sim 6$, close to the tail of the tentative second *reionization epoch* (Becker *et al.*, 2001; Fan *et al.*, 2001) of the universe (the first reionization having tentatively occurred at $z \sim 20^{+10}_{-9}$; Bennett *et al.*, 2003). Such a study could therefore produce results that are not merely incremental in our understanding of the cosmic star formation history, galaxy evolution, and the ionization history of the universe, and that can be used to place meaningful constraints on theoretical models (e.g., Somerville and Livio, 2003).

VII. THE HUBBLE CONSTANT

A. A brief background

Ever since Edwin Hubble's pioneering measurements in the 1920s (and to some extent even before that, since Vesto Slipher's measurements, that started in 1912; e.g., Slipher, 1917), we knew that we live in an expanding universe (Hubble, 1929; Hubble and Humason, 1931). In the standard big bang theory the universe expands uniformly, with the recession velocity being related to the distance through the Hubble law, $v = H_0 d$. More generally, based on the "Cosmological Principle" (the assumption that the universe is homogeneous and isotropic on large scales), the expansion is governed by the Friedmann equation in the context of general relativity

$$\left(\frac{\dot{R}}{R}\right)^2 \equiv H^2 = \frac{8\pi G\rho_M}{3} - \frac{kc^2}{R^2} + \frac{\Lambda c^2}{3} \quad . \tag{24}$$

Here $R(t)$ is the scale factor, $H = \dot{R}/R$ measures expansion rate (with H_0, the "Hubble Constant," giving the rate at present), ρ_M is the mass density, k is the curvature parameter and Λ is Einstein's cosmological constant (which represents the energy density of the vacuum). Commonly, the density of matter and that associated with the vacuum are represented by the density parameters (at present) $\Omega_M = 8\pi G\rho_M/3H_0^2$ and $\Omega_\Lambda = \Lambda c^2/3H_0^2$, with which the Friedmann equation can be expressed as $\frac{kc^2}{R_o^2} = H_0^2(\Omega_M + \Omega_\Lambda - 1)$. The Hubble Constant is thus the key parameter in determining the age of the universe (with Ω_M and Ω_Λ also playing a role). Similarly, physical processes such as the growth of structure and the nucleosynthesis of light elements (H, D, ^{3}He, ^{4}He, Li), as well as critical epochs in the Universe's history, such as the transition from a radiation-dominated to a matter-dominated universe, depend on the cosmic expansion rate and thereby on the value of H_0. It should therefore come as no surprise that the determination of the value of the Hubble Constant became a major observational goal for the past eight decades.

The first value for the Hubble Constant may have actually been derived by Lemaître (1927), who, on the basis of Slipher's radial velocity measurements and Hubble's mean absolute magnitude for galaxies ("nebulae") obtained $H_0 = 526$ km s^{-1} Mpc^{-1}. The next set of values by Hubble (1929) and Hubble and Humason (who produced a velocity-distance relation up to $V \sim 20000$ km s^{-1} and obtained $H_0 = 559$ km s^{-1} Mpc^{-1} in 1931), were all around 500 km s^{-1} Mpc^{-1}, with an uncertainty stated rather naively as "of the area of ten percent." About twenty years passed before Baade (1954) revised the distance to nearby galaxies, recognizing that Hubble confused two classes of "standard candles" (Population I Cepheids and Population II W Virginis stars), thereby reducing the value of H_0 by about a factor two (a revision suggested also by Behr, 1951). The value of the Hubble Constant first reached the range of values accepted today through the work of Alan Sandage (1958). Sandage demonstrated that Hubble mistakenly identified H II regions as bright stars, and he [Sandage] was able to revise the value further to $H_0 \simeq 75$ km s^{-1} Mpc^{-1} (recognizing that the uncertainty could still be by a factor 2).

In the three decades that followed, published values of the Hubble Constant varied by about a factor of two between ~ 100 and 50 km s^{-1} Mpc^{-1}. Table 1, adapted from Trimble

(1997), summarizes the early history of the constant.

Generally, since redshifts (and therefore radial velocities) can be determined relatively readily (this is, of course, not true for the most distant or faintest objects), the problem of determining the Hubble Constant has always been a problem of determining accurate astronomical distances. The availability of new instrumentation, and the Hubble Space Telescope in particular, have allowed for a dramatic improvement in distance determinations.

B. Distance indicators and methods

Direct trigonometric parallaxes that use the Earth's orbit around the Sun as a baseline for triangulation can only be used to the nearest stars. Consequently, other distance indicators and methods had to be used for extragalactic distances (see, e.g., Jacoby *et al.*, 1992; Trimble, 1997, for excellent reviews). The most common of these employ "standard candles," geometrical properties, physical properties, or various correlations to determine distances.

Standard candles are simply based on the fact that the flux of radiation decreases as an inverse square law. Objects with either a constant luminosity or whose luminosity can be related to a distance-independent measurable property (such as an oscillation period) are good standard-candle candidates. The best known and probably most reliable in this class are the Cepheid variables. Their potential as standard candles on the basis of their period-luminosity (P-L) relation was first recognized by Henrietta Leavitt in 1912, and they were used by Hubble (1925) to determine distances to Local Group galaxies. The physical processes responsible for the P-L relation are broadly understood. Near ionization zones (in this case, primarily $\mathrm{He}^{+} \rightleftarrows \mathrm{He}^{++}$), gas can absorb heat under compression and release it after maximum compression. This leads to an instability strip in the effective temperature-luminosity plane, which, in the case of Cepheids, is very narrow in temperature. The pulsation period depends on the mass and radius as

$$P \sim \frac{1}{(G\rho)^{\frac{1}{2}}} \sim M^{-\frac{1}{2}} R^{\frac{3}{2}} \quad . \tag{25}$$

The luminosity of the star (which is determined ultimately by nuclear reactions that depend on the density and temperature) is proportioned to a power of the mass, $L \sim M^k$. Since, however, we also have (for black body radiation) $L \sim R^2 T^4$, we obtain $P \sim L^{\frac{3k+2}{4}} T^{-3}$, or a period-luminosity-color relation. One of the "Key Projects" of HST has been to measure H_0 based on a Cepheid calibration of a number of secondary distance determination methods.

The results of this project have been described in a series of some 30 papers (see Freedman *et al.*, 2001, and references therein). The main goals of the project have been: (i) To discover Cepheids in a sample of relatively nearby galaxies (with distances $\lesssim 20$ Mpc) and to determine distances to these galaxies. (ii) To determine H_0 through several secondary distance indicators, to all of which Cepheid calibration is applied; and perhaps most importantly, (iii) To determine the uncertainties in all the methods by comparing the distances obtained from them to Cepheid-based distances. As a part of the project, the uncertainties in the Cepheid P-L relation itself (and its dependence on other factors such as metallicity) have been investigated.

The Key Project used Cepheid calibration to 31 galaxies, of which 18 have been observed and analyzed in the context of the project, and to which archival data, and data on the nearby galaxies M31, M33, IC 1613, NGC 300, and NGC 2403 have been added.

The Key Project used the following secondary methods based on Cepheid distances: Type Ia Supernovae, the Tully-Fisher Relation, the Fundamental Plane for elliptical galaxies, Surface Brightness Fluctuations and Type II Supernovae. Let me describe very briefly the physical basis for each one of these methods.

1. *Type Ia supernovae*

Type Ia supernovae at peak brightness are extremely bright, with $M_B \simeq M_V \simeq -19.3 + 5\log(H_0/60)$, and they show a relatively low dispersion, $\sigma(M_B) \sim 0.33$ (e.g., Branch, 1998). Furthermore, there exists a relatively tight correlation between their peak luminosity and light-curve shape (or rate of decline, with brighter supernovae declining more slowly; Hamuy *et al.*, 1996; Phillips, 1993; Riess *et al.*, 1996). The homogeneity may be related to the fact that Type Ia supernovae represent thermonuclear disruptions of mass-accreting white dwarfs, when the latter reach the Chandrasekhar limit (e.g., Livio, 2001). The luminosity-light-curve relation may be the result of the following (e.g., Arnett, 2001). The peak luminosity of a supernova Type Ia is proportional to the mass of ^{56}Ni that is produced. A higher mass of ^{56}Ni, however, also results in more heating and concomitantly a higher opacity (due mainly to UV lines). Consequently, a slower development of the light curve results.

2. The Tully-Fisher relation

For spiral galaxies, that are known to have flat rotation curves, there is an observationally-determined relationship (Tully and Fisher, 1977) between the total luminosity and the maximum rotational velocity (both corrected for inclination effects), of the form $L \sim V_{\rm max}^3$ (in the I band). The scatter around this relation is about ± 0.3 mag (e.g., Giovanelli *et al.*, 1997).

There is no precise physical understanding of the Tully-Fisher relation. Very broadly, a similar relation can be obtained from the following argument (e.g., Eisenstein and Loeb, 1996). Consider a galaxy (of mass M) collapsing from a spherical cloud. At epoch $t_{\rm coll}$, the turnaround radius R_t is

$$R_t \sim M^{\frac{1}{3}} t_{\rm coll}^{\frac{2}{3}} \quad . \tag{26}$$

After virialization, the energy is

$$E \sim M\sigma^2 \sim GM^2/R_t \quad , \tag{27}$$

where σ is the velocity dispersion. Combining the above and assuming that the galaxy forms from a single collapse gives

$$\sigma \sim (M/t_{\rm coll})^{\frac{1}{3}} \quad . \tag{28}$$

Therefore, if all galaxies collapse at the same epoch and the mass-to-light ratios, M/L, do not vary significantly, we obtain $L \sim \sigma^3$.

3. The fundamental plane for elliptical galaxies

Large spectrophotometric surveys conducted during the mid-1980s (e.g., Djorgovski and Davis, 1987; Dressler *et al.*, 1987) revealed that, for elliptical galaxies, a tight correlation exists between the effective radius, R_e, the effective surface brightness, SB_e, and the central velocity dispersion σ, of the form (the "fundamental plane")

$$\log R_e = \alpha \log \sigma + \beta SB_e + \delta \quad . \tag{29}$$

Here R_e is in kpc, σ in km s^{-1}, and SB_e in mag arcsec^{-2}. The value of δ depends on H_0, since the calculation of the effective radius in kpc uses the Hubble constant.

The physical origin of the fundamental plane relation can be understood on the basis of the following simple considerations (e.g., Treu *et al.*, 2001). We can define an effective (virial) mass by

$$M = \sigma^2 R_e/G \quad . \tag{30}$$

Let us also assume that the mass-to-light ratio satisfies

$$\frac{M}{L} \sim M^{\delta} \quad . \tag{31}$$

We then obtain

$$L \sim \sigma^{2(1-\delta)} R_e^{(1-\delta)} \quad , \tag{32}$$

which reduces to the fundamental plane relation for reasonable values of δ ($\delta \sim 0.25$).

4. Surface brightness fluctuations

The method of surface brightness fluctuations was developed by Tonry and Schneider (1988) and Tonry *et al.* (1997, 2000). The method basically makes use of the obvious fact that the ability to resolve stars within galaxies is distance dependent. More specifically, for every region of a galaxy one can measure the average flux per pixel, g, and the pixel-to-pixel rms, σ. Since the flux obtained in a pixel is received from N stars of average flux $\bar{f}$, we have $g = N\bar{f}$ and $\sigma = \sqrt{N}\bar{f}$. A galaxy which is twice as distant appears twice as smooth as the closer galaxy. Consequently, the average stellar flux is given by $\bar{f} = \sigma^2/g$ (and $\bar{f}$ scales as the inverse of the square of the distance).

5. The Expanding Photosphere Method of Type II supernovae

Type II supernovae result from the collapse of stars more massive than about 8 $M_{\odot}$. Generally, Type II supernovae are fainter than Type Ia supernovae, and they also exhibit a considerably larger range in their luminosities, making them poorer standard candles. Nevertheless, Type II supernovae have been used as distance indicators, through an application of the Expanding Photosphere Method (e.g., Kirshner and Kwan, 1974; Schmidt *et al.*, 1994).

The basic idea is simple. The angular size of the photosphere is given by (for $z \ll 1$)

$$\Theta = \frac{R}{D} = \left(\frac{f_\lambda}{\zeta_\lambda^2 \pi B_\lambda(T)}\right)^{\frac{1}{2}} \quad , \tag{33}$$

where T is the color temperature, f_λ is the flux density, $B_\lambda(T)$ is the Planck function, and ζ_λ represents the dilution effects of scattering atmospheres (derived from model atmospheres).

The photospheric radius is given by

$$R = v(t - t_o) + R_o \quad , \tag{34}$$

where v is the expansion velocity (measured from the absorption minima of optically thin lines), and the initial radius, R_o, can be neglected at all but the earliest times. Combining the above yields

$$t = D\left(\frac{\Theta}{v}\right) + t_o \quad , \tag{35}$$

making it possible to determine both the distance and the time of the explosion from a few measurements of t, v, Θ.

C. The results

Table 2, adapted from Freedman *et al.* (2001), lists the values of H_0 obtained from the different methods, based on the Cepheid distances of the Key Project. Of the five calibrated methods, it is clear that the Fundamental Plane is somewhat of an outlier. Combining these results, the H_0 Key Project obtained the value $H_0 = 72 \pm 3 \pm 7$ km s^{-1} Mpc^{-1}, where the first quoted error is random and the second is systematic. Using three different weighting schemes, all the results were found to be consistent with $H_0 = 72 \pm 8$ km s^{-1} Mpc^{-1}.

In an independent work, Allan Sandage, Gustav Tammann, Abijhit Saha, and collaborators used a Cepheid calibration of the peak brightness of Type Ia Supernovae (using HST), and thereby determined H_0 directly from the Hubble diagram of the latter. This effort resulted to date in nine supernovae Ia with "normal" spectra (i.e., the spectra characterizing about 60% of all Type Ia supernovae) to which Cepheid distances are known (Table 3, adapted from Saha *et al.*, 2001). After corrections based on the decline rate and colors, the weighted average of the absolute magnitudes of the calibrators was fitted to a sample of 35 more distant, "normal" Type Ia supernovae. After further correcting for some systematic errors, Saha *et al.* (2001) obtained $H_0 = 58.7 \pm 6.3$(internal) km s^{-1} Mpc^{-1} (for cosmological parameters $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$).

Unlike the uncertainty by a factor of two that has plagued this field for decades, therefore, the observations with HST have reduced the uncertainty in the value of the Hubble constant to about 15%.

The values obtained by both the Key Project and the Sandage-Tammann-Saha team are also consistent with other recent measurements based on combining the Sunyaev-Zeldovich effect with x-ray flux measurements of clusters, and on time delays in gravitational lensing. The systematics in these methods are still estimated to be at the 20–25% level, and the value obtained for the Hubble constant is $H_0 \sim 60$ km s^{-1} Mpc^{-1} (e.g., Reese *et al.*, 2002, 2000; Schecter, 2000).

Analysis of Sunyaev-Zeldovich effect (Sunyaev and Zeldovich, 1970) and x-ray data provides a direct method for determining distances to galaxy clusters. Clusters of galaxies are known to contain hot intracluster gas (at $kT \sim 10$ keV), trapped in the clusters' potential wells. Photons from the cosmic microwave background that pass through a cluster have a finite probability (optical depth $\tau \sim 0.01$) to interact with energetic electrons in the intracluster gas. The inverse Compton scattering that ensues boosts the energy of the microwave background photon, generating a small distortion (a decrement in frequencies $\lesssim 218$ GHz and an increment above this value) in the spectrum of the microwave background. The Sunyaev-Zeldovich effect is proportional to the integral of the pressure along the line of sight ($\int n_e T_e dl$). Since the x-ray emission from the intracluster medium is proportional to a different power of the density, $\int n_e^2 \Lambda dl$ (where Λ is the cooling function), a combination of the two measurements (given some assumptions about the cluster geometry) can be used to determine the distance to the cluster, independently of the distance scale ladder that is based on standard candles.

Finally, the Wilkinson Microwave Anisotropy Probe (WMAP) found a Hubble constant of $H_0 = 72 \pm 5$ km s^{-1} Mpc^{-1}. When the WMAP data were combined with the Key Project results, finer-scale cosmic microwave background experiments, and other large-scale structure and Lyman α forest data, the best-fit value of the Hubble constant was $H_0 = 71^{+4}_{-3}$ km s^{-1} Mpc^{-1} (Spergel *et al.*, 2003).

Type Ia supernovae have played a key role not only in the determination of the *age* of the universe (through H_0), but also in the determination of the universe's *geometry*, and the dynamics of the cosmic expansion. This came about through a combination of ground-based and HST observations in which both careful planning and serendipity played a part.

VIII. THE ACCELERATING UNIVERSE

In 1998, two teams of astronomers, working independently, presented evidence that the expansion of the universe is *accelerating* (Perlmutter *et al.*, 1999; Riess *et al.*, 1998). This evidence was based primarily on the faintness (by about ~ 0.25 mag) of distant Type Ia supernovae (SNe Ia), compared to their expected brightness in a universe decelerating under its own gravity.

The first suggestions that high-redshift Type Ia supernovae could be used to determine the rate of cosmic deceleration came in the late 1970s (Colgate, 1979; Tammann, 1979; Wagoner, 1977), but the actual discovery of an even moderate redshift ($z = 0.31$) SN Ia (SN 1988U) had to wait for a decade (Nørgaard-Nielsen *et al.*, 1989). The samples of high-redshift supernovae started to become sufficiently large to place constraints on cosmological parameters through the efforts of the Supernova Cosmology Project, led by Saul Perlmutter, and the High-z Supernova Search Team, led by Brian Schmidt. By 1998 the two teams gathered sufficient data to be able to show that the universe is characterized by a matter density paremater satisfying $\Omega_M < 1$ (i.e., that the universe is not closed by matter; Garnavich *et al.*, 1998; Perlmutter *et al.*, 1998). However, the real shocker was still to come.

A. The observations

The two supernova search teams use a similar method to find their supernovae. They take two deep images separated by about a month, subtract the first-epoch image from the second-epoch one, and search for sources above a certain threshold in the difference image. Once a candidate SN is identified, the SN type is determined by its spectrum (if that can be taken; in a few cases one has to rely on the host galaxy type—only SNe Ia were found so far in ellipticals). Supernovae at relatively high redshift are then monitored photometrically to construct their light curves. An image of the host galaxy is obtained at a later time (after a year or more), and subtracted to obtain an accurate measurement of the SN brightness.

The Hubble Space Telescope has proved to be crucial especially for the highest-redshift supernovae. There, the ability to resolve and pinpoint the SN location on the host (including in cases in which the SN was found close to the galactic nucleus) was essential for a correct determination of the SN magnitudes. With samples of a few dozen SNe Ia in hand, the two teams compared their measured distances (derived from the luminosity-distance relation, $F = L/4\pi D_L^2$; where D_L is the distance and L, F are the intrinsic luminosity and observed flux, respectively) with the distances expected for the observed redshifts, for different cosmological models (e.g., Carroll, Press, and Turner, 1992). The latter is given by

$$\begin{aligned} D_L = & \; cH_0^{-1}(1+z)|1-\Omega_M-\Omega_\Lambda|^{-1/2}\mathrm{sinn}\Big\{|1-\Omega_M-\Omega_\Lambda|^{1/2}\times \\ & \textstyle\int_0^z dz[(1+z)^2(1+\Omega_M z)-z(2+z)\Omega_\Lambda]^{-1/2}\Big\} \end{aligned} \tag{36}$$

where sinn denotes sinh for $\Omega_M + \Omega_\Lambda \leq 1$ and sin for $\Omega_M + \Omega_\Lambda > 1$. The results for the likelihood of the cosmological parameters Ω_M and Ω_Λ are shown in Fig. (28). As can be seen from the figure, the results favor values of $\Omega_M \simeq 0.3$, $\Omega_\Lambda \simeq 0.7$ and a *negative* deceleration parameter (a is the scale factor, Ω_0 is the sum of today's energy densities, and $w \equiv P_\Lambda/\rho_\Lambda$ characterizes the dark energy "equation of state"; see section VIII D)

$$q_0 \equiv \frac{-\ddot{a}(t_0)a(t_0)}{\dot{a}^2(t_0)} = \frac{\Omega_0}{2} + \frac{3}{2}w\Omega_\Lambda < 0 \quad , \tag{37}$$

corresponding to an accelerating universe.

B. Alternative interpretations

Clearly, the interpretation of an accelerating cosmic expansion cannot go unchallenged (see also Riess, 2000; Turner, 2000). The two most serious challenges to acceleration are, in principle, evolution/progenitors and dust extinction.

The possibility that SNe Ia undergo some type of cosmic evolution cannot be easily dismissed. After all, almost everything else (e.g., the host galaxies themselves, the metallicity) evolves. The inference of acceleration, on the other hand, assumes that the local and high-redshift supernovae are drawn from the same statistical sample. I should explain that by evolution, I do not mean that somehow the processes governing supernova explosions are necessarily changing with redshift, but rather that when observing the high-redshift universe we are sampling younger galaxies, and consequently the population from which the supernovae are drawn is typically younger.

Since SNe Ia represent thermonuclear disruptions of mass-accreting white dwarfs (Hillebrandt and Niemeyer, 2000; Livio, 2001), changes in the C/O ratio in the white dwarfs (as a result of lower metallicities in the distant past), for example, could produce some inhomogeneity in SNe Ia light curves (Umeda *et al.*, 1999). However, observations tend to show that even if such an effect exists, it is insignificant. In fact, existing observations of local galaxies (e.g., Cappellaro *et al.*, 1997; Riess *et al.*, 1999; van den Bergh, 1994) already span a wider range of metallicities and host properties than that expected on the average between the sample of galaxies at $z \sim 0$ and $z \sim 0.5$.

To my knowledge, there is only one known evolutionary effect that is physically meaningful, that can mimic accelerated expansion. This is the effect of the metallicity on the density at the point of central carbon ignition (the trigger of SNe Ia). Generally, a lower metallicity (as expected at high-z) will result in a lower central density (Nomoto *et al.*, 1997). This is because a lower metallicity results in a lower abundance of ^{21}Ne, that is responsible for much of the neutrino cooling (via the so-called local URCA shell process, involving the ^{21}Ne-^{21}F pair). A lower metallicity therefore reduces the cooling and leads to an earlier ignition. Due to the lower white dwarf binding energy, the light curve exhibits a more rapid development, and a lower implied maximum brightness. However, as I noted above, the local (low-z) sample of galaxies does not appear to exhibit any significant metallicity-dependent effect.

Dust extinction is another natural candidate for making the distant supernovae appear dimmer. In fact, the observed decrease in the brightness (~ 0.25 mag) of the distant sample would require only a $\sim 25\%$ increase in the extinction. Both teams used colors to correct for dust extinction (like in sunsets, ordinary dust also reddens the light). Furthermore, observations from the optical to the infrared of SN 1999Q (at $z = 0.46$) showed that a large extinction to this supernova is highly improbable (Riess *et al.*, 2000), making the dust hypothesis untenable.

Another potential uncertainty is associated with the progenitors of SNe Ia. The progenitor systems of SNe Ia are not known without doubt (Livio, 2001). In particular, the two leading progenitor-system candidates are double white dwarf systems (that merge to produce the explosion; Iben and Tutukov, 1984; Webbink, 1984) and systems in which the white dwarf accretes from a normal companion (Nomoto, 1982; Whelan and Iben, 1973). The possibility therefore exists, in principle, that the local and distant samples are dominated by different progenitors, with one class producing somewhat dimmer supernovae. This possibility should certainly be considered, especially in view of the fact that Li *et al.* (2001) find a relatively high ($\sim 40\%$) fraction of peculiar SNe Ia among the local sample.

Yungelson and Livio (2000) have shown that while it is possible, in principle, that one class of progenitors (e.g., double white dwarf systems) dominates the local sample and another (a white dwarf with a normal companion) the high-z one, such a transition is not likely, because of the following reason. If such a transition were to occur, one would expect that around the transition point (at $z \sim 1$), SNe Ia would be composed of an equal mix of the two classes. This is inconsistent with the observations, that show that the high-z sample is actually extremely homogeneous (Li *et al.*, 2001). Consequently, it is highly unlikely that the interpretation of an accelerating universe is an artifact of the existence of two classes of progenitors (see Livio, 2001, for a more complete discussion).

Other effects, such as those resulting from gravitational lensing, will require some further consideration once larger databases of high-redshift supernovae become available. Generally, it can be expected that most lines of sight to detected supernovae do not pass through large mass concentrations. Consequently, light paths are being bent *out* of the line of sight—resulting in deamplification. In rare cases however, the line of sight may cross significant matter concentrations resulting in strong amplification. The picture that emerges, therefore, is that of a brightness distribution in which the peak is shifted toward a lower value, but with a tail of bright objects (e.g., Holz, 1998). Effects of this type can be averaged out once deep data from many lines of sight become available, perhaps with the proposed wide-field imager, the Super Nova/Acceleration Probe (SNAP).

C. Supernova 1997ff—serendipity and careful planning

In the spirit of "extraordinary claims require extraordinary proof," the magnitude of the discovery of an accelerating universe requires an extremely careful elimination of potential contaminating astrophysical effects. A pervasive screen of "gray" dust, for example, that can dim the light while leaving little imprint on the spectral energy distribution, has been suggested as an alternative explanation for the unexpected faintness of high-z SNe Ia (Aguirre, 1999). While gray dust, e.g., quasispherical grains larger then 0.1 μm, could be made to reproduce some of the observations, some measurements do disfavor a 30% visual opacity due to gray dust at the $\sim 2.5\sigma$ level (Riess *et al.*, 2000). Nevertheless, both the possibility of gray dust and, in particular, the ever-existing challenge of luminosity evolution, did leave the case for an accelerating universe somewhat short of compelling by 2001. The strongest evidence so far against alternatives to an accelerating universe came from SN 1997ff.

In 1997, Ron Gilliland and Mark Phillips reobserved the Hubble Deep Field North with

HST, with the goal of detecting high-z supernovae (Gilliland, Nugent and Phillips, 1999). They discovered two supernovae: SN 1997fg, at a redshift of $z = 0.95$ was found in a late-type galaxy, while SN 1997ff was found to be hosted by an elliptical galaxy. The initial, photometrically determined redshift, placed the latter galaxy at $z \sim 0.95$–1.32 with considerable uncertainty. Serendipitously, the Guaranteed Time Observer (GTO) program of Thompson *et al.* (1999) included imaging the host of SN 1977ff with the near infrared camera, NICMOS, on board HST. One of these exposures, amazingly enough, was actually taken within hours of the discovery of the SN. Furthermore, six months after the GTO program, Dickinson *et al.* (2003, in preparation; GO 7817) reobserved the field with NICMOS. The combination of all the HST and ground-based observations (in the U, B, V, I, J, H, and K bands) allowed for an improved photometric redshift determination for the host of $z = 1.65 \pm 0.15$ (Budavári *et al.*, 2000), making SN 1997ff the farthest known supernova to date. A probability density function based on the partial supernovae light curve in three bands also gave $z \simeq 1.7$, as did a tentative determination of the redshift based on the spectrum of the host taken with the Keck telescopes (see Riess *et al.*, 2001, for a full description).

The importance of detecting a SN Ia at such a high redshift cannot be overemphasized. Even with the suggested value of the vacuum energy density of $\Omega_\Lambda \simeq 0.7$, at $z \gtrsim 1$ attractive gravity would still have dominated over the cosmic repulsion, resulting in a *decelerated* expansion. Consequently, SNe Ia in the redshift range $z \sim 1$–2 should appear *brighter* relative to SNe in a coasting universe. On the other hand, any effect that increases monotonically with redshift, as would be expected from dust extinction and simple evolutionary effects, would predict that the SNe Ia would be *dimmer* with increasing redshift. SN 1997ff was found to be brighter by ~ 1.1 mag (at the $> 99.99\%$ confidence level) than expected for an alternative source of dimming (e.g., dust or evolution) beyond $z \sim 0.5$. Figure 29 shows a redshift-distance relation in which the points are redshift-binned data from Perlmutter *et al.* (1999) and Riess *et al.* (1998), together with a family of curves for flat cosmological models. The transition from a decelerating to an accelerating phase occurs (if the "dark energy" is represented by a cosmological constant; see Section VIII D) at $z_{tr} = (2\Omega_\Lambda/\Omega_M)^{1/3} - 1$. As can be seen from the figure, the only cosmological model that is consistent with *all* the data (including SN 1997ff) is one in which $\Omega_M \simeq 0.35$, $\Omega_\Lambda \simeq 0.65$.

Measurements of the power spectrum of the cosmic microwave background (e.g., Abroe *et al.*, 2002; de Bernardis *et al.*, 2002; Hu *et al.*, 2001; Netterfield *et al.*, 2002) provided strong evidence for a geometrically flat ($\Omega_M + \Omega_\Lambda \simeq 1$) universe. When combined with several estimates of Ω_M from mass to light ratios, x-ray temperature of intracluster gas, numbers, and dynamics of clusters of galaxies, all giving $\Omega_M \simeq 0.2$–0.3 (e.g., Bahcall *et al.*, 2000; Bahcall, Fan, and Cen, 1997; Carlberg *et al.*, 1996; Strauss and Willick, 1995), the inescapable conclusion was that there is a dark energy component of $\Omega_\Lambda \simeq 0.7$—consistent with the value found from high-redshift supernovae. Most recently, the best fit cosmological model has been determined by the Wilkinson Microwave Anisotropy Probe (WMAP). In combination with other large-scale structure work, the WMAP results imply $\Omega_{\rm tot} = 1.02 \pm 0.02$, $\Omega_\Lambda \simeq 0.73$ (Bennett *et al.*, 2003; Spergel *et al.*, 2003).

D. The nature of "dark energy"

It is beyond the scope of the present article to discuss dark energy in detail and the reader is referred to the excellent review by Peebles and Ratra (2002). However, since this is arguably the most dramatic discovery involving HST I would like to make a few points.

In general relativity, the stress-energy tensor of the vacuum, $T^\Lambda_{\mu\nu}$, can be written as

$$T^\Lambda_{\mu\nu} = \rho_\Lambda g_{\mu\nu} \quad , \tag{38}$$

where ρ_Λ is constant (proportional to Einstein's cosmological constant) and $g_{\mu\nu}$ is the metric. In a Minkowski flat spacetime this can be written simply as the "equation of state"

$$P_\Lambda = -\rho_\Lambda \quad . \tag{39}$$

More generally, if the equation of state is written as (for example, Canuto *et al.*, 1977; Ratra and Peebles, 1988; Sahni and Starobinsky, 2000; Turner and White, 1997),

$$P_\Lambda = w\rho_\Lambda \quad , \tag{40}$$

then the dark energy density behaves (for a constant w) like $\rho_\Lambda \sim a^{-3(1+w)}$ (where a is the scale factor). Since the scale factor satisfies the equation (in units in which $c = 1$)

$$\frac{\ddot{a}}{a} = \frac{4\pi G}{3}(\rho + 3P) \quad , \tag{41}$$

the dark energy contribution will result in an *accelerating* universe if $w < -\frac{1}{3}$.

It is interesting to note that from a purely fluid dynamical point of view, any equation of state of the type $P = w\rho$ is in fact *dynamically unstable* for any negative value of w except for $w = -1$ (corresponding to a cosmological constant). This can be easily seen from perturbations on the equations for momentum and energy conservation ($<\rho>$, $<P>$ denote mean values)

$$\delta\dot{\rho} + (<\rho> + <P>)\nabla \cdot \vec{v} = 0 \tag{42}$$

$$(<\rho> + <P>)\dot{\vec{v}} + \left(\frac{dP}{d\rho}\right)\nabla\delta\rho = 0 \quad . \tag{43}$$

Defining the speed of sound, $c_s^2 = \dfrac{dP}{d\rho}$, these two equations can be combined to form

$$\delta\ddot{\rho} = c_s^2\nabla^2\delta\rho \quad , \tag{44}$$

which is clearly unstable for negative c_s^2 (except for $w = -1$, when the combination $<\rho> + <P>$ vanishes, as expected from the fact that the vacuum is the same for all inertial observers).

Thus, formally speaking, from a fluid-dynamical point of view only a cosmological constant yields a stable solution.

One can certainly take the dark energy to be associated with a uniform scalar field ϕ (a "quintessence" field; see below), that rolls down a potential $V(\phi)$ at a rate determined by

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \quad , \tag{45}$$

where $V'(\phi) = \dfrac{dV}{d\phi}$. In this case,

$$\begin{aligned} \rho_\phi &= \frac{1}{2}\dot{\phi}^2 + V(\phi) \\ P_\phi &= \frac{1}{2}\dot{\phi}^2 - V(\phi) \quad . \end{aligned} \tag{46}$$

For a sufficiently slowly varying dynamical component, such that the kinetic energy is much smaller than the potential, $\dot{\phi}^2 \ll V(\phi)$, one even obtains a field energy that mimics the cosmological constant ($P_\phi \simeq -\rho_\phi$). Generally, quintessence solutions allow for different equations of state, and even values of w that are time variable. Nevertheless, I find the stability property expressed in eq. (44) sufficiently intriguing to warrant a deeper examination of the *cosmological constant* possibility, before that is abandoned in favor of other dynamical solutions. This point of view received some further support from the recent determination by WMAP that $w < -0.78$ (Bennett *et al.*, 2003).

There are two main problems with the implied energy density ρ_V of the dark energy (e.g., Weinberg, 2001). (i) Why is its value not $\sim$ 120 orders of magnitude larger (as expected from fluctuations in the gravitational field up to the Planck scale)? (ii) Why now? (Namely, why $\Omega_\Lambda \sim \Omega_M$ now, even though Ω_Λ may be associated with a cosmological constant, while Ω_M declined continuously from the initial singularity to its present value).

An interesting curiosity to note is that even though taking graviton energies up to Planck scale, M_P, misses the value of the dark energy density by $\sim 10^{120}$, and taking them up to the supersymmetry scale, $M_{\rm SUSY}$, misses by $\sim 10^{55}$, a scale of $M_V \sim (M_{\rm SUSY}/M_P)M_{\rm SUSY}$ actually does give the right order of magnitude! While I am not aware at present of any theory that produces this scale in a natural way (although see, e.g., Arkani-Hamed *et al.*, 2000), I find this coincidence worth following up.

Much of the efforts to resolve the above two problems revolved not around a cosmological constant, but rather around the behavior of quintessence fields. In particular, attention has concentrated on "tracker" solutions, in which the final value of the quintessence energy density is independent of fine-tuning of the initial conditions (e.g., Albrecht and Skordis, 2000; Zlatev, Wang, and Steinhardt, 1998). For example if one takes a potential of the form

$$V(\phi) = \phi^{-\alpha} M^{4+\alpha} \quad , \tag{47}$$

where $\alpha > 0$ and M is an adjustable constant, and the field is initially much smaller than the Planck mass, then for $\rho_M \gg V(\phi)$, $\dot{\phi}^2$, we find that ρ_M decreases initially faster ($\sim t^{-2}$) than $\phi(t)(\sim t^{-2\alpha/(2+\alpha)})$.

Eventually, however, a transition to a ρ_ϕ-dominated universe occurs (and ρ_ϕ decreases as $t^{-2/(4+\alpha)}$). In other words, the quintessence answer to the question: why is the dark energy density so small? is simply: because the universe is very old. Nevertheless, simple potentials of the form (47) do not offer a clear solution to the "why now?" problem. In fact, to have $\rho_\phi \sim \rho_M$ (and of order of the critical density) at the present time requires fine-tuning the parameters so that (for example, Weinberg, 2001)

$$M^{4+\alpha} \simeq (8\pi G)^{-1-\alpha/2} H_0^2 \quad , \tag{48}$$

with no simple explanation as to why this equality should hold.

In order to overcome this requirement for fine-tuning, some versions of the quintessence models choose potentials in which the universe has periodically been accelerating in the past (the dark energy has periodically dominated the energy density; e.g., Dodelson, Koplinghat & Stewart 2000). Dodelson *et al.* (2000) have shown, for example, that with a potential of the form

$$V(\phi) = V_o \exp\left(-\lambda\phi\sqrt{8\pi G}\right)\left[1 + A\sin\left(\nu\phi\sqrt{8\pi G}\right)\right] \quad , \tag{49}$$

one can obtain solutions in which the dark energy density tracks the ambient energy density of the universe and satisfies observational constraints.

A different approach to the problems associated with the dark energy density has been through anthropic considerations (e.g., Kallosh and Linde, 2002; Vilenkin, 1995; Weinberg, 2001). The key assumption in this class of models is that *some* constants of nature, in particular the cosmological constant, and possibly the density contrast at the time of recombination, $\sigma_{\rm rec}$, are in fact random variables, whose range of values and a priori probabilities are nevertheless determined by the laws of physics. In this picture, some values of the constants which are allowed in principle, may be incompatible with the very existence of observers. Assuming a *principle of mediocrity*—that we should expect to find ourselves in a universe *typical* of those that allow the emergence of intelligent life—Garriga, Livio, and Vilenkin (2000) were able to show that the "why so small?" and "why now?" questions find a natural explanation. While I personally believe that anthropic considerations should only be used as a last resort, there is no denial of the fact that most versions of "eternal inflation," the notion that once inflation starts it never stops, unavoidably produce an ensemble of "pocket" universes (Guth, 1981; Linde, 1986; Steinhardt, 1983; Vilenkin, 1983). Once such an infinite ensemble is believed to exist, the problem of defining probabilities on it, and the concept of "mediocrity" become real (e.g., Linde, Linde, and Mezhlumian, 1995; Vilenkin, 1998).

In spite of the apparent "success" of the anthropic argumentation in solving the two main problems associated with dark energy, the search for a fundamental explanation is not, and should not, be abandoned. An interesting, if speculative, new direction is provided by alternative theories of gravity. For example, models in which ordinary particles are localized on a three-dimensional surface (3-brane) embedded in infinite volume extradimensions to which gravity can spread, have been developed (Deffayet, Dvali, and Gabadadze, 2002). In a particular version of these models, the Friedmann equation (eq. 24) is replaced by

$$H^2 + \frac{k}{a^2} = \left(\sqrt{\frac{\rho}{3M_p^2} + \frac{1}{4r_c^2}} + \epsilon\frac{1}{2r_c^2}\right)^2 \quad , \tag{50}$$

where $\epsilon = \pm 1$, and r_c represents a crossover scale. At distances shorter than r_c (which can be of astronomical size, e.g., $r_c \sim cH_0^{-1}$), observers on the brane discover the familiar Newtonian gravity. At large cosmological distances, however, the force-law of gravity becomes five-dimensional (as gravity spreads into the extra dimensions) and weaker. The dynamics are governed by whether ρ/M_p^2 is larger or smaller than $1/r_c^2$. While highly speculative at this stage, some aspects of these alternative theories of gravity can be experimentally tested (e.g., by lunar ranging experiments; Dvali, Gruzinov, and Zaldarriaga, 2002). Also, the relative lack of power on large scales (in particular, absence of any correlated signal on scales larger than 60 degrees) found by the Wilkinson Microwave Anisotropy Probe (Spergel *et al.*, 2003), may indicate a breakdown of canonical gravity on large scales (or, more speculatively, a finite universe).

Luckily, observations that are already being performed with HST, and observations with the proposed Super Nova/Acceleration Probe (SNAP) will provide some crucial information on the nature of dark energy. In particular, if the equation of state parameter w is constant, then a transition from a decelerating phase to an accelerating phase occurs at

$$z_{\rm tr} = [(1+3w)(\Omega_M - 1)/\Omega_M]^{-1/3w} - 1 \quad . \tag{51}$$

For example, for $\Omega_M = 0.27$ (Bennett *et al.*, 2003) and a cosmological constant ($w = -1$), this gives $z_{\rm tr} \simeq 0.76$. Equation (51) shows that if Ω_M can be determined independently to within a few percent (a goal that can perhaps be achieved by a conbination of cosmic microwave background and large scale structure measurements), then observations of a large sample of Type Ia supernovae in the redshift range 0.8–2 can place very meaningful constraints on the dark energy equation of state. The situation is more ambiguous if w is time-dependent, but even then, a sufficiently large sample of SNe Ia will give useful information. At the time of this writing, a search for SNe Ia conducted in the GOODS fields has already yielded 10 SNe Ia, with six of them in the redshift range $z \sim 1$–2 (Principal Investigator: Adam Riess). This, together with a few current ground-based searches at lower redshifts, is an excellent first step. The survey proposed by SNAP, of 15 square degrees, could yield light curves and spectra for over 2000 SNe Ia at redshifts up to $z \sim 1.7$. With this sample in hand, one could reach the elimination or at least control of the sources of systematic uncertainties needed to discriminate between different dark-energy models.

IX. EPILOGUE

The Hubble Space Telescope, together with many other space-based and ground-based observatories, have given us a much clearer picture than we ever had before of the universe we live in. We now know not only that other planetary systems are quite common, but that the cosmic expansion is accelerating. We know the age of the universe, its geometry, and its composition. Observations of stars start to reach the level of detail previously characterizing only solar physics. A general picture of the formation of structure in the universe, and many clues for how galaxies form and evolve, start to emerge. We know that supermassive black holes reside at the centers of most galaxies. This is, however, far from marking the end of HST's contributions. The recent installation of the Advanced Camera for Surveys, and the two future (currently planned for 2004) instruments, the Cosmic Origins Spectrograph (COS) and the Wide Field Camera 3 (equipped with an infrared channel), promise that the coming seven years will be at least as exciting as the first thirteen were.

Acknowledgments

This review has benefited from discussions with more colleagues than I can practically thank here. I would like, however, to extend special thanks to John Bally, Stefano Casertano, Mark Dickinson, Gia Dvali, Harry Ferguson, Andy Fruchter, Mauro Giavalisco, Ron Gilliland, Andy Ingersoll, Ken Lanzetta, Nino Panagia, Saul Perelmutter, Adam Riess, Kailash Sahu, Rachel Somerville, Massimo Stiavelli, Roeland van der Marel, and Alex Vilinkin for very helpful discussions.

References

Abraham, R. G., R. S. Ellis, A. C. Fabian, N. R. Tanvir, and K. Glazebrook, 1999a, Mon. Not. R. Astron. Soc. **303**, 641.

Abraham, R. G., M. R. Merrifield, R. S. Ellis, N. R. Tanvir, and J. Brinchman, 1999b, Mon. Not. R. Astron. Soc. **308**, 569.

Abraham, R. G., N. R. Tanvir, B. X. Santiago, R. S. Ellis, K. Glazebrook, and S. van den Bergh, 1996, Mon. Not. R. Astron. Soc. **279**, L47.

Abroe, M. E., A. Balbi, J. Borrill, *et al.*, 2002, Mon. Not. R. Astron. Soc. **334**, 11.

Adams, F. C., D. S. Graff, and D. D. Richstone, 2001, Astrophys. J. Lett. **551**, L31.

Aguirre, A., 1999, Astrophys. J. **525**, 583.

Albrecht, A. and C. Skordis, 2002, Phys. Rev. Lett. **84**, 2076.

Andersen, M. I., J. Hjorth, and J. Gorosabel, 2002, Astron. Astrophys. (in press).

Arkani-Hamed, N., L. J. Hall, C. Colda, and H. Murayama, 2000, Phys. Rev. Lett. **85**, 4434.

Armitage, P. J., 2000, Astron. Astrophys. **362**, 968.

Armitage, P. J., M. Livio, S. H. Lubow, and J. E. Pringle, 2002, Mon. Not. R. Astron. Soc. **334**, 248.

Arnett, D., 2001, in *Supernovae and Gamma-Ray Bursts*, edited by M. Livio, N. Panagia, and K. Sahu (Cambridge University, Cambridge, England), p. 250.

Baade, W., 1954, IAU Trans. **8**, 397.

Bahcall, J. N., S. Kirhakos, D. H. Saxe, and D. P. Schneider, 1997, Astrophys. J. **479**, 642.

Bahcall, J. N., S. Kirhakos, and D. P. Schneider, 1994, Astrophys. J. Lett. **435**, L11.

Bahcall, N. A., R. Cen, R. Davé, J. P. Ostriker, and Q. Yu, 2000, Astrophys. J. **541**, 1.

Bahcall, N. A., X. Fan, and R. Cen, 1997, Astrophys. J. Lett. **485**, L53.

Balick, B., 1987, Astron. Astrophys. **94**, 671.

Balick, B. and A. Frank, 2002, Annu. Rev. Astron. Astrophys. **40**, 439.

Bally, J., 2003, in *A Decade of HST Science*, edited by Mario Livio, Keith Noll, and Massimo Stiavelli, (Cambridge University, Cambridge, England), p. 44.

Bally, J., C. R. O'Dell, and M. J. McCaughrean, 2000, Astron. J. **119**, 2919.

Baumgardt, H., P. Hut, J. Makino, S. McMillan, and S. Portegies Zwart, 2002, eprint astro-ph/0210133.

Becker, R. H., X. Fan, R. L. White, *et al.*, 2001, Astron. J., **122**, 2850.

Begelman, M. and Z. Y. Li, 1992, Astrophys. J. **397**, 187.

Behr, A., 1951, Astr. Nachr. **279**, 97.

Benitez, N., T. Broadhurst, R. Bouwens, J. Silk, and P. Rosati, 1999, Astrophys. J. **515**, 65.

Bennett, C. L., M. Halpern, G. Hinshaw, *et al.*, 2003, Astrophys. J. (submitted).

Bershady, M. A., A. Jangren, and C. J. Conselice, 2000, Astron. J., **119**, 2645.

Bertoldi, F. and C. F. McKee, 1990, Astrophys. J. **354**, 529.

Binney, J. and G. A. Mamon, 1982, Mon. Not. R. Astron. Soc. **200**, 361.

Binney, J. and S. Tremaine, 1987, *Galactic Dynamics*, (Princeton University, Princeton), p. 428.

Biretta, J. A., W. B. Sparks, and F. Macchetto, 1999, Astrophys. J. **520**, 621.

Bjorkman, J. E. and J. P. Cassinelli, 1993, Astrophys. J. **409**, 429.

Blandford, R. and R. Znajek, 1977, Mon. Not. R. Astron. Soc. **179**, 433.

Blandford, R. D. and D. G. Payne, 1982, Mon. Not. R. Astron. Soc. **199**, 883.

Bloom, J. S., S. R. Kulkarni, P. A. Price, *et al.*, 2002, Astrophys. J. Lett. **572**, L45.

Bond, H. E., R. Ciardullo, L. K. Fullton, and K. G. Schaefer, 1996 (unpublished).

Boroson, T. A. J. B. Oke, and R. F. Green, 1982, Astrophys. J. **263**, 32.

Boss, A. P., 2000, Astrophys. J. Lett. **536**, L101.

Bouwens, R. J., L. Cayon, and J. Silk, 1997, Astrophys. J. Lett. **489**, L21.

Branch, D., 1998, Annu. Rev. Astron. Astrophys. **36**, 17.

Brinchmann, J., R. Abraham, and D. Schade, *et al.*, 1998, Astrophys. J. **499**, 112.

Brown, T. M., D. Charbonneau, R. L. Gilliland, R. W. Noyes, and A. Burrows, 2001, Astrophys. J. **552**, 699.

Budavári, T., A. S. Szalay, A. J. Connolly, I. Casbai, and M. Dickinson, 2000, Astron. J., **120**, 1588.

Burrows, A., M. S. Marley, and C. M. Sharp, 2000, Astrophys. J. **531**, 438.

Burrows, C. J., J. Krist, J. J. Hester, *et al.*, 1995, Astrophys. J. **452**, 680.

Burrows, C. J., K. R. Stapelfeldt, A. M. Watson, *et al.*, 1996, Astrophys. J. **473**, 437.

Butler, R. P., S. S. Vogt, G. W. Marcy, D. A. Fischer, G. W. Henry, and K. Apps, 2000, Astrophys. J. **545**, 504.

Calzetti, D. and T. M. Heckman, 1999, Astrophys. J. **519**, 27.

Calzetti, D., A. L. Kinney, and T. Storchi-Bergmann, 1994, Astrophys. J. **429**, 582.

Camenzind, M., 1990, in *Reviews of Modern Astronomy 3*, edited by G. Klare, (Springer, Berlin), p. 259.

Canuto, V., P. J. Adams, S.-H. Hsieh, and E. Tsiang, 1977, Phys. Rev. D **16**, 1643.

Cappellaro, E., M. Turatto, D. Yu. Tsvelkov, O. S. Bartunov, C. Pollas, R. Evans, and M. Hamuy, 1997, Astron. Astrophys. **322**, 431.

Carlberg, R. G., H. K. C. Yee, E. Ellingson, R. Abraham, P. Grabel, S. Morris, and C. J. Pritchet, 1996, Astrophys. J. **462**, 32.

Carlson, R., P. Weissman, M. Segura, *et al.*, 1995, Geophys. Res. Lett. **22**, L1557.

Carroll, S. M., W. H. Press, and E. L. Turner, 1992, Annu. Rev. Astron. Astrophys., **30**, 499.

Charbonneau, D., T. M. Brown, D. W. Latham, and M. Mayor, 2000, Astrophys. J. Lett. **529**, L41.

Charbonneau, D., T. M. Brown, R. W. Noyes, and R. L. Gilliland, 2002, Astrophys. J. **568**, 377.

Chevalier, R. A. and D. Luo, 1994, Astrophys. J. **435**, 815.

Churchwell, E., M. Feli, D. O. S. Wood, and M. Massi, 1987, Astrophys. J. **321**, 516.

Chyba, C., P. Thomas, and K. Zahnle, 1993, Nature **361**, 40.

Colbert, E. J. M. and R. F. Mushotzky, 1999, Astrophys. J. **519**, 89.

Cole, S., C. G. Lacey, C. M. Baugh, and C. S. Frenk, 2000, Mon. Not. R. Astron. Soc. **319**, 168.

Colgate, S., 1979, Astrophys. J. **232**, 404.

Conselice, C. J., 1997, Publ. Astron. Soc. Pac. **109**, 1251.

Conselice, C. J., 2001, in *Deep Fields*, edited by S. Cristiani, A. Renzini, and R. E. Williams, (Springer, Berlin), p. 91.

Conselice, C. J., M. A. Bershady, and A. Jangern, 2000, Astrophys. J., **529**, 886.

Corradi, R. L. M., M. Livio, B. Balick, U. Munali, and H. E. Schwarz, 2001, Astrophys. J. **553**, 211.

Costa, E., F. Frontera, and J. Heise, 1997, Nature **387**, 783.

Crotts, A. and S. R. Heathcote, 1991, Nature **350**, 683.

D'Amico, N., A. Possenti, L. Fici, R. N. Manchester, A. G. Lyne, F. Camilo, and J. Sarkissian, 2002, Astrophys. J. Lett. **570**, L89.

Damineli, A., P. S. Conti, and D. F. Lopes, 1997, New Astron. **2**, 107.

Damineli, A., A. Kaufer, B. Wolf, O. Stahl, D. F. Lopes, and F. X. de Araujo, 2000, Astrophys. J. Lett. **528**, L101

Davidson, K. and R. M. Humphreys, 1997, Annu. Rev. Astron. Astrophys. **35**, 1.

Davies, M. B. and S. Sigurdson, 2001, Mon. Not. R. Astron. Soc. **324**, 612.

de Bernardis, P., P. A. R. Ade, J. J. Bock, *et al.*, 2002, Astrophys. J. **564**, 559.

Deffayet, C., G. Dvali, and G. Gabadadze, 2002, eprint astro-ph/0105068.

Dickinson, M., 1995, in *Galaxies in the Young Universe*, edited by H. Hippelein, H. J. Meisenheimer, and H.-J. Röser, (Springer, Berlin), p. 144.

Dickinson, M., 2000a, Phil. Trans. R. Soc. Lond. A **358**, 2001.

Dickinson, M., 2000b, in *Building Galaxies*, XIX Rencontre de Moriond, edited by F. Hammer, *et al.* (World Scientific, Singapore), p. 257.

Dickinson, M., C. Papovich, H. C. Ferguson, and T. Budavári, 2003, Astrophys. J. (in press).

Djorgovski, G. and M. Davis, 1987, Astrophys. J. **313**, 59.

Dobrovolskis, A. R., 1990, Icarus **88**, 24.

Dodelson, S., M. Koplinghat, and E. Stewart, 2000, Phys. Rev. Lett. **85**, 5276.

Dressler, A., D. Lynden-Bell, D. Burstein, R. L. Davies, S. M. Faber, R. Terlevich, and G. Wegner, 1987, Astrophys. J. **313**, 42.

Dressler, A., A. J. Oemler, W. B. Sparks, and R. A. Lucas, 1994, Astrophys. J. Lett. **435**, L23.

Dressler, A. and D. O. Richstone, 1988, Aastrophys. J. **324**, 701.

Dressler, A. and D. O. Richstone, 1990, Astrophys. J. **348**, 120.

Dull, J. D., H. N. Cohn, P. M. Lugger, B. W. Murphy, P. O. Seitzer, P. J. Callanan, R. G. M. Rutten, and P. A. Charles, 2002, eprint astro-ph/0210588.

Dvali, G., A. Gruzinov, and M. Zaldarriaga, 2002, eprint hep-ph/0212069.

Dwarkadas, V. V., R. A. Chevalier, and J. M. Blondin, 1996, Astrophys. J., **457**, 773.

Ebisuzaki, T., J. Makino, T. G. Tsuru, *et al.*, 2001, Astrophys. J. Lett. **562**, L19.

Eggen, O. J., D. Lynden-Bell, and A. Sandage, 1962, Astrophys. J. **136**, 748.

Eichler, D., M. Livio, T. Piran, and D. Schramm, 1989, Nature **340**, 126.

Eisenstein, D. J., and A. Loeb, 1996, Astrophys. J. **459**, 432.

Eskridge, P. B., J. A. Frogel, R. W. Pogge, *et al.*, 2000, Astron. J. **119**, 536.

Fabbiano, G., 1989, Annu. Rev. Astron. Astrophys. **27**, 87.

Fan, X., V. K. Narayanan, R. L. Lupton, *et al.*, 2001, Astron. J., **122**, 2833.

Ferguson, H. C., 1998, Rev. Mod. Astron. **11**, 83.

Ferguson, H. C., M. Dickinson, and C. Papovich, 2002, Astrophys. J. Lett., **569**, L65.

Ferguson, H. C., M. Dickinson, and A. Williams, 2000, Annu. Rev. Astron. Astrophys., **38**, 667.

Ferguson, H. C., and S. S. McGaugh, 1995, Astrophys. J. **440**, 470.

Ferrarese, L. and D. Merritt, 2000, Astrophys. J. Lett. **539**, L9.

Field, G. and A. Ferrara, 1995, Astrophys. J. **438**, 957.

Filippenko, A. V. and W. L. W. Sargent, 1986, Astron. J. **91**, 691.

Fishman, G. J., 2001, in *Supernovae and Gamma-Ray Bursts*, edited by M. Livio, N. Panagia, and K. Sahu (Cambridge University, Cambridge, England), p. 9.

Ford, H. C, R. J. Harms, Z. I. Tsvetanov, G. F. Hartig, L. L. Dressel, *et al.*, 1994, Astrophys. J.

Lett. **435**, L27.

Frail, D. A., S. R. Kulkarni, R. Sari, *et al.*, 2001, Astrophys. J. Lett. **562**, L55.

Frail, D. A., S. R. Kulkarni, M. Wieringa, G. Taylor, and G. Moriarty-Schieven, 1999, in *Gamma-Ray Bursts*, AIP Conf. Proc. 526, edited by M. Kippen (AIP, New York), p. 298.

Franceschini, A., L. Silva, G. Fasano, L. Granato, A. Bressan, S. Arnouts, and L. Danese, 1998, Astrophys. J. **506**, 600.

Frank, A., B. Balick, V. Icke, and G. Mellema, 1993, Astrophys. J. Lett. **404**, L25.

Freedman, W. L., B. F. Madore, B. K. Gibson, *et al.*, 2001, Astrophys. J. **553**, 47.

Fruchter, A., 2002, APS **Apr.Y2**, 004.

Fruchter, A. S. and R. N. Hook, 2002, Publ. Astron. Soc. Pac. **114**, 144.

Fruchter, A. S., S. Thorsett, M. Metzger, *et al.*, 1999, Astrophys. J. Lett. **519**, L13.

Galama, T. J., P. M. Vreeswijk, and J. van Paradijs, 1998, Nature **395**, 670.

Gallego, J., J. Zamorano, A. Aragon-Salamanca, and M. Rego, 1995, Astrophys. J. Lett. **455**, L1

Garcia-Segura, G., N. Langer, M. Rózyczka, and J. Franco, 1999, Astrophys. J. **517**, 767.

Garnavich, P. M., R. P. Kirshner, P. Challis, *et al.*, 1998, Astrophys. J. Lett. **493**, L53.

Garnavich, P. M., K. Z. Stanek, L. Wyrzykowski, *et al.*, 2003, Astrophys. J. **582**, 924.

Garriga, J., M. Livio, and A. Vilenkin, 2000, Phys. Rev. D **61**, 023503.

Gebhardt, K., R. Bender, G. Bower, *et al.*, 2000, Astrophys. J. Lett. **539**, L13.

Gebhardt, K., R. M. Rich, and L. C. Ho, 2002, Astrophys. J. Lett. **578**, L41.

Genzel, R., C. Pichon, A. Eckart, O. Gerhard, and T. Ott, 2000, Mon. Not. R. Astron. Soc. **317**, 348.

Gerhard, O. E., 1993, Mon. Not. R. Astron. Soc. **265**, 213.

Gerssen, J., R. P. van der Marel, K. Gebhardt, P. Guhathakurta, R. C., Peterson, and C. Pryor, 2002, Astron. J. **124**, 3270.

Gerssen, J., R. P. van der Marel, K. Gebhardt, P. Guhathakurta, R. C., Peterson, and C. Pryor, 2003, Astron. J., (in press).

Ghez, A. M., B. L. Klein, M. Morris, and E. E. Becklin, 1998, Astrophys. J. **509**, 678.

Giavalisco, M., 2002, Annu. Rev. Astron. Astrophys. **40**, 579.

Giavalisco, M., M. Livio, R. C. Bohlin, F. D. Macchetto, and T. P. Stecher, 1996, Astron. J. **112**, 369.

Gilliland, R. L., T. M. Brown, P. Guhathakurta, *et al.*, 2000, Astrophys. J. Lett. **545**, L47.

Gilliland, R. L., P. E. Nugent, and M. M. Phillips, 1999, Astrophys. J. **521**, 30.

Giovanelli, R., M. P. Haynes, N. P. Vogt, G. Wegner, J. J. Salzer, L. N. Da Costa, and W. Freudling, 1997, Astron. J. **113**, 22.

Gonzalez, G., 1997, Mon. Not. R. Astr. Soc. **285**, 403.

Goodman, J., 1986, Astrophys. J. Lett. **308**, L47.

Graham, J., I. DePater, J. Jernigan, M. Livio, and M. Brown, 1995, Science **267**, 1320.

Greiner, J., S. Klose, A. Zeh, *et al.*, 2001, GCN **1166**, 1.

Griffiths, R. E., K. V. Ratnatunga, L. W. Neuschaefer, *et al.*, 1994, Astrophys. J. **437**, 67.

Guth, A. H., 1981, Phys. Rev. D **23**, 347.

Haehnelt, M. G. and G. Kauffmann, 2000, Mon. Not. R. Astron. Soc. **318**, L35.

Hammel, H. B., R. F. Beebe, A. P. Ingersoll, *et al.*, 1995, Science **267**, 1288.

Hamuy, M., M. M. Phillips, N. B. Suntzeff, R. A. Schommer, J. Maza, and R. Avilés, 1996, Astron. J. **112**, 2398.

Harms, R. J., H. C. Ford, Z. I. Tsvetanov, *et al.*, 1994, Astrophys. J. Lett. **435**, L35.

Henney, W. J. and C. R. O'Dell, 1999, Astron. J. **118**, 2350.

Henry, G. W., G. W. Marcy, R. P. Butler, and S. S. Vogt, 2000, Astrophys. J. Lett. **529**, L41.

Hillebrand, L. A., S. E. Strom, N. Calvet, K. M. Merrill, I. Gatley, R. B. Makidon, M. R. Meyer, and M. F. Skrutskie, 1998, Astron. J. **116**, 1816.

Hillebrandt, W. and J. C. Niemeyer, 2000, Annu. Rev. Astron. Astrophys. **38**, 191.

Hillier, D. J., K. Davidson, K. Ishibashi, and T. Gull, 2001, Astrophys. J. **553**, 837.

Holz, D. E., 1998, Astrophys. J. Lett. **506**, L1.

Hopkins, A. M., A. J. Connolly, and A. S. Szalay, 2000, Astron. J., **120**, 2843.

Howell, J. H., R. Guhathakurta, and R. L. Gilliland, 2000, Publ. Astron. Soc. Pac. **112**, 775.

Hu, W., M. Fukugita, M. Zaldarriaga, and M. Tegmark, 2000, Astrophys. J. **549**, 669.

Hubble, E., 1929, Proc. Nat. Acad. Sci. **15**, 168.

Hubble, E. and M. L. Humason, 1931, Astrophys. J. **74**, 43.

Hubble, E. P., 1925, Astrophys. J. **62**, 409.

Hubble, E. P., 1926, Astrophys. J. **64**, 321.

Hutchings, J. B. and B. Campbell, 1983, Nature **303**, 584.

Hutchings, J. B. and T. Janson, and S. G. Neff, 1989, Astrophys. J. **342**, 660.

Iben, I. Jr., 1985, Quart. J. Roy. Astron. Soc. **26**, 1.

Iben, I. Jr. and M. Livio, 1993, Publ. Astro. Soc. Pac. **105**, 1373.

Iben, I. Jr. and A. V. Tutukov, 1993, Astrophys. J. **418**, 343.

Icke, V., G. Mellema, B. Balick, F. Eulderink, and A. Frank, 1992, Nature, **355**, 524.

Im, M., R. E. Griffiths, A. Naim, *et al.*, 1999, Astrophys. J. **510**, 82.

Ingersoll, A. P. and H. Kanamori, 1996, in *The Collision of Comet Shoemaker-Levy 9 and Jupiter*, edited by K. S. Noll, H. A. Weaver, and P. D. Feldman, (Cambridge University, Cambridge, England), p. 329.

Ishibashi, K., M. F. Corcoran, K. Davidson, J. H. Swank, R. Petre, S. A. Drake, A. Damineli, and S. White, 1999, Astrophys. J. **524**, 983

Iwamoto, K., P. A. Mazzali, and K. Nomoto, 1998, Nature **395**, 672.

Jacoby, G. H., D. Branch, R. Ciardullo, R. L. Davies, W. H. Harris, M. J. Pierce, C. J. Pritchett, J. L. Tonry, and D. L. Welch, 1992, Publ. Astron. Soc. Pac. **104**, 599.

Jeltema, J. E., C. R. Canizares, D. A. Buote, and G. P. Garmire, 2002, Astrophys. J. (submitted).

Johnstone, D., D. Hollenbach, and J. Bally, 1998, Astrophys. J., **499**, 758.

Kahn, F. D., 1982, in *Planetary Nebulae*, IAU Symp. 103, edited by D. R. Flower (Reidel, Dordrecht), p. 305.

Kalas, P., J. Larwood, B. A. Smith, and A. Schultz, 2000, Astrophys. J. Lett. **530**, L133.

Kallosh, R. and A. Linde, 2002, eprint hep-th/0208157.

Kauffmann, G., J. Colberg, A. Diaferio, and S. D. M. White, 1999, Mon. Not. R. Astron. Soc. **303**, 188.

Kauffmann, G., S. D. M. White, and B. Guideroni, 1993, Mon. Not. R. Astron. Soc. **264**, 201.

King, A. R., M. B. Davies, M. J. Ward, G. Fabbiano, and M. Elvis, 2001, Astrophys. J. Lett. **552**, L109.

Kirhakos, S., J. N. Bahcall, D. P. Schneider, and J. Kristian, 1999, Astrophys. J. **520**, 67.

Kirshner, R. P. and J. Kwan, 1974, Astrophys. J. **193**, 27.

Königl, A., 1989, Astrophys. J. **342**, 208.

Kormendy, J., 1988, Astrophys. J. **325**, 128.

Kormendy, J. and D. Richstone, 1995, Annu. Rev. Astron. Astrophys. **33**, 581.

Kouveliotou, C., C. A. Meegan, G. J. Fishman, N. P. Bhat, M. S. Briggs, T. M. Koshut, W. S. Paciesas, and G. N. Pendleton, 1993, Astrophys. J. Lett. **413**, L101.

Kristian, J., 1973, Astrophys. J. Lett. **179**, L61.

Krolik, J. H., 1999a, Astrophys. J. Lett. **515**, L73.

Krolik, J. H., 1999b, *Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment*, (Princeton University Press, Princeton).

Kukula, M. J., J. S. Dunlop, D. H. Hughes, G. Taylor, and T. Boroson, 1996, in *Quasar Hosts*, edited by D. Clements and I. Perez-Fournon, (Springer, Berlin), p. 177.

Kulkarni, S. R., S. G. Djorgovski, S. C. Odewahn, *et al.*, 1999, Nature **398**, 389.

Kwok, S., 1982, Astrophys. J. **258**, 280.

Lada, C. J., A. A. Muench, K. E. Haisch, *et al.*, 2000, Astron. J. **120**, 3162.

Lanzetta, K. M., D. V. Bowen, D. Tytler, and J. K. Webb, 1995, Astrophys. J. **442**, 538.

Lanzetta, K. M., N. Yahata, S. Pascarelle, H.-W. Chen, and A. Fernández-Soto, 2002, Astrophys. J. **570**, 492.

Lauer, T. R., S. M. Faber, C. R. Lynds, *et al.*, 1992, Astron. J. **103**, 703.

Laughlin, G., 2000, Astrophys. J. **545**, 1064.

Lemaître, G., 1927, Ann. Acad. Sci. Bruxelles **47A**, 49.

Li, W., A. V. Filippenko, R. R. Treffers, A. G. Riess, J. Hu, and Y. Qiu, 2001, Astrophys. J. **546**, 734.

Lilly, S. J., O. Le Févre, F. Hammer, and D. Crampton, 1996, Astrophys. J. Lett. **460**, L1.

Lin, D. N. C., P. Bodenheimer, and D. C. Richardson, 1996, Nature, **380**, 606.

Linde, A., 1986, Phys. Rev. Lett. **A1**, 81.

Linde, A., D. Linde, and A. Mezhlumian, 1995, Phys. Lett., **B345**, 203.

Lira, P., R. Johnson, and A. Lawrence, 2002, Mon. Not. R. Astron. Soc. (submitted).

Lissauer, J. J., 1987, Icarus **69**, 249.

Livio, M., 1994, in *Circumstellar Media in the Late Stages of Stellar Evolution*, edited by R. Clegg *et al.* (Cambridge University, Cambridge, England), p. 35.

Livio, M., 1996, in *Evolutionary Processes in Binary Stars*, edited by R. A. M. J. Wijers, M. B. Davies, and C. A. Tout (Kluwer, Dordrecht), p. 141.

Livio, M., 2000, in *Cosmic Explosions*, AIP Conference Proceedings No. 522, edited by S. S. Holt and W. W. Zhang (AIP, Melville), p. 275.

Livio, M., 2001, in *Supernovae and Gamma-Ray Bursts*, edited by M. Livio, N. Panagia, and K. Sahu (Cambridge University, Cambridge, England), p. 334.

Livio, Mario, S. Michael Fall, and Piero Madau, 1998, *The Hubble Deep Field* (Cambridge University, Cambridge, England).

Livio, M. and N. Soker, 2001, Astrophys. J. **552**, 685.

Livio, M. and E. Waxman, 2000, Astrophys. J. **538**, 187.

Luo, D., R. McCray, and J. Slavin, 1994, Astrophys. J. **430**, 264.

Lynden-Bell, D., 1969, Nature **223**, 690.

Macchetto, F. D., A. Marconi, D. J. Axon, A. Capetti, W. Sparks, and P. Crane, 1997, Astrophys. J. **489**, 579.

MacLow, M.-M., 1996, in *The Collision of Comet Shoemaker-Levy 9 and Jupiter*, edited by K. S. Noll, H. A. Weaver, and P. D. Feldman, (Cambridge University, New York), p. 157.

Madau, P., H. C. Ferguson, M. E. Dickinson, M. Giavalisco, C. C. Steidel, and A. Fruchter, 1996, Mon. Not. R. Astron. Soc. **283**, 1388.

Magorrian, J., S. Tremaine, D. Richstone, *et al.*, 1998, Astron. J. **115**, 2285.

Marcy, G. W., W. D. Cochran, and M. Mayor, 2000, in *Protostars and Planets IV*, edited by V. Mannings, A. P. Boss, and S. Russell, (Univ. Arizona, Tucson), p. 1285.

Mayer, L., T. Quinn, J. Wadsley, and J. Stadel, 2002, Science **298**, 1756.

Mayor, M. and D. Queloz, 1995, Nature **378**, 355.

Mazeh, T., D. Naef, G. Torres, *et al.*, 2000, Astrophys. J. Lett., **532**, L55.

McCarthy, P. J., R. G. Carlberg, H.-W. Chen, *et al.*, 2001, Astrophys. J. Lett. **560**, L131.

McCray, R., 2003, in *A Decade of HST Science*, edited by Mario Livio, Keith Noll, and Massimo Stiavelli, (Cambridge University, Cambridge, England), p. 64.

Meaburn, J., 1988, Mon. Not. R. Astron. Soc. **233**, 791.

Mellema, G., 1995, Mon. Not. R. Astron. Soc. **277**, 173.

Mészáros, P., 2002, Annu. Rev. Astron. Astrophys. **40**, 137.

Mészáros, P. and M. J. Rees, 1997, Astrophys. J. **476**, 232.

Mészáros, P. and M. J. Rees, 1999, Mon. Not. R. Astron. Soc. **306**, L39.

Metcalfe, N., T. Shanks, A. Campos, R. Fong, and J. P. Gardner, 1996, Nature **383**, 236.

Metzger, M., S. G. Djorgovski, S. Kulkarni, C. Steidel, and K. Adelberger, 1997, Nature **387**, 878.

Meurer, G. R., T. M. Heckman, and D. Calzetti, 1999, Astrophys. J. **521**, 64.

Moustakas, L. A. and R. S. Somerville, 2002, Astrophys. J. **577**, 1.

Najita, J., 2000, in *Unsolved Mysteries in Stellar Evolution*, edited by M. Livio (Cambridge University, Cambridge, England), p. 25.

Nakano, S., 1993, IAU Circular **5800**.

Netterfield, C. B., P. A. R. Ade, J. J. Bock, *et al.*, 2002, Astrophys. J. **571**, 604.

Nicholson, P., T. Gierasch, T. Hayward, *et al.*, 1995, Geophys. Res. Lett. **22**, 1613.

Nomoto, K., 1982, Astrophys. J. **253**, 798.

Nomoto, K., K. Iwamoto, N. Nakasato, F. K. Thielemann, F. Brochwitz, T. Young, T. Shigeyama, T. Tsufimoto, and Y. Yoshii, 1997, in *Thermonuclear Supernovae*, edited by P. Ruiz-Lapuente, R. Canal, and J. Isern, (Kluwer, Dordrecht), p. 349.

Nørgaard-Nielsen, H. V., L. Hansen, H. E. Jorgensen, A. A. Salamanca, R. S. Ellis, and W. J. Couch, 1989, Nature **339**, 523.

O'Dell, C. R., 2001, Annu. Rev. Astron. Astrophys. **39**, 99.

O'Dell, C. R., B. Balick, A. R. Hajian, W. J. Henney, and A. Burkett, 2003, R. Mx. A. C. **15**, 29.

O'Dell, C. R., Z. Wen, and X. Hu, 1993, Astrophys. J. **410**, 696.

O'Dell, C. R. and S. K. Wong, 1996, Astron. J. **111**, 846.

Ogilvie, G. I., and M. Livio, 2001, Astrophys. J. **553**, 158.

Ostriker, J. P., 2000, Phys. Rev. Lett. **84**, L5258.

Owocki, S. P., S. R. Crammer, and J. M. Blondin, 1994, Astrophys. J. **424**, 887.

Paczynski, B., 1986, Astrophys. J. Lett. **308**, L43.

Paczynski, B., 1998, Astrophys. J. Lett. **494**, L45.

Panagia, N., 2002, in *Multifrequency Behavior of High Energy Cosmic Sources*, 2001 Frascati Workshop, in press.

Panaitescu, A., and P. Kumar, 2001, Astrophys. J. **554**, 667.

Parks, G. K., 1991, *Physics of Space Plasmas: An Introduction* (Addison-Wesley, New York).

Pearce, F. R., A. Jenkins, C. S. Frenk, S. D. M. White, P. A. Thomas, H. M. P. Conchman, J. A. Peacock, and G. Efstathiou, 2001, Mon. Not. R. Astron. Soc. **326**, 649.

Peebles, P. J. E. and B. Ratra, 2002, eprint astro-ph/0207347.

Pei, Y. and S. M. Fall, 1995, Astrophys. J. **454**, 69.

Pei, Y., S. M. Fall, and M. G. Hauser, 1999, Astrophys. J., **522**, 604.

Perlmutter, S., G. Aldering, M. Della Valle, *et al.*, 1998, Nature **391**, 51.

Perlmutter, S., G. Aldering, G. Goldhaber, *et al.*, 1999, Astrophys. J. **517**, 565.

Phillips, M. M., 1993, Astrophys. J. Lett. **413**, L105.

Podsiadlowski, Ph., S. Rapport, and Z. Han, 2002, Mon. Not. R. Astron. Soc. (submitted).

Pollack, J. B., O. Hubickys, P. Bodenheimer, J. J. Lissauer, M. Podolak, and Y. Greenzweig, 1996, Icarus **124**, 62.

Prialnik, D., 2000, *An Introduction to the Theory of Stellar Structure and Evolution*, (Cambridge University, Cambridge, England), p. 174.

Price, P. A., E. Berger, D. E. Reichert, *et al.*, 2002, Astrophys. J. Lett. **572**, L51.

Pringle, J. E., 1989, Mon. Not. R. Astron. Soc. **238**, 37.

Pun, C. S. J., 1997, eprint http://oposite.stsci.edu//pubinfo/PR/97/03.html/.

Rasio, F. A. and M. Livio, 1996, Astrophys. J. **471**, 366.

Ratra, B. and P. J. E. Peebles, 1988, Phys. Rev. D **37**, 3406.

Rees, M. J., 1978, Nature **275**, 516.

Rees, M. J., 1984, Annu. Rev. Astron. Astrophys. **22**, 471.

Reese, E. D., J. E. Carlstrom, M. Joy, J. J. Mohr, L. Grego, and W. L. Holzapfel, 2002, Astrophys. J. **581**, 53.

Reese, E. D., J. J. Mohr, J. E. Carlstrom, M. Joy, L. Grego, G. P. Holder, W. L. Holzapfel, J. P. Hughes, S. K. Patel, and M. Donahue, 2000, Astrophys. J. **533**, 38.

Reid, I. N., 2002, Publ. Astron. Soc. Pac. **114**, 306.

Reipurth, B., 1999, in *A General Catalog of Herbig-Haro Objects*, second edition, http://casa.colorado.edu/hhcat/.

Reipurth, B., D. Devine, and J. Bally, 1998, Astron. J. **116**, 1396.

Reipurth, B. and S. Heathcote, 1993, in *Astrophysical Jets*, edited by D. Burgarella, M. Livio, and C. P. O'Dea, (Cambridge University, Cambridge, England), p. 35.

Rhoads, J., 1997, Astrophys. J. Lett. **487**, L1.

Riess, A. G., 2000, Publ. Astron. Soc. Pac. **112**, 1284.

Riess, A. G., A. V. Filippenko, P. Challis, *et al.*, 1998, Astron. J. **116**, 1009.

Riess, A. G., A. V. Filippenko, M. C. Liu, *et al.*, 2000, Astrophys. J. **526**, 62.

Riess, A. G., R. P. Kirshner, B. P. Schmidt, *et al.*, 1999, Astron. J. **117**, 707.

Riess, A. G., P. E. Nugent, R. L. Gilliland, *et al.*, 2001, Astrophys. J. **560**, 49.

Riess, A. G., W. H. Press, and R. P. Kirshner, 1996, Astrophys. J. **473**, 88.

Roberts, T. P., M. R. Good, M. J. Ward, R. S. Warwick, P. T. O'Brien, P. Lira, and A. D. P. Hands, 2001, Mon. Not. R. Astron. Soc. **325**, L7.

Roche, N., K. Ratnatunga, R. E. Griffiths, M. Im, and A. Naim, 1998, Mon. Not. R. Astron. Soc.

293, 157.

Saha, A., A. Sandage, G. A. Tammann, A. E. Dolphin, J. Christensen, N. Panagia, and F. D. Machetto, 2001, Astrophys. J. **562**, 314.

Sahni, V. and A. Starobinsky, 2000, Int. J. Mod. Phys. D **9**, 373.

Sahu, K. C., M. Livio, L. Petro, F. D. Macchetto, J. van Paradijs, C. Kouveliotou, G. J. Fishman, C. A. Meegan, P. J. Groot, and T. J. Calama, 1997, Nature **387**, 476.

Salamanca, I., E. Rol, N. Tanvir, and L. Kaper, 2002, GCN 1443.

Salaris, M. and A. Weiss, 1998, Astron. Astrophys. **335**, 943.

Sandage, A., 1958, Astrophys. J. **127**, 513.

Sargent, W. L. W., P. J. Young, A. Boksenberg, K. Shortridge, C. R. Lynds, and F. D. A., Hartwick, 1978, Astrophys. J. **221**, 731.

Schecter, P. L., 2000, eprint astro-ph/0009048.

Schmidt, B. P., R. P. Kirshner, R. G. Eastman, M. M. Phillips, N. B. Suntzeff, M. Hamuy, J. Maza, and R. Avilés, 1994, Astrophys. J., **432**, 42.

Schneider, G., B. A. Smith, E. E. Becklin, D. W. Koerner, R. Meier, D. C. Hines, P. J. Lowrance, R. J. Terrile, R. I. Thompson, and M. Rieke, 1999, Astrophys. J. Lett. **513**, L127.

Seager, S. and D. D. Sasselov, 2000, Astrophys. J. **537**, 916.

Sekanina, Z., P. W. Chodas, and D. K. Yeomans, 1994, Astron. Astrophys. **304**, 296.

Selwood, J. A., 2000, in *Dynamics of Galaxies: from the Early Universe to the Present*, ASP Conf. Ser. 197, edited by F. Combes, G. A. Mamon, and V. Charmandaris (ASP, San Francisco), p. 3.

Shoemaker, C. S., E. M. Shoemaker, and D. Levy, 1993, IAU Circular **5725**.

Shu, F. H., D. Johnstone, and D. Hollenbach, 1993, Icarus **106**, 92.

Shu, F. H., J. Najita, E. Ostriker, F. Wilkin, S. Ruden, and S. Lizano, 1994, Astrophys. J. **429**, 781.

Silk, J. and M. J. Rees, 1998, Astron. Astrophys. **331**, L1.

Slipher, V. M., 1917, Proc. Amer. Phil. Soc. **56**, 403.

Soker, N. and M. Livio, 1989, Astrophys. J. **339**, 268.

Soker, N. and S. Rappaport, 2001, Astrophys. J. **557**, 256.

Somerville, R. S. and M. Livio, 2003, Astrophys. J. (in press); eprint astro-ph/0303017.

Somerville, R. S., J. R. Primak, and S. M. Faber, 2001, Mon. Not. R. Astron. Soc. **320**, 504.

Speck, A. K., M. Meixner, D. Fong, P. R. McCullough, D. E. Moser, and T. Veta, 2002, Astron. J.

123, 346.

Spergel, D. N. and P. J. Steinhardt, 2000, Phys. Rev. Lett. **84**, 3760.

Spergel, D. N., L. Verde, H. V. Peiris, *et al.*, 2003, Astrophys. J. (in press); eprint astro-ph/0302209.

Spruit, H. C., 1996, in *Evolutionary Processes in Binary Stars*, edited by R. Wijers, M. Davies, and C. Tout (Cambridge University, Cambridge, England), p. 249.

Stanway, E. R., A. J. Bunker, and R. G. McMahon, 2003, Mon. Not. R. Astron. Soc. (in press); eprint astro-ph/0302212.

Steidel, C. C., K. L. Adelberger, M. Giavalisco, M. Dickinson, and M. Pettini, 1999, Astrophys. J. **519**, 1.

Steinhardt, P. J., 1983, in *The Very Early Universe*, edited by G. W. Gibbons, S. Hawking, and S. T. C. Siklos (Cambridge University, Cambridge, England), p. 251.

Strauss, M. A. and J. A. Willick, 1995, Phys. Rep. **261**, 271.

Sunyaev, R. A. and Y. B. Zeldovich, 1970, Comments Astrophys. Space Phys. **2**, 66.

Tammann, G. A., 1979, in *ESA/ESO Workshop on Astronomical Uses of the Space Telescope*, edited by F. Macchetto, F. Pacini, and M. Tarenghi (ESO, Geneva), p. 229.

Terman, J. L., R. E. Taam, and L. Hernquist, 1994, Astrophys. J. **422**, 729.

Terman, J. L., R. E. Taam, and L. Hernquist, 1995, Astrophys. J. **445**, 367.

Thompson, R., M. Rieke, G. Schneider, D. Hines, R. Sahai, *et al.*, 1997, ST ScI Press Release.

Thompson, R. I., L. J. Storrie-Lombardi, P. J. Weymann, M. J. Rieke, G. Schneider, E. Stobie, and D. Lytle, 1999, Astron. J. **117**, 17.

Throop, H., 2000, Ph.D. thesis (University of Colorado, Boulder).

Tinsley, B. M. and J. Gunn, 1976, Astrophys. J. **302**, 52.

Tonry, J. L., J. P. Blakeslee, E. A. Aghar, and A. Dressler, 1997, Astrophys. J. **475**, 399.

Tonry, J. L., J. P. Blakeslee, E. A. Aghar, and A. Dressler, 2000, Astrophys. J. **530**, 625.

Tonry, J. L. and D. P. Schneider, 1988, Astron. J. **96**, 807.

Tremaine, S., K. Gebhardt, R. Bender, *et al.*, 2002, Astrophys. J. **574**, 740.

Treu, T. and M. Stiavelli, 1999, Astrophys. J. Lett. **524**, L27.

Treu, T., M. Stiavelli, G. Bertin, S. Casertano, and P. Møller, 2001, Mon. Not. R. Astron. Soc. **326**, 237.

Trimble, V., 1997, Space Sci. Rev. **79**, 793.

Tully, R. B. and J. R. Fisher, 1977, Astron. Astrophys. **54**, 661.

Turner, M. S., 2000, Phys. Rep. **334**, 619.

Turner, M. S. and M. White, 1997, Phys. Rev. D **56**, R4439.

Umeda, H., K. Nomoto, C. Kobayashi, I. Hachisu, and M. Kato, 1999, Astrophys. J. Lett. **522**, L43.

van den Bergh, S., 1994, Astrophys. J. Suppl. Ser. **92**, 219.

van den Bergh, S., R. G. Abraham, L. F. Whyte, M. R. Merrifield, P. B. Eskridge, J. A. Frogel, and R. Pogge, 2002, Astron. J. **123**, 2913.

van der Marel, R., 1994, Mon. Not. R. Astron. Soc. **270**, 271.

van der Marel, R. P., J. Gerssin, P. Guhathakurta, R. C. Peterson, and K. Gebhardt, 2002, Astron. J. **124**, 3255.

van Paradijs, J., P. Groot, T. Galama, *et al.*, 1997, Nature **386**, 686.

Vidal-Madjar, A., A. Lecavelier des Etangs, J.-M. Désert, G. E. Ballester, R. Ferlet, G. Hébrard, and M. Mayor, 2003, Nature **422**, 143.

Vilenkin, A., 1983, Phys. Rev. D **27**, 2848.

Vilenkin, A., 1995, Phys. Rev. Lett. **74**, 846.

Vilenkin, A., 1998, Phys. Rev. Lett. **81**, 5501.

Vishniac, E. T., 1983, Astrophys. J. **274**, 152.

Wagoner, R. V., 1977, Astrophys. J. Lett. **214**, L5.

Weaver, H. A., M. F. A'Hearn, C. Arpigny, *et al.*, 1995, Science **267**, 1282.

Weaver, H. A., P. D. Feldman, M. F. A'Hearn, *et al.*, 1994, Science **263**, 787.

Webbink, R. F., 1984, Astrophys. J. **227**, 355.

Weinberg, S., 2001, in *Sources and Detection of Dark Matter and Dark Energy in the Universe*, edited by D. B. Cline, (Springer, Berlin), p. 18.

Whelan, J. and I. Iben Jr., 1973, Astrophys. J. **186**, 1007.

Windhorst, R. A., S. P. Driver, E. J. Ostrander, *et al.*, 1995, in *Galaxies in the Young Universe*, edited by H. Hippelein, (Springer, Berlin), p. 265.

Woosley, S. E., 1993, Astrophys. J. **405**, 273.

Wyckoff, S., T. Gehren, D. C. Morton, R. Albrecht, P. A. Wehinger, and A. Boksenberg, 1980, Astrophys. J. **242**, 59.

Yeomans, D. K. and P. W. Chodas, 1993, IAU Circular **5807**.

Yungelson, L. R. and M. Livio, 2000, Astrophys. J. **528**, 108.

Zahnle, K., 1996, in *The Collision of Comet Shoemaker-Levy 9 and Jupiter*, edited by K. S. Noll, H. A. Weaver, and P. D. Feldman, (Cambridge University, Cambridge, England), p. 183.

Zahnle, K. and M.-M. MacLow, 1994, Icarus **108**, 1.

Zeldovich, Ya. B. and I. D. Novikov, 1964, Sov. Phys. Dokl. **9**, 195.

Zepf, S. E., 1997, Nature **390**, 377.

Zlatev, I., L. Wang, and P. J. Steinhardt, 1998, Phys. Rev. Lett. **82**, 896.

TABLE I The early history of the Hubble constant. Unless otherwise shown, stated error bars were 10% or less, or not given.

When	Who	Numerical value
1927	Lemaitre	600
1929	Hubble	530–513–465
1931	Hubble & Humason	558
1936	Hubble	526
1946	Mineur	330
1951	Behr	250
1952	Baade, Thackeray	270
1956	Humas, Mayall & Sandage	180
1958	Holmberg	134
1958	Sandage	150–75–38
1959	McVittie	227–143
1960	Sersic	125
1960	van den Bergh	125
1960	van den Bergh	125
1961	Ambartsumyan	140–60
1961	Sandage	113–85
1964	de Vaucouleurs	125
1968–69	de Vaucouleurs	100
1969	van den Bergh	110–83
1968–76	Sandage & Tammann	50
1972	Sandage	55
1979	de Vaucoulers	100

TABLE II H_0 from secondary methods (the Key Project)

Method	H_0	Error (random, systematic) (%)
36 Type Ia SN, $4000 < cz < 30,000$ km s^{-1}	71	$\pm 2 \pm 6$
21 TF clusters, $1000 < cz < 9000$ km s^{-1}	71	$\pm 3 \pm 7$
11 FP clusters, $1000 < cz < 11,000$ km s^{-1}	82	$\pm 6 \pm 9$
SBF for 6 clusters, $3800 < cz < 5800$ km s^{-1}	70	$\pm 5 \pm 6$
4 Type II SN, $1900 < cz < 14,200$ km s^{-1}	72	$\pm 9 \pm 7$

TABLE III Mean absolute B, V, and I magnitudes of nine SNe Ia with known Cepheid distances, without and with corrections for decline rate and color

SN	Galaxy	$(m-M)^o$	M_B^0	M_V^o	M_I^o	Δm_{15}	$(B-V)^o$	$M_B^{\rm corr}$	$M_V^{\rm corr}$	$M_I^{\rm corr}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1937C	1C 4182	28.36(12)	−19.56 (15)	19.54 (17)	...	0.87 (10)	−0.02	−19.39 (18)	−19.37 (17)	...
1960F	NGC 4496A	31.03 (10)	−19.56 (18)	−19.62 (22)	...	1.06 (12)	0.06	−19.67 (18)	−19.65 (22)	...
1972E	NGC 5253	28.00 (07)	−19.64 (16)	−19.61 (17)	−19.27 (20)	0.87 (10)	−0.03	−19.44 (16)	−19.42 (17)	−19.12 (20)
1974G	NGC 4414	31.46 (17)	−19.67 (34)	−19.69 (27)	...	1.11 (06)	0.02	−19.70 (34)	−19.69 (27)	...
1981B	NGC 4536	31.10 (12)	−19.50 (18)	−19.50 (16)	...	1.10 (07)	0.00	−19.48 (18)	−19.46 (16)	...
1989B	BGC 3627	30.22 (12)	−19.47 (18)	−19.42 (16)	−19.21 (14)	1.31 (07)	−0.05	−19.42 (18)	−19.41 (16)	−19.20 (14)
1990N	NGC 4639	32.03 (22)	−19.39 (26)	−19.41 (24)	−19.14 (23)	1.05 (05)	0.02	−19.39 (26)	−19.38 (24)	−19.02 (23)
1998bu	NGC 3368	30.37 (16)	−19.76 (31)	−19.69 (26)	−19.43 (21)	1.08 (05)	−0.07	−19.56 (31)	−19.55 (36)	−19.31 (21)
1998aq	NGC 3982	31.72 (14)	−19.56 (21)	−19.48 (20)	...	1.12 (03)	−0.08	−19.35 (24)	−19.34 (23)	...
Straight mean			−19.57 (04)	−19.55 (04)	−19.26 (06)			−19.49 (04)	−19.47 (04)	−19.16 (06)
Weighted mean			−19.56 (07)	−19.53 (06)	−19.25 (09)			−19.47 (07)	−19.46 (06)	−19.19 (09)

FIG. 1 Hubble Space Telescope in orbit photographed by the crew of the Space Shuttle Columbia, just after being released during mission STS-109 to service and upgrade Hubble. March 9, 2002. Credit: NASA. http://spaceflight.nasa.gov/gallery/images/shuttle/sts-109/html/s109e5700.html

FIG. 2 Comet P/Shoemaker-Levy 9 (1993e) "String of Pearls," HST/WFPC2, January 1994. Credit: NASA, H. A. Weaver (JHU) and T. E. Smith (STScI).
http://hubblesite.org/newscenter/archive/1994/13/

FIG. 3 Series of images showing the impact of Comet P/Shoemaker-Levy 9 (1993e) fragment W on Jupiter, HST/WFPC2, July 22, 1994. Credit: NASA and the HST Comet Impact Team.

FIG. 4 The light curve from the impact of fragment R at 2.3 μ, adapted from Graham *et al.* (1995) and Zahnle (1996).

FIG. 5 The viewing geometry of a typical SL-9 impact (see text; adapted from Zahnle, 1996).

FIG. 6 Flat Projection of Comet P/Shoemaker-Levy 9 (1993e) Fragments D (left) and G Impacts on Jupiter, HST/WFPC2, July 18, 1994. Two rings of propagating waves can be seen. Credit: NASA and the HST Comet Impact Team. http://hubblesite.org/newscenter/archive/1994/36/

FIG. 7 Globular Cluster 47 Tucanae (NGC 104), HST/WFPC2, July 1999. Credit: NASA and R. Gilliland (STScI). http://hubblesite.org/newscenter/archive/2000/33/

FIG. 8 Light curve of the slight dimming of the star HD 209458 due to a planet passing directly in front of it, HST/STIS, April-May 2000. Credit: NASA, T. M. Brown, D. Charbonneau, R. L. Gilliland, R. W. Noyes, & A. Burrows.

FIG. 9 Montage of four images showing jets from young stars HST/WFPC2.
Top Left: HH 47 Credit: NASA and J. Morse (University of Colorado).
http://hubblesite.org/newscenter/archive/1995/24/
Top Right: DG Tau B Credit: NASA, C. Burrows (STScI), J. Krist (STScI), K. Stapelfeldt (JPL), and the WFPC2 Science Team. http://hubblesite.org/newscenter/archive/1999/05/
Bottom Left: HH 34 Credit: NASA, J. Hester (Arizona State University), and the WFPC2 Investigation Definition Team. http://hubblesite.org/newscenter/archive/1995/24/
Bottom Right: HH 1-2 Credit: NASA, J. Hester (Arizona State University), and the WFPC2 Investigation Definition Team. http://hubblesite.org/newscenter/archive/1995/24/

FIG. 10 HH 30 disk/jet, HST/WFPC2. Credit: NASA, C. Burrows (STScI), J. Krist (STScI), K. Stapelfeldt (JPL), and the WFPC2 Science Team.
http://hubblesite.org/newscenter/archive/1999/05/

FIG. 11 A bright one-sided microjet emerging from a proplyd located below Orion's Bright Bar. Adapted from Bally *et al.* (2000).

FIG. 12 Protoplanetary disks (Proplyds) in the Orion Nebula (M42), HST/WFPC2. Credit: NASA, C. R. O'Dell (Vanderbilt University), and M. McCaughrean (Max-Planck-Institute for Astronomy). http://hubblesite.org/newscenter/archive/1995/45/

FIG. 13 Evolutionary tracks of 1 $M_\odot$, 5 $M_\odot$, and 25 $M_\odot$ stars in the Luminosity-Effective Temperature (Hertzspring-Russell) diagram. Thick segments mark long, nuclear burning, evolutionary phases (from Iben 1985; adapted from Prialnik, 2000).

FIG. 14 The Luminous Blue Variable star Eta Carinae, HST/WFPC2, 1994. Credit: NASA and J. Hester (Arizona State University). http://hubblesite.org/newscenter/archive/1994/09/

FIG. 15 Rings around Supernova 1987A, HST/WPFC2, February 1994. Credit: NASA and C. Burrows (ESA/STScI). http://hubblesite.org/newscenter/archive/1994/22/

FIG. 16 Planetary Nebula MyCn18, The Hourglass Nebula, HST/WFPC2. Credit: NASA, R. Sahai, J. Trauger (JPL) and the WFPC2 Science Team.
http://hubblesite.org/newscenter/archive/1996/07/

FIG. 17 Nebula surrounding the symbiotic star system He2−104, the Southern Crab Nebula, HST/WFPC2, May 1999. Credit: NASA and R. Corradi (Instituto de Astrofisica de Canarias, Tenerife, Spain), M. Livio (Space Telescope Science Institute), U. Munari (Osservatorio Astronomico di Padova-Asiago, Italy), H. Schwarz (Nordic Optical Telescope, Canarias, Spain).
http://hubblesite.org/newscenter/archive/1999/32/

FIG. 18 A schematic distribution of the planetary nebulae morphology in the rotation (Ω–magnetic energy density (σ) plane (see text; adapted from Garcia-Segura *et al.*, 1999).

FIG. 19 An HST-WFPC2 image of the circumstellar ring around supernova 1987A, obtained in May 2002 with narrow band filters that include Hα 6563 Å and [NII] 6584 Å emission line radiation. We can identify a number of bright spots that correspond to gaseous protuberances on the inner side of the ring that are currently being hit by the supernova ejecta, thus becoming hot and luminous. Because of the narrow width of the filters employed to gather this image, the ring and the spots appear bright since they are mostly radiating in emission lines. On the other hand, both the supernova itself (a wide patch at the center of the ellipse) and a star (projected by chance near spot #6) are not visible in this image because most of their energy is radiated in the form of a rather smooth continuum. Credits: Nino Panagia (ESA/STScI), on behalf of the SINS Collaboration.

FIG. 20 Jet and disk of stars and gas in the active galaxy M87, HST/WFPC2, 1994. Credit: NASA and H. Ford (JHU). http://hubblesite.org/newscenter/archive/1994/23/

FIG. 21 Black hole mass versus bulge luminosity (left), and the luminosity-weighted dispersion (right). Adopted from Gebhardt *et al.* (2000).

FIG. 22 Schematic gamma-ray burst from internal shocks and afterglow from external shock, arising from a relativistic jet produced by the collapse of a massive star. Internal shocks produce γ-rays and neutrinos, external shocks produce γ-rays, X-rays, optical and radio emission. Adapted from Mészáros (2002).

FIG. 23 Hubble Deep Field North, HST/WFPC2, December 1995. Credit: NASA, R. Williams (STScI) and the HDF Team. http://hubblesite.org/newscenter/archive/1996/01/

FIG. 24 Hubble Deep Field South, HST/WFPC2, October 1998. Credit: NASA, R. Williams (STScI) and the HDF Team. http://hubblesite.org/newscenter/archive/1998/41/

FIG. 25 Optical and near-infrared images of distant galaxies in the Hubble Deep Field North. For each galaxy, the left panel shows a composite of WFPC2 images, while the right panel shows a new-infrared view with NICMOS. Courtesy of Mark Dickinson.

FIG. 26 The star formation rate density versus redshift derived from the UV luminosity density. Adapted from Ferguson, Dickinson, and Williams (2000).

FIG. 27 The redshift evolution of the co-moving stellar mass density. The vertical extent of the boxes shows the range of systematic uncertainty. The bottom two solid lines (on the right-hand scale) show the result of integrating the star-formation-rate histories traced by the rest-frame UV light, with and without corrections for dust extinction. The top two solid lines and the dashed line show theoretical predictions from semi-analytical galaxy evolution models (Cole *et al.*, 2000; Somerville, Primak, and Faber, 2001). Adapted from Dickinson *et al.* (2003).

FIG. 28 Joint confidence intervals for (Ω_M, Ω_Λ) from SNe Ia. Regions representing specific cosmological scenarios are illustrated.

FIG. 29 Hubble diagram of SNe Ia minus an empty (i.e., $\Omega_M + \Omega_\Lambda = 0$) universe compared to various cosmological and astrophysical models. The points are redshift-binned data from Riess *et al.* (1998) and Perlmutter *et al.* (1999). The observations of SN 1997ff are inconsistent with monotonic evolutionary or dust effects that could mimic an accelerating universe. Adapted from Riess *et al.* (2001).