

SCIENCE WITH THE HUBBLE SPACE TELESCOPE IN THE NEXT DECADE: THE CASE FOR AN EXTENDED MISSION

S. M. Fall, R. J. Allen, S. V. W. Beckwith, R. A. Brown,
S. Casertano, M. E. Dickinson, R. L. Gilliland, J. E. Rhoads, A. G. Riess,
W. B. Sparks, and M. Stiavelli

1. Introduction

The Hubble Space Telescope (HST) has been the premier tool for astronomical research since its launch in 1990 and especially since the correction of spherical aberration in 1993. HST has brought the excitement of scientific discovery to the professional research community and to the public at large in a way that no previous scientific enterprise ever has. Demand for HST began high and has increased to the point that only one requested program in ten can now be executed. Discoveries and inspiring images have poured in in such numbers that it would be impossible to catalog them all. With each new servicing mission and instrument, HST has become a rejuvenated, more powerful telescope, and thus able to tackle more challenging and fundamental scientific programs. In all these respects, HST has exceeded the hopes of even the greatest optimists involved in its early planning and development.

The planned lifetime of HST—initially 15 years, now possibly a little longer—was set mainly by programmatic (budgetary) and technical (engineering) issues. Science played a less definite role. Scientific discovery, by its nature, is unpredictable, and no-one could be certain in advance how much HST would achieve. But we are certain now. We also have a pretty good idea what more could be achieved if the lifetime of HST were extended and its capabilities augmented with one or more new instruments. This report is a summary of some preliminary thinking along these lines. Our intent is to stimulate more thought and discussion of the scientific benefits of extending the HST mission. Some of the ideas presented here require further study before they can be regarded as definitive. But even at this stage, it appears that a compelling case can be made to extend the HST mission.

We are aware that some of the new scientific programs for HST outlined here might also be addressed by several future missions now under discussion. However, these are special-purpose missions, usually focused on only one or two major scientific objectives. Because HST is a general-purpose observatory, it is able to pursue a larger number of similar programs, and also has greater flexibility to respond to unanticipated future discoveries by itself and other observatories. HST is already on orbit and in operation, significant advantages

over missions now only in the planning stages. Furthermore, HST has a proven record of accomplishment as a tool for public outreach. Finally, an extended HST mission would ensure continuity of research in all branches of astronomy that require high-resolution optical and/or infrared observations until the launch of the James Webb Space Telescope (JWST) and the Terrestrial Planet Finder (TPF), now planned for *ca.* 2010 and 2015, respectively.

This report outlines several major scientific objectives as prime examples for an extended HST mission. Several other programs appear promising but have not yet been developed to the point that they can be included here. These merit further study before a definitive case is presented. Even so, we foresee that an extended HST mission may be the most cost-effective way to achieve some of the major scientific objectives of the NASA Origins Program.

2. Key Science Themes of the Next Decade

Four key themes that generate much of the excitement in astronomy today are dark energy, dark matter, young galaxies, and extrasolar planets. It is highly likely they will continue to dominate forefront research for the next decade or more.

The recently installed Advanced Camera for Surveys (ACS) and the planned Wide Field Camera 3 (WFC3) were designed largely around programs to study dark energy, dark matter, and young galaxies. Much of the work in these themes is inherently statistical and benefits from large samples of faint objects. The Great Observatories Origins Deep Survey (GOODS), a combination of HST Treasury and Space Infrared Telescope Facility (SIRTF) Legacy Programs, which began in late 2002, is a prime example of this type of survey. The HST part of the program will cover two regions of sky, each approximately $10' \times 16'$ (15 ACS fields), in four wavelength bands to a depth within 0.5–0.8 mag of that of the original Hubble Deep Fields (HDFs) in a total of 400 orbits. Another major HST program, the Ultra-Deep Field (UDF), scheduled to begin in early 2003, will probe a single ACS field to about 1.3 mag fainter than the HDFs, in about 400 orbits.

Several important scientific problems could be solved with ACS and WFC3 with even larger surveys (thousands of orbits). And the installation on HST of an Ultra-Wide Field Camera (UWFC), with a much larger field of view (FOV), would enable an additional suite of issues to be addressed.

ACS now has a coronagraph that is barely capable of detecting a planet like Jupiter in an orbit like that of Jupiter around the nearest star, α Cen (unlikely, though, since this star is multiple). The installation of a new coronagraph, with wavefront correction by a deformable mirror, would enable the direct detection of planets with these properties around

many nearby stars, out to distances approaching 10 pc, thus opening for exploration a volume 500 times greater than that accessible with the ACS coronagraph.

In the next two sections of this report, we present concrete examples of major scientific programs that could be accomplished with large amounts of time with existing and planned instruments (ACS and WFC3) and with new instruments (UWFC and a high-contrast coronagraph). In the remainder of this section, we present the scientific motivation for these programs.

With HST, the Hubble relation between redshift and apparent magnitude can be determined up to $z \approx 1.6$ using supernovae of Type Ia (SNe Ia) as standard candles. Recent studies of this kind with a combination of HST and ground-based telescopes have revealed that the expansion of the universe is now in an accelerating phase that began at $z \approx 0.7$. Expansion of this sort requires some form of vacuum or dark energy with negative pressure. The acceleration might be caused by the so-called cosmological constant Λ (Einstein’s “biggest blunder”) with a corresponding density parameter $\Omega_\Lambda \approx 0.7$. This result is so important that it requires confirmation and further study, including reduction of the statistical and any systematic errors. The Hubble relation for SNe Ia is consistent with measurements of the first Doppler peak in the cosmic microwave background radiation and of the present-day motions of galaxies, which indicate that the total density parameter is $\Omega_T \approx 1.0$, while that of matter, both luminous and dark, is $\Omega_M \approx 0.3$.

The cosmic acceleration has come as a complete surprise, with hardly any underpinnings from theoretical physics. It has become customary to describe it in terms of the parameter w in an assumed equation of state for the dark energy (X) of the form $p_X = w\rho_X$. If the universe were dominated by a cosmological constant, domain walls, or cosmic strings, it would have, respectively, $w = -1$, $-2/3$, or $-1/3$. A popular model for the dark energy is a decaying scalar field called quintessence, a modern cousin of the field that caused inflation. This type of model can be recognized by an equation-of-state parameter in the range $-1.0 \lesssim w \lesssim -0.5$, together with significant evolution, i.e., $dw/dz \neq 0$. In fact, the value of dw/dz is directly related to the scale length of the potential $V(\phi)$ of the field ϕ giving rise to the dark energy. A true cosmological constant Λ , instead, would have $dw/dz = 0$. Distinguishing between these possibilities is a major challenge for both astrophysics and particle physics in the next decade. This requires accurate measurements of many SNe Ia over a wide range of redshifts, $0.3 \lesssim z \lesssim 1.6$.

The evolution of the dark energy also determines the ultimate fate of the universe. A constant equation-of-state parameter w leads to continued acceleration and eventual loss of causal connection between galaxies. Evolution of w at different (non-zero) rates and with different signs can lead to future deceleration and collapse or unstable acceleration

that eventually tears apart progressively smaller bound systems, from clusters of galaxies to atomic nuclei.

The distribution of galaxies in space may be pictured as a “cosmic web” of clusters, voids, and filaments or superclusters. This distribution is often characterized statistically in terms of its low-order correlation functions or their Fourier transforms, the power spectra. These functions have now been measured very accurately at $z = 0$ and moderately accurately up to $z \approx 3$. It is widely believed that the cosmic web of large-scale structure results from the gravitational amplification of small density fluctuations in the early universe [and now visible as small temperature fluctuations in the cosmic microwave background (CMB) radiation]. However, the mass in galaxies ($\Omega_{\text{gal}} \approx 0.01$) is only a small fraction of the mass in all forms of matter ($\Omega_M \approx 0.3$) and the distribution of the former (luminous matter) may or may not trace that of the latter (mostly dark matter). Indeed, it is theoretically expected that galaxies form preferentially in regions of high density, an effect usually referred to as biasing, and that, on the scales of individual galaxies, non-gravitational effects, principally pressure gradients and radiative cooling in the gas, would strongly influence the distribution of luminous matter.

The best way to probe the distribution of dark matter is by gravitational lensing. The most striking manifestations of gravitational lensing are the multiple images of background sources seen in projection near objects of high surface density, such as the cores of galaxies and clusters of galaxies, an effect known as strong lensing. Objects of lower surface density, such as the outer parts of galaxies and clusters of galaxies and the large-scale distribution of matter itself, merely distort the images of background sources. This effect, known as weak lensing or cosmic shear, is revealed by the apparently coherent orientations of otherwise randomly oriented galaxies. Statistical measures of cosmic shear are directly related to statistical measures of the intervening distribution of dark matter. Observations of cosmic shear are still in their infancy and have yet to probe the dark matter distribution on large scales or at high redshifts and thus to test directly for the growth of structure by gravitational instability. Since the point spread function (PSF) of HST is much sharper and more stable than that of ground-based telescopes, it can measure the shapes and orientations of smaller and fainter galaxies. This allows weak lensing to be measured to higher redshifts and, if the surveys are large enough, with more accuracy by HST than from the ground.

Galaxies are often regarded as the basic building blocks of the universe and an understanding of how they formed as one of the most important and challenging themes in astrophysics. It is generally believed that the formation of galaxies involves the gravitational clustering and merging of dark matter halos in a hierarchical, bottom-up sequence—the small-scale, nonlinear part of the gravitational instability described above. The hierarchical

picture is at least qualitatively consistent with HST observations, especially those in the HDFs, which indicate that galaxies were smaller, more numerous, and interacted more often in the past (at $z \gtrsim 2$) than at present. The other processes thought to be important in galaxy formation, interstellar and radiative processes and especially star formation, are less well understood. This makes it difficult to predict reliably the observable properties of young galaxies, and progress in this field therefore requires a close interplay between theory and observation. From the HDFs, we already have a glimpse of the evolution of galaxies at $z \lesssim 3$, and from GOODS and other near-term surveys, this may be extended to $z \lesssim 6$. JWST will be powerful enough to explore the pregalactic era at $z \gtrsim 10$. The crucial intervening period, $6 \lesssim z \lesssim 10$, is likely to be the epoch of the first galaxies and active galactic nuclei (AGN) and may be accessible to HST with ultra-deep and/or ultra-wide field surveys (depending on the luminosities and space densities of the sources, which are uncertain for the reasons mentioned above).

The period $6 \lesssim z \lesssim 10$ (possibly $6 \lesssim z \lesssim 15$; see below) is special in cosmic history for another reason; it is the likely epoch of reionization, probably the most significant event since the (re)combination of the primordial H+He plasma at $z \sim 1000$. Following recombination, the universe became, in effect, a giant, cooling HI region, transparent below 1 Rydberg but opaque above it. Today, the universe is once again mostly ionized, a giant HII region, with a temperature above 10^4 K and is relatively transparent at all wavelengths. The transition between these phases of the intergalactic medium (IGM) required a large output of ionizing radiation, from young stars and/or AGN in galaxies and/or pregalactic objects. The first sources would create isolated HII regions, but as the number and luminosities of the sources increased, the HII regions would begin to overlap until eventually they were all connected and reionization was complete. It might be said that the universe then experienced a “renaissance” following its “dark ages.” Reionization is also expected to inhibit or impede the formation of small galaxies, because photoionized gas is too warm to collapse and form stars.

Recent observations of the most distant quasars with the Keck Telescope indicate that reionization ended by $z \approx 6$. More recent measurements of the polarization of the CMB with the Wilkinson Microwave Anisotropy Probe (WMAP) indicate that reionization may have begun much earlier, at $z \approx 15$ (albeit with a large uncertainty). The high temperature of the IGM measured in quasar absorption line systems suggests that much of the heating due to reionization occurred at $z \ll 15$. (Earlier heating of the IGM would be counteracted by adiabatic cooling as the universe expands.) Some theoretical models suggest that reionization may even have occurred twice, for example at $z \approx 15$ and then again at $z \approx 7$, but this possibility depends on several unknown factors. To resolve these issues, it is crucial to determine which types of sources dominate the production of ionizing photons (whether they were galaxies or AGN, how luminous they were, etc.) and to explore the effects of reionization

on the IGM and the formation and evolution of galaxies. This requires a combination of deep and wide surveys from space in the wavelength range 0.8–1.4 μm . These wavelengths include the crucial Ly α emission line at redshifts $6 \lesssim z \lesssim 10$. Indeed, the three-dimensional structure of the reionization surface (the interface between HI and HII regions) can be mapped by measuring the strengths of Ly α emission and absorption in low-resolution spectra of many faint sources as functions of their redshifts and positions on the sky.

Another major theme of astronomy over the next decade will be the discovery and study of planets around other stars. We now know of the existence of about a hundred of these extrasolar planets from the Doppler shifts of their parent stars, but we have not yet seen the reflected light or thermal emission from any of the planets themselves. The ultimate goals of this line of exploration are to understand our solar system in its astrophysical context, to study the atmospheres and surfaces of new planets, and to seek environments compatible with the presence of life—if not signs of life itself.

TPF is the NASA mission planned specifically with direct imaging of extrasolar planets in mind. NASA is considering optical coronagraphy on an HST-like telescope as one approach to TPF. A high-contrast coronagraph on HST would take important steps in this direction long before TPF is launched. Indeed, such an instrument, CODEX, was proposed in 1997 and rated “selectable” by the NASA peer review process. In the following, we use CODEX to illustrate the capabilities of a high-contrast coronagraph on HST.

The statistics of short-period microlensing events in rich star fields will provide a census of planets of various masses, including Earth-mass planets. Another technique for finding planets relies on the fact that, in rare cases, when the orbital planes are almost exactly parallel to our lines of sight, planets will periodically occult their parent stars. The differences between the spectra of the stars during and outside these transit events then provides information about the atmospheres of the planets (thickness, composition, etc.). This technique has already been applied successfully to a giant planet orbiting the star HD 209458b using the Space Telescope Imaging Spectrograph (STIS). As more transiting planets are discovered, it will be possible to obtain similar information about their atmospheres.

The most direct, but also the most difficult, method of detecting extrasolar planets is by their reflected starlight. HST, with the existing coronagraph on ACS, may just be able to obtain images of gas giants around a few nearby stars, thanks to a new method described later in this report, although this will require a favorable combination of circumstances. But recent advances in deformable mirror technology for correcting wavefront errors and in pupil shapes and masks for suppressing diffracted starlight allow the construction of a new coronagraph with performance gains of 3–4 orders of magnitude. (This design was the basis of the proposed CODEX instrument.) If HST were equipped with such a coronagraph,

it could find and study twins of Earth around the nearest stars, planets like Uranus and Neptune on Mars/Jupiter-like orbits around stars to 5 pc, as well as planets like Jupiter and Saturn to 10 pc. In addition, other astrophysical studies that require imaging with high dynamic range, such as quasar feeding zones and protoplanetary nebulae, would be enabled with an advanced coronagraph.

Even crude spectra or multicolor photometry with a CODEX-like instrument would provide some information about the atmosphere of a planet, whether it was more like those in the inner or outer parts of our solar system. Under very favorable conditions, it might even be possible to detect some of the biomarkers in the atmosphere of a planet, such as the spectral features of H₂O and/or CH₄. Temporal variations in the brightness of the planet would provide some information about rotation periods, major surface features, and even meteorological activity.

3. Science with Current and Planned Instruments

In this section, we present examples of major scientific programs that could be carried out by HST with its current and planned instruments if its lifetime were extended by several years.

3.1. Dark Energy

Measuring and understanding the dark energy may be the greatest challenges remaining at the crossroads of particle physics and astrophysics. Observations of Type Ia supernovae over the redshift range $0.3 \lesssim z \lesssim 1.6$ are the best means we have to determine the nature of the dark energy and its evolution. The varieties of physics that may explain the dark energy (X) are often diagnosed in terms of the equation-of-state parameter $w = p_X/\rho_X$. The expansion of the universe is governed by the pressures p_i and densities ρ_i of all its constituents; the cosmic scale factor a evolves according to $\ddot{a}/a \propto -\sum_i(3p_i + \rho_i)$. Thus, for $z \ll 1000$, the acceleration or deceleration is determined mainly by w and the density parameters of the dark energy and dark matter, Ω_X and Ω_M (since p_M is negligible). The Hubble diagram for SNe Ia constrains w because the observed brightness of a supernova at a redshift z depends on the expansion history of the universe between z and now. From existing data, the allowed range is $-1.5 < w < -0.7$.

GOODS is the first HST program designed to discover and measure high-redshift SNe Ia for studies of the dark energy. In this program, ACS has made 100 repeat pointings and

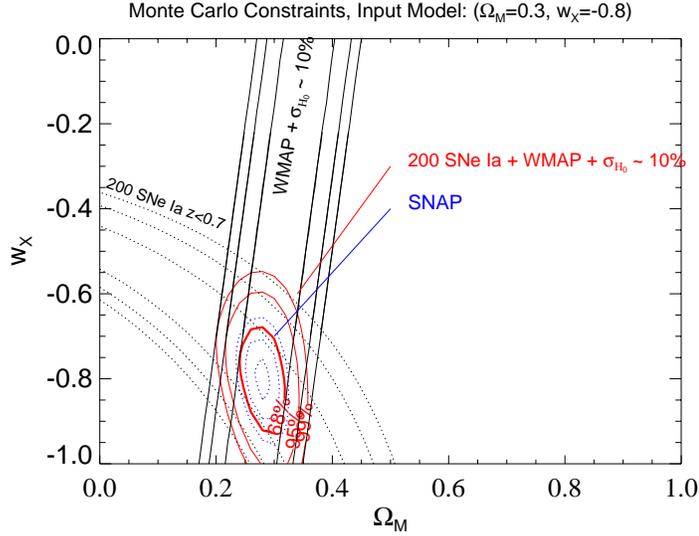


Fig. 3.1-1. Confidence contours (1, 2, and 3σ enclosed) for the equation-of-state parameter w_X and the matter-density parameter $\Omega_M = 1 - \Omega_X$ with the 200 SNe Ia expected from the program described in §3.1. The dotted contours show the constraints based on the supernovae alone and the bold contours for SN plus WMAP in the present. We assume a redshift limit of $z < 0.7$.

detected 35 new supernovae, half of which have been confirmed as Type Ia. The redshift range of the SNe Ia is $0.3 < z < 1.8$, with a median value $z \approx 1.0$. Based on this rate, one SN Ia, on average, is found in every five ACS pointings. This means searches for supernovae with ACS are relatively inefficient (80% overhead); it would take about a year of observation to collect a sample of 200 SNe Ia. However, as with GOODS, it is possible to combine the discovery of SNe Ia with other large-area ACS surveys. For example, the 3600-orbit weak-lensing program described in the next subsection (§3.2) is expected to find ~ 300 SNe Ia with optimal timing of the observations.

Maintaining the statistical power of such a large sample of SNe Ia requires controlling systematic errors to a few percent per redshift bin of $\delta z = 0.1$, a difficult task from the ground. HST has always been and still remains the most precise and reliable platform from which to perform such measurements. With its well-characterized PSF, transmission curve, zeropoints, and ability to resolve supernovae from their host galaxies, HST is unmatched in its ability to measure distant SNe Ia. In addition, only HST can reliably discover and measure SNe Ia at $z \gtrsim 1$.

The first step in determining the nature of the dark energy is to constrain its time-averaged equation of state. To reach this goal, it is sufficient to observe SNe Ia at relatively low redshifts ($z \lesssim 1$) and to rely on WMAP for complementary constraints at higher redshifts.

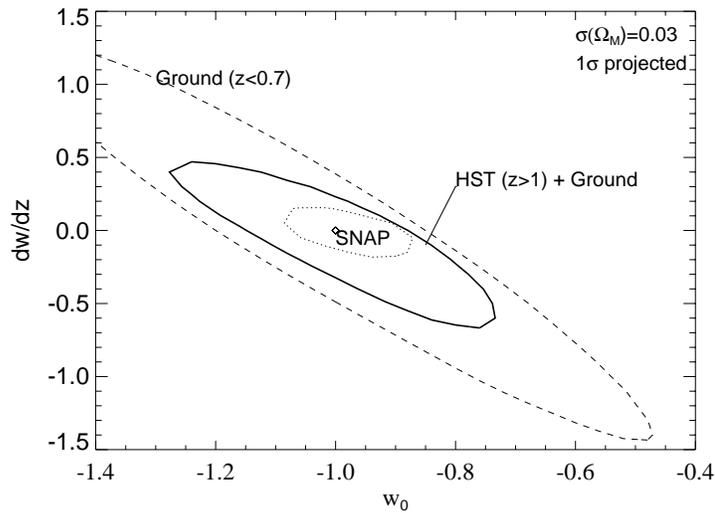


Fig. 3.1-2. Confidence contours with 39% probability enclosed (1σ projected on each axis) for the equation-of-state parameter w_0 and its derivative dw/dz at $z = 0$ for three different dark-energy programs. The Ground program is modeled after the ESSENCE and CFHT programs and includes 200 SNe Ia at $0.3 < z < 0.7$. The HST+Ground program includes the previous sample and 200 SNe Ia at $1.0 < z < 1.6$ collected in 1–2 years with HST. The SNAP program includes 2000 SNe Ia at $0.3 < z < 1.6$. Each program assumes knowledge of Ω_M to 10% (or equivalently, the WMAP constraints) and irreducible systematic uncertainties appropriate for either ground- or spaced-based observations as defined by the SNAP collaboration.

Monte Carlo simulations suggest that we can constrain w to better than 10% with HST observations of 200 SNe Ia at $0.3 < z < 0.7$ when combined with measurements from WMAP. Figure 3.1-1 shows the results of these simulations. A survey of this quality could begin to determine whether the acceleration of the universe is driven by a cosmological constant ($w = -1$) or by some other form of dark energy ($w \neq -1$), a discovery that might point the way toward new physics. In principle, ground-based programs, such as the approved Project ESSENCE and CFHT Legacy Survey, can also provide these SNe Ia, but the systematic uncertainties are better controlled in space.

A more challenging, but potentially more powerful, test of the dark energy is based on the evolution (or lack of it) of the equation of state. As discussed in §2, the two leading models for the dark energy, a cosmological constant and quintessence, differ both in the value of w and in its derivative dw/dz . Tight constraints on these parameters near $w = -1$ and $dw/dz = 0$ would be strong evidence for a cosmological constant (and strong motivation to the theoretical physics community to explain it.) In contrast, quintessence and other models (e.g., supergravity, string theory, hidden dimensions, time variation of physical constants,

etc.) have dark energies with evolving equations of state (with $|dw/dz| \sim 1$ in some cases). A non-zero measurement of dw/dz would be evidence for some of these models and would rule out a cosmological constant even if $w = -1$ were found at some redshift.

To constrain both the equation of state and its evolution, it is crucial to sample SNe Ia over a wide range of redshifts. A first, but significant, probe of the evolution of w could come from a combination of HST-unique and ground-based observations. As an example of such a pilot program, we assume that 200 SNe Ia at $1.0 < z < 1.6$ are observed with HST (as in the programs discussed above) and that 200 SNe Ia at $z < 0.7$ are observed from the ground (as in the ESSENCE and CFHT programs). The results are shown in Figure 3.1-2. Evidently, the combined HST+Ground program provides much better constraints on the joint estimates of w and dw/dz than the ground-based program alone, although neither of these competes with the expected results from the Supernova Acceleration Probe (SNAP). The HST+Ground program could, however, make an early discovery of a large variation in the equation of state and thus provide an important clue in our quest to understand the dark energy and the fate of the universe.

3.2. Dark Matter

HST will uniquely study the development of large-scale structure in response to dark matter gravity from $z \sim 3$ to 0 using weak lensing of background galaxies.

An observing program with HST (ACS+WFC3) capable of measuring the clustering of dark matter on scales up to $30'$, and of following its evolution in redshift over the range $0 \lesssim z \lesssim 3$, will require a total of 3600 orbits (4500 with ACS only), and cover the spectral range from U to z (B to z with ACS only). The program will observe 1 square degree at the same effective depth as the GOODS program, $I_{AB} = 26.5$.

The data thus collected will yield an accurate measurement of lensing power on scales ranging from a few arcsec to $30'$, with $S/N \sim 5$ to 15 depending on scale, in four independent redshift intervals (see Figure 3.2-1). The power measured as a function of redshift and angular scale probes the growth of structure on different scales from $z = 3$ to the present, and discriminates between different cosmological models.

Unlike purely geometry-based cosmological tests, such as the luminosity distance-redshift relation probed by SNe Ia, cosmic shear measurements directly probe physical processes that are responsible for the growth of structure. For reference, a 7 Mpc cube in the present universe radiates on average $L^* = 2 \times 10^{10} L_{\odot}$, equivalent to a single bright galaxy. A comoving cube of this size subtends $7'$ at $z = 1$ and $4'$ at $z = 3$. Thus, shear measurements at angles

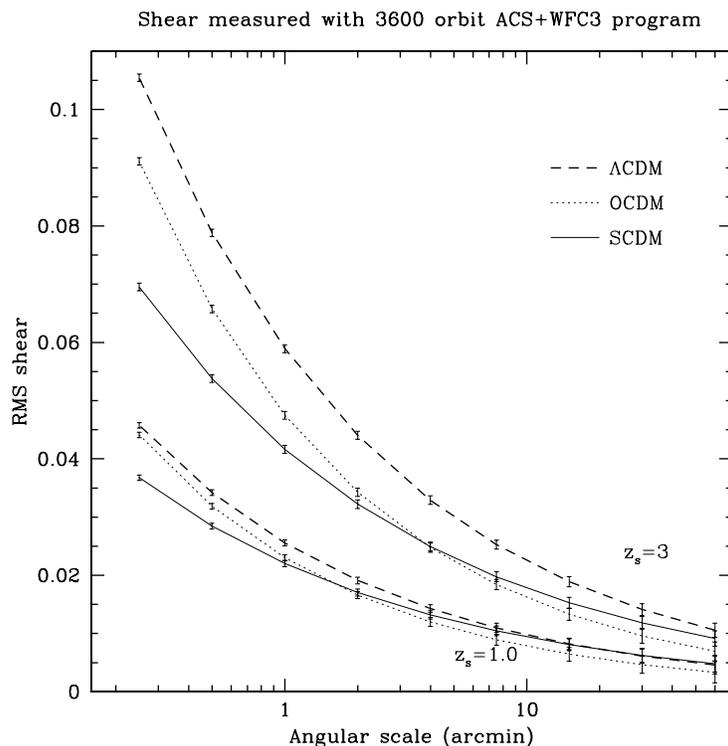


Fig. 3.2-1. Expected shear measurements in a 3600-orbit weak-lensing program covering 1 square degree with ACS and WFC3. The error bars indicate the expected 1σ uncertainties in the RMS shear on the given scale, for a standard (Λ) cosmology, an open cosmology, and a flat cosmology, respectively. Each set of curves pertains to one quarter of the sample, centered at the indicated redshift.

$\gtrsim 10'$ are required to probe the scales on which the growth of structure is in the linear regime and driven by gravity alone. Non-linear and non-gravitational effects are expected to become significant at smaller angles, corresponding to the precollapse scales of individual galaxies. Measurements at angles $\lesssim 1'$ will probe the postcollapse structure of galactic halos and follow their evolution from $z = 3$ to the present. These observations will also reveal any dark halos not associated with luminous material, an indication of “failed” galaxies.

Such measurements are uniquely enabled by space-based observations. Ground-based observations have been quite successful in sampling cosmic shear on scales from $1'$ to $10'$, but only for galaxies at $z < 1$. On larger scales, the signature of cosmic shear is small enough to compete with (so far) irreducible systematics from the ground; on smaller scales or at higher redshifts, too few galaxies are accessible from the ground for reliable measurements. Space-based measurements of cosmic shear already exist, although their impact has been thus far limited by the relatively small area covered; with a deep, extended, multiband survey taking

about one year of ACS+WFC3 time, space-based data will probe regions of parameter space that will remain otherwise inaccessible for the foreseeable future.

The required observations cover 1 square degree, corresponding to approximately 360 ACS pointings or 490 WFC3 pointings, for a total of 3600 orbits. The tiling pattern will be chosen carefully to ensure that both ACS and WFC3 cover the full area efficiently, despite their different sizes, and will result in about 10 orbits per pointing with ACS and about 7.3 orbits per pointing with WFC3; below we assume 7 orbits per pointing with WFC3 to account for slight inefficiencies of the tiling pattern. Of the 10 ACS orbits per pointing, 5 will be allocated to F850LP (z), 2.5 to F775W (i), and 2.5 to F606W (V), giving limiting magnitudes of 26.9, 26.8, and 27.5, respectively. Of the 7 WFC3 orbits per pointing, 4.5 will be allocated to F300X (U) and 2.5 to F475X (B), giving limiting magnitudes of 25.9 and 26.5, respectively. With appropriate timing, the z -band observations will also find the supernovae required by the Dark Energy project (§3.1), effectively searching the equivalent of 1440 ACS fields (360 fields times four search epochs) and yielding ~ 300 SNe Ia.

3.3. Young Galaxies

Reionization was the last major phase transition for most of the baryonic matter in the universe. Polarization measurements of the CMB by WMAP indicate that substantial ionization had begun as early as $z \sim 15$. On the other hand, spectra of $z > 6$ quasars show opaque Gunn-Peterson troughs, indicating that reionization was not complete until $z \approx 6$. The high temperature of the IGM at $z \sim 4$ requires considerable heating at $z \ll 15$, also suggesting that a large part of the reionization was relatively recent.

Beyond these constraints, we know very little about the reionization process. Key unresolved questions include: Were the ionizing photons primarily produced by young stars in galaxies, or by accretion onto black holes in early AGN? How luminous were these sources, and how long did they last? What was the history of reionization—was it a single event, or did it consist of several episodes, driven by physically distinct generations of objects in the early universe?

One fundamental issue that could be addressed with an extended HST mission is the nature of the ionizing sources, at least during the later phases of reionization, at $6 \lesssim z \lesssim 10$. A handful of galaxies and quasars at $z \approx 6$ are now being discovered, but current samples are woefully inadequate to characterize their properties in a statistically meaningful way. Galaxies at $z > 6$ are expected (and indeed, are found) to be very faint, and the bulk of the population is too faint for detection by ground-based telescopes (see Figure 3.3-1). The

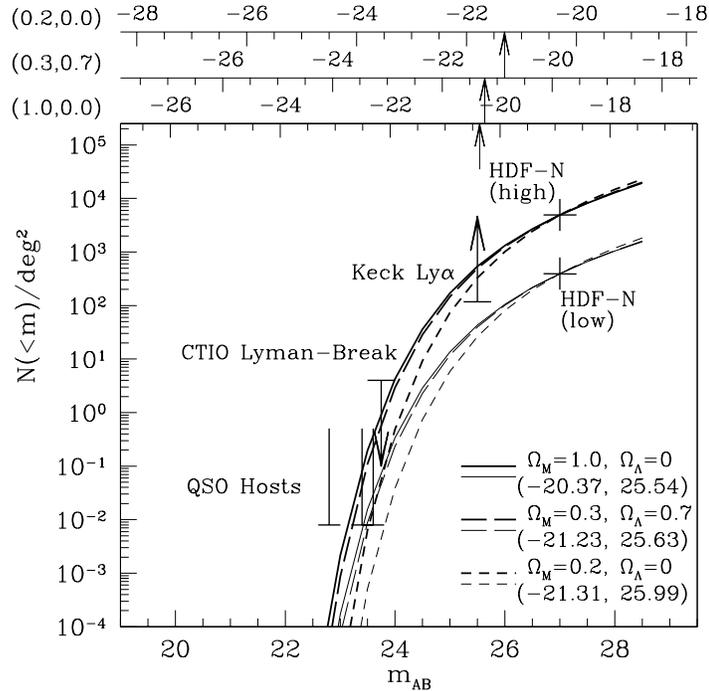


Fig. 3.3-1. Expected cumulative surface density of galaxies at $5.5 \leq z \leq 6.5$. The results for a flat universe with a cosmological constant are shown as long-dashed lines. Those for a flat universe without a cosmological constant and an open universe are shown as solid and short-dashed lines, respectively. The various symbols highlight different surveys. Two thick and thin sets of lines correspond to high and low normalizations, respectively. The absolute magnitude scales are given in the top bars for the three cosmologies. In the low-density model, the surface density of objects brighter than 26 is as low as 100 per square degree, i.e., one per three ACS fields. (From Yan et al. 2002, ApJ 580, 725)

emerging light shifts to near-infrared wavelengths, where HST has a powerful advantage thanks to the dark sky background and high angular resolution, which helps when observing these compact objects.

The most ambitious HST observing programs yet to image distant galaxies are GOODS and the UDF, with ~ 400 orbits each. Like the HDFs, both ACS programs survey the distant universe in four filters, providing photometric guidance to galaxy redshifts and important information about stellar populations. GOODS covers an area ≈ 60 times larger than the HDF-North with roughly half its sensitivity for extended galaxies. The UDF observes a single ACS pointing, but goes about six times deeper than GOODS.

Perhaps the most basic question relevant to reionization is how much UV light was

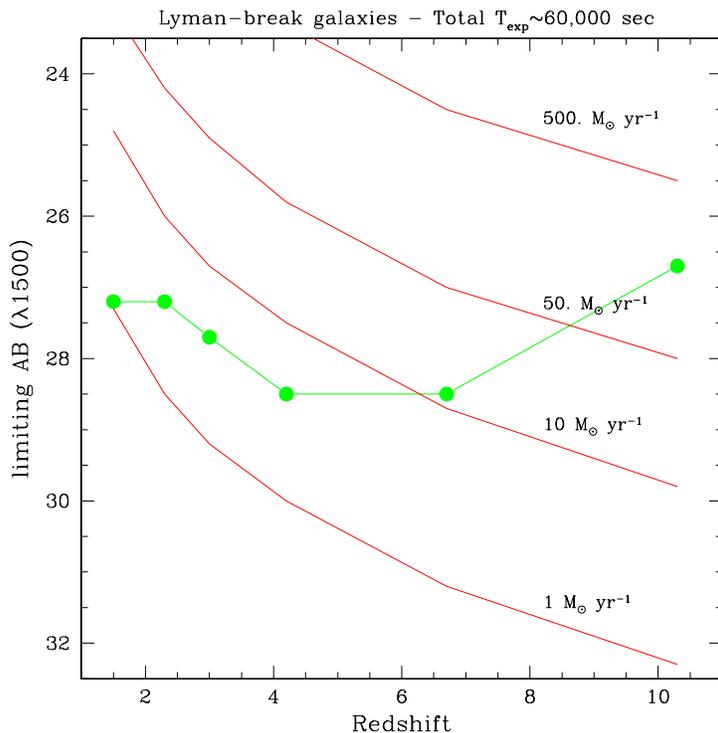


Fig. 3.3-2. Sensitivity of a combined ACS and WFC3 program with 25 orbits per filter for the detection of Lyman-break galaxies. Such a survey could detect galaxies forming stars at $10 M_{\odot} \text{ yr}^{-1}$ at $z = 7$.

produced by galaxies at $z > 6$. Early results from GOODS suggest that galaxies at $z \approx 6$ were less luminous than their descendants at $z \approx 3$, and hence that only a few of the brightest objects will be identified down to the GOODS flux limit ($m_z \approx 26.7$). Measuring the total UV emission requires observations of galaxies fainter than L^* , the characteristic bend in their luminosity function. The UDF may be deep enough to probe these faint luminosities, but only over a small area and a single sightline. Because of this, it will detect only a few dozen typical objects. Both theory and observations show that the most luminous young galaxies should be highly clustered, and a single UDF may not provide a representative sample.

An ambitious program combining the capabilities of ACS and the WFC3 IR channel would allow robust measurements of $z > 6$ luminosity functions (see Figures 3.3-2 and 3.3-3), thus establishing the most likely sources of reionizing photons. By combining red and near-IR data, this survey could provide useful constraints to at least $z \approx 10$. Such a survey will need the depth to measure sources two magnitudes below L^* , in order to constrain the faint-end slope of the luminosity function. At the same time, it will need to span several lines of sight in order to measure the rarest and brightest sources and to overcome uncertainties

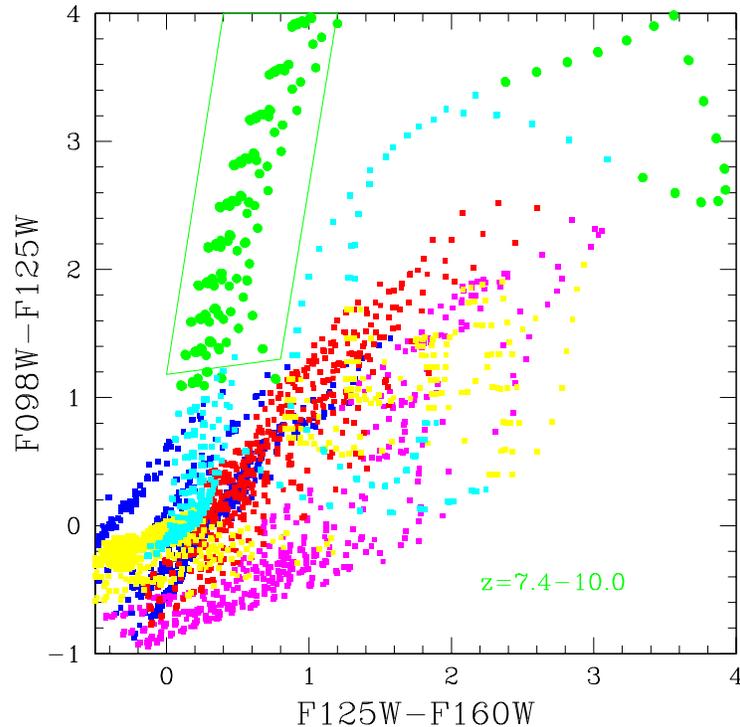


Fig. 3.3-3. Example of a Lyman-break selection criterion in the WFC3-IR color-color diagram ($F098W - F125W$ vs $F125W - F160W$) for galaxies at $z \geq 7.4$ (large green circles). The other symbols represent objects at lower redshifts. Objects in the area outlined by the green line have a high probability of being at high redshifts.

introduced by cosmic variance in the source density. Finally, it should combine broad-band imaging (which maximizes sensitivity to starlight) with slitless grism spectroscopy (which identifies emission lines from both star forming regions and AGN) in order to constrain the luminosity functions of Lyman-break galaxies, less luminous emission-line galaxies, and quasars.

A baseline survey strategy would go to a depth of 50 orbits each in the i , z , J , and H filters. For $z \approx 6$ sources, this will achieve a continuum sensitivity competitive with the UDF, since the three reddest filters can be combined. The current best guess (but only a guess) for the magnitudes of typical (L^*) Lyman-break galaxies at $6 \lesssim z \lesssim 8$, based on GOODS data, is m_z or $m_J \approx 27$ to 28. Exposure times of ~ 50 orbits with ACS and WFC3 should be ample to identify L^* galaxies out to $z \gtrsim 8$ and to measure the faint-end slope of the luminosity function of Lyman-break galaxies at $z \approx 6$. AGN point sources could be identified to still lower luminosities and higher redshifts. A solid angle coverage of ~ 220 square arcminutes, divided among at least five widely separated fields, would provide good

statistics of the brightest sources at all redshifts, together with robustness to cosmic variance. This corresponds to 20 ACS fields and 50 WFC3-IR fields. Thus, the imaging portion of such a survey would total 7000 orbits, although this could be reduced to 5000 orbits with a tiling scheme that uses ACS and WFC3-IR in parallel.

A survey like the one outlined above, if carefully phased, would yield significant byproducts, in particular a wealth of high-redshift supernovae of both Types I and II up to $z \approx 3$. The detection of Type II SNe would allow for a derivation of the star formation history of the universe independent of the UV luminosity density and unbiased with respect to surface brightness selection effects.

An extended HST mission could also be vital in testing the pace of reionization. The first direct observations of the reionization process have come from the Gunn-Peterson trough, i.e., Ly α absorption by neutral gas in the IGM. Since Ly α is such a strong transition, a Gunn-Peterson trough will be produced by even a small neutral fraction: $x \geq 10^{-4}$ for a hypothetical, homogeneous IGM, and $x \geq 1\%$ for a realistic, clumpy IGM. However, the most rapid and interesting part of the reionization process is the overlap phase, when the Stromgren spheres around individual ionizing sources merge into a volume-filling ionized medium. This transition occurs at a much higher neutral fraction, $x \sim 10\text{--}50\%$.

Extensive grism observations with ACS and WFC3-IR can provide a complementary test of reionization that is applicable at these larger neutral fractions. Resonant scattering of Ly α photons renders low-luminosity Ly α lines effectively invisible for $x \gtrsim 10\%$. Ly α source counts will therefore show a sharp drop around this neutral fraction. This drop in Ly α source counts provides a robust, direct way of determining the reionization redshift. This test can be applied at $6.5 \lesssim z \lesssim 8$ with about 80 orbits per field with the G102 grism on WFC3-IR. For $z < 6.5$, the ACS grism will work equally well with a comparable integration time. A total of 1000 orbits with G102 and about 500 with ACS+G800L will provide a sample of ~ 50 Ly α emitters per unit redshift from $z = 6$ to $z = 8.5$ in the absence of neutral hydrogen scattering. This is sufficient to detect the drop in Ly α counts induced by neutral hydrogen at the 5σ level, providing a new and robust probe of the reionization process.

3.4. Extrasolar Planets

Even with the current and planned suite of instruments, HST offers unique capabilities for contributions to the emergent study of extrasolar planets (albeit quite limited in comparison to what could be done with a new coronagraph—see §4.4). The existing ACS coronagraph, in conjunction with an innovative observing technique, holds the promise of

making the first detections of extrasolar planets by direct imaging. Success here would provide not only science results of the first rank, but also crucial guidance for future instrument and mission designs. STIS has already detected two atmospheric constituents on HD 209458b, the first extrasolar planet found to transit its host star, and more detections on this and other stars are likely within reach with this technique. Again, these results are of inherent science interest and also provide experience that will be valuable in planning future missions to explore extrasolar planets.

3.4.1. Direct Searches with the ACS Coronagraph

The image quality of HST with the recently installed ACS is now sufficient to attempt the detection of Jupiter-type planets around some of the nearest stars. This is a crucial stage in our progression towards locating and identifying terrestrial planets around other stars. The region of parameter space open to an ACS search requires that the systems be close analogues of our solar system. This makes them exciting targets for TPF follow-up, and, critically, allows us for the first time to validate our ability to image and characterize extrasolar planets.

The High Resolution Camera (HRC), one of the three cameras that comprise ACS, offers high sensitivity across the entire optical and near-UV wavelength range. The FOV is $26'' \times 29''$, and the images are essentially diffraction limited in angular resolution (typically 50 mas). Not only does the camera offer very high angular resolution, but it also contains a module known as the Aberrated Beam Coronagraph. This is a true Lyot coronagraph, differing from a conventional implementation only in that the focal plane masks lie in the spherically aberrated focal plane of HST. This results in (1) rather large coronagraphic focal plane spots, 1.8 and 3'' in diameter, and (2) some additional light leakage from the occulted star due to the extended PSF wings. Downstream optics correct for the spherical aberration in the usual way, and the resulting suppression of the PSF scattered-light halo is about a factor of 10 better than the already excellent PSF profile of HRC without the coronagraph in the beam.

The problem of imaging a Jupiter-type extrasolar planet is that it is 10^{-9} times the brightness of the star around which it orbits. At 1 pc distance, the Jovian orbit would have an apparent semi-major axis of 5''; at 5 pc, it would be only 1''. The observational difficulty is dominated by stray light from the bright star, arising from diffraction and scattering in the HST optics.

Conservatively, coronagraphic observations are made of a star with a presumed faint

companion by first observing the target, and then observing either a reference star for improved PSF subtraction, or else by re-observing the target with the telescope at a different roll angle. In this way, residual structure in the stray light from the star can be effectively subtracted. The limiting factor in this approach is that, without adaptive optics, small changes in telescope focus (“breathing”) and in centering the star behind the coronagraphic spot cause small changes in the residual stray light of the bright star. The uncertainties in the HST observations are quickly dominated not by photon statistics, but by these systematic errors. We calculate that, to order of magnitude, within a reasonably short observing time, the ACS coronagraph can see companions approximately 10^{-8} times the brightness of the star, but without significant gain from increased observing time.

[A dedicated coronagraph like CODEX could (1) avoid the necessity to work in the aberrated beam, hence work with smaller focal plane spots, (2) do a better job of correcting for mid-frequency surface errors and hence decrease stray light dramatically, and (3) correct for focus changes through an observation and reduce time-dependent limitations.]

Fortuitously, in the search for point-source companions like extrasolar planets, this is not the end of the story. If, instead of observing a target and a reference, we observe the target at a series of different wavelengths using the ACS tunable (ramp) filters, we can turn the wavelength dependence of the complex coronagraphic PSF to advantage. By including this spectral dimension in an otherwise conventional coronagraphic imaging observation, and building on the fact that all the residual structure has angular scale proportional to wavelength while the extrasolar planet remains fixed in the image plane, we may eliminate the stray light of the star while retaining the light of the extrasolar planet. This method takes us into a domain in which we are limited only by the number of photons collected. That is, the signal is that of the extrasolar planet, and the noise is simply the photon or Poisson noise of the stray-light halo (as opposed to its residual structure). These concepts are illustrated in Figure 3.4-1.

This observing technique has the additional advantage that a low-resolution spectrum of the companion is acquired as part of the process. We may hope to detect molecular absorption troughs, atmospheric scattering, or eventually, such important features as the chlorophyll edge that may indicate the presence of life.

Since we must use a narrow tunable filter to implement this technique, the photon-collection rate is somewhat reduced. Nevertheless, we calculate that with a modest observing time (of order 20 orbits per star) we can see companions that are a few times 10^{-9} of the host star. For the case of α Cen A and B, we can observe all the way to 10^{-9} or slightly better, and hence observe a true Jovian analogue if one is present. For Sirius, Procyon and Altair, we are within a factor of 2 to 20 of seeing a Jupiter-type planet in a modest amount of

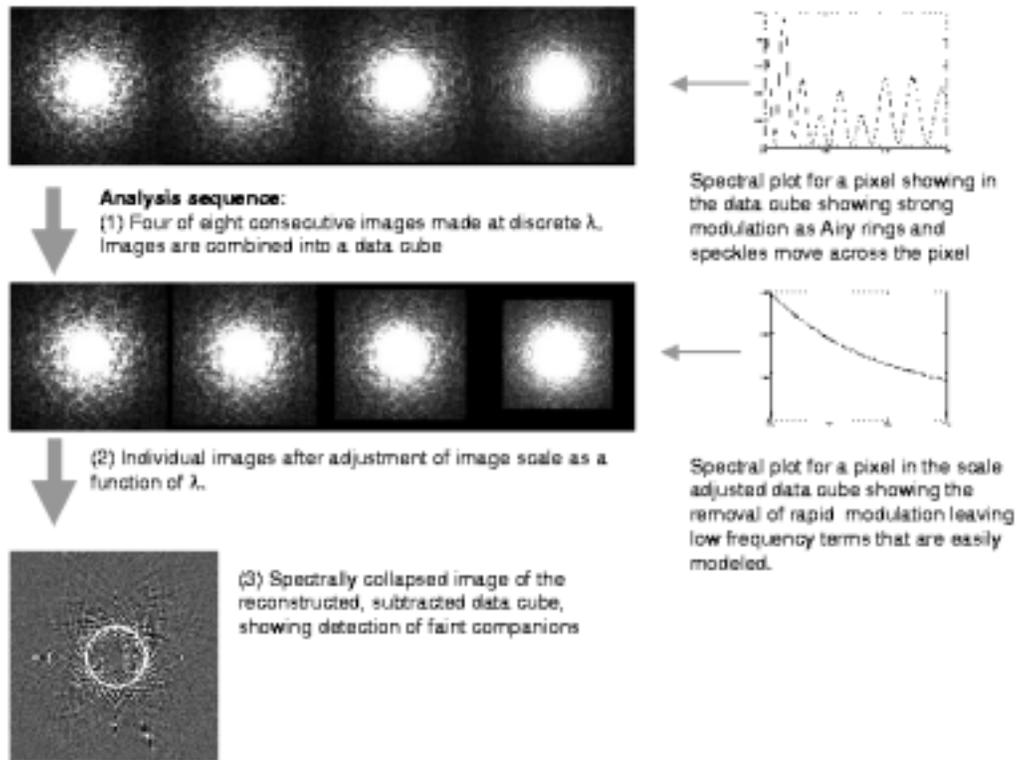


Fig. 3.4-1. Illustration of the spectral deconvolution approach to coronagraphy. The top panel shows observations at four different wavelengths, the middle panel the same after adjusting to the same image scale, and the bottom panel the collapsed image revealing faint companions.

observing time. At this level, details such as ring systems (like that of Saturn), albedo, phase functions, and simply size, can bring an extrasolar planet into the realm of detectability.

Since this technique gives us improved detectability with increased exposure time, in an extended HST mission, one might devote several hundred orbits to a single star. In such circumstances, we may already, with existing instrumentation and observing techniques, hope to image and characterize extrasolar planets similar to those of our solar system around seven stars: τ Cet, ϵ Eri, and those already mentioned. An additional 10 or so stars would be within a factor of a few of Jovian detectability, and a large number (approximately 25) would be within a factor of 10 of Jovian detectability. And of course, with similar, long observations of the most favorable stars (α Cen in particular), we may even approach terrestrial planet detectability.

This technique is new, and in the process of validation, but it has the promise to move HST to the forefront of direct imaging searches for extrasolar planets until dedicated missions such as TPF become a reality.

3.4.2. Atmospheric Spectroscopy During Transits

The observations of HD 209458b with STIS in April and May 2000 were the first probes of an extrasolar planetary atmosphere, a remarkable achievement for an instrument not designed with this science in mind. By accumulating ultra-high S/N spectra on this bright, $V = 7.6$ mag star within the 3-hour planet transit (repeating every 3.52 days), and comparing them to spectra outside of transit, it was possible to isolate contributions from the planetary atmosphere in the resonance doublet of sodium near 5890 Å. It seems likely that follow-up work planned in Cycle 11, covering a broader wavelength domain, will return information on other atmospheric species (perhaps including water) and further stimulate the emerging studies of extrasolar planetary atmospheres. Test observations with the Near-Infrared Camera and Multi-Object Spectrograph (NICMOS) were obtained in December 2002 to determine high- S/N capabilities in the IR; a similar test with STIS in 1999 served as the enabling calibration for the HD 209458 observations. The NICMOS test, while not definitive, suggests that nearly Poisson-limited results can be expected in the IR as well, which should enable secure detection of water and other species with strong IR bands. A considerable success is already in hand with HD 209458b observed with STIS, and extensions of the technique are being developed with continuing observations. It seems likely that around 40 orbits dedicated to HD 209458b (with the benefit of hindsight only now being developed) would suffice for an extensive probe of its atmosphere, quantifying abundances of several atomic and molecular species, setting limits on the height of cloud decks, and perhaps measuring the albedo as a function of wavelength, thus determining the energy input to the planet.

What opportunities will arise in the remaining years of HST for studies of extrasolar planetary atmospheres? Clearly, we need to observe more than just one planet. HD 209458 may well be the brightest star with a transiting gas giant, but it seems likely that a few more examples brighter than $V = 10$ and with periods of 3–7 days will be found in the next few years. Observations of slightly fainter stars hosting transiting planets will not be as time consuming as might at first be surmised. At $V = 7.6$, HD 209458 is very bright for instruments designed to study faint, fuzzy things near the edge of the visible universe; in particular, readout overheads take over half the available visibility time with STIS, and over 90% with NICMOS. As fainter targets are observed, the on-target efficiency will be regained. Thus at $V = 10$ (about 10 times fainter than HD 209458), with STIS, it will be

possible to obtain comparable science in 5 times the observing time—perhaps 200 orbits for an extensive study of the atmosphere of each new, transiting extrasolar planet, or 2000 orbits for ten planets. However, a sufficient number and variety of targets to exploit this technique fully will only be discovered by a dedicated mission such as Kepler, perhaps within the year following its launch, now planned for 2007. Therefore this program will certainly benefit from, and may even require, an extended HST mission.

4. Science with New Instruments

In this section, we present examples of major scientific programs that could be carried out by HST if it were equipped with two new instruments, a wide-field camera (UWFC) and a high-contrast coronagraph, and if the mission were extended by several years.

UWFC will increase dramatically the instantaneous solid angle that HST can observe, making possible surveys that would otherwise require years or even decades to complete. These surveys are the basis for our exploration of the distant universe and searches for rare objects. The only proposed mission with similar capabilities is SNAP. In the discussion below, we assume UWFC has an FOV of $10' \times 10'$, sensitivity similar to that of ACS, and a wavelength coverage from $0.4 \mu\text{m}$ to at least $1.4 \mu\text{m}$, and ideally to $1.8 \mu\text{m}$.

The advanced coronagraph provides HST with a new capability, improving high-contrast imaging by several orders of magnitude. No other facility planned or proposed for launch before TPF has this capability. Giant planets and debris around nearby stars, dust shells around distant giant stars, and the zone near the central engines of AGN all fall in the observational domain that would be opened up by the coronagraph. In the following discussion, we assume this instrument has the capabilities of CODEX, the coronagraph proposed for HST in 1997.

4.1. Dark Energy

A factor of 10 increase in FOV (UWFC over ACS) would transform HST into a multiplexing supernova harvester. This would enable even more ambitious studies of the dark energy than those described in §3.1. After creating first-epoch images, single pointings of UWFC would yield ~ 2 SNe Ia on average. Chance fluctuations, however, would provide many supernova-rich fields, with 3 or 4 SNe Ia. Optimization of the search and follow-up would return 3 to 4 SNe Ia light curves in parallel. Thus, an allocation of just 2 months of HST could yield more than 200 SNe Ia, or a tenth of the sample SNAP proposes to collect.

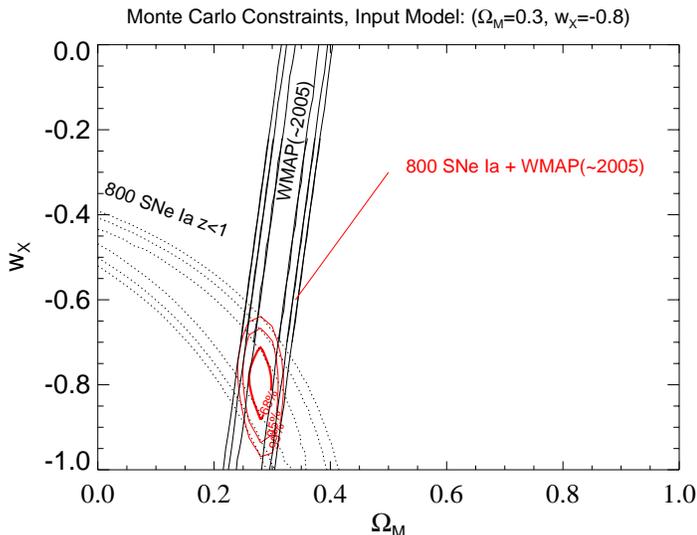


Fig. 4.1-1. Confidence contours (1, 2, and 3σ enclosed) for the equation-of-state parameter w_X and the matter-density parameter $\Omega_M = 1 - \Omega_X$ with the 800 SNe Ia expected from the program described in §4.1. The dotted contours show the constraints based on the supernovae alone and the bold contours for SN plus WMAP in the future. We assume a redshift limit of $z < 1.0$.

A year of HST time dedicated to this project could attain half the SNAP sample and much of its discriminating power.

For the initial goal of constraining the time-averaged value of the equation-of-state parameter w , a large sample of SNe Ia at low redshifts with well controlled systematics can be combined with high-redshift information from WMAP. In Figure 4.1-1, we show the results of a Monte Carlo simulation of 800 SNe Ia at $z < 1$ together with the projected constraints from WMAP in ~ 2005 . This combined program would be potent, providing a measurement of w with $\sim 5\%$ precision, almost rivaling SNAP.

As discussed in §2, key information about the dark energy comes from the evolution of its equation of state (or lack thereof). In particular, a measurement of dw/dz would distinguish between a cosmological constant and a decaying scalar field, the two leading models of the dark energy. Space-based observations are crucial for this enterprise, because only they can discover high-redshift supernovae and follow their fading light curves. By expanding the search and follow-up program with UWFC described above to higher redshifts, we can make relatively precise measurements of both w and dw/dz . This is shown by the Monte Carlo simulation in Figure 4.1-2 of a one-year program to observe 1000 SNe Ia at $0.3 < z < 1.6$. The results expected from HST almost rival those from SNAP. With either mission, we could readily discriminate among models of the dark energy and come closer to knowing the fate of

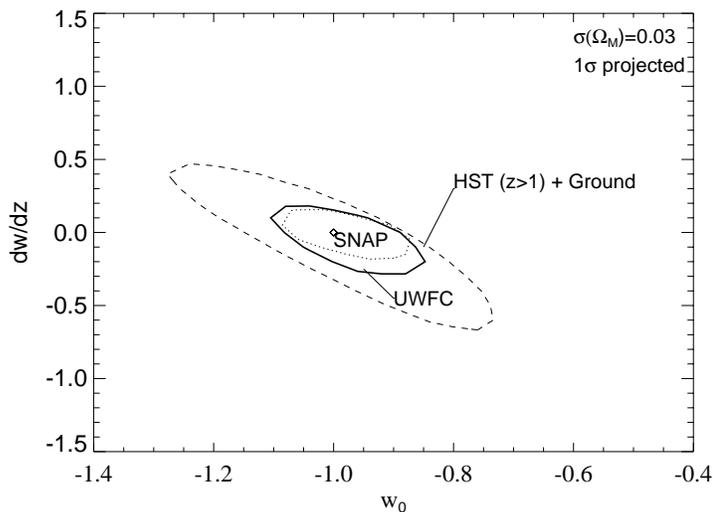


Fig. 4.1-2. Confidence contours with 39% probability enclosed (1σ projected on each axis) for the equation-of-state parameter w_0 and its derivative dw/dz at $z = 0$ for three different dark-energy programs. The HST+Ground and SNAP programs are the same as in Figure 3.1-2. The UWFC program collects 1000 SNe Ia at $0.3 < z < 1.6$ in about a year with HST alone. Each program assumes knowledge of Ω_M to 10% (or equivalently, the WMAP constraints) and irreducible systematic uncertainties appropriate for either ground- or spaced-based observations as defined by the SNAP collaboration.

the universe, whether it will expand forever or eventually contract and end in a “big crunch.”

The progress that can be made with HST toward understanding the dark energy in the near future without waiting for SNAP is a little analogous to the situation with the CMB. Balloon- and ground-based studies have been well worth the effort, even if WMAP and Planck ultimately provide definitive measurements. Surprises are possible at any new level of precision. In studies of the dark energy, HST could soon begin to answer some of the most fundamental questions in particle physics and astrophysics.

4.2. Dark Matter

With an FOV of 100 square arcmin and sensitivity comparable to ACS, UWFC will be about 10 times faster, and therefore the challenging weak-lensing program outlined for ACS in §3.2 would be feasible in about 450 orbits, comparable to the size of current Treasury programs.

A wide-field imager such as UWFC would enable a program with even greater impact.

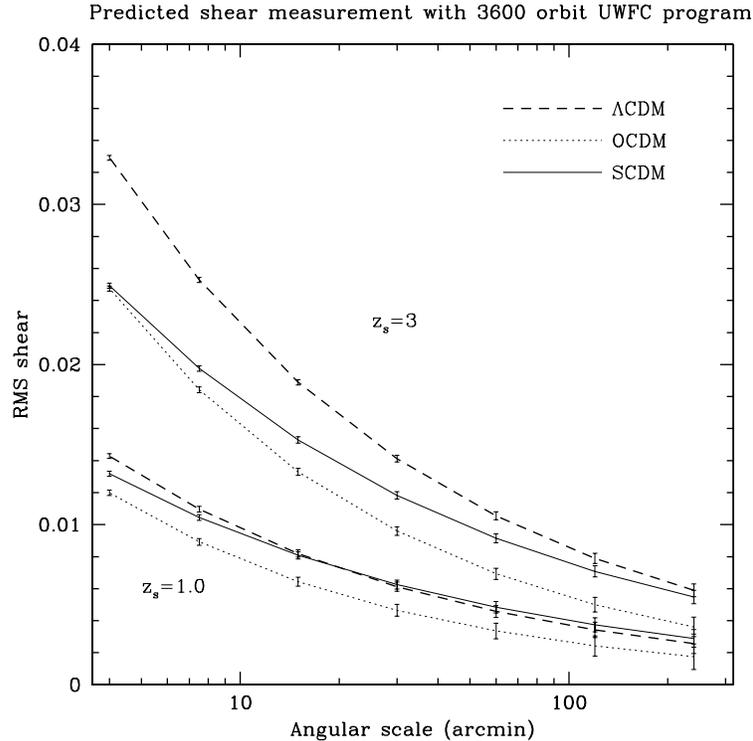


Fig. 4.2-1. Expected shear measurements in a 3600-orbit weak-lensing program covering 25 square degrees with UWFC. The error bars indicate the expected 1σ uncertainties in the RMS shear on the given scale, for a standard (Λ) cosmology, a flat cosmology, and an open cosmology, respectively. Each set of curves pertains to one quarter of the sample, centered at the indicated redshift.

In order to take maximum advantage of the wide-field characteristics of UWFC, an optimized weak-lensing program would cover 25 square degrees in 900 pointings, about 300 times the area covered by GOODS. Four orbits per pointing will be split evenly between V and I , reaching a depth $I_{AB} \sim 26.3$ at the $S/N \sim 20$ necessary for reliable shape measurements. Coordinated ground-based observations can be used to obtain photometric redshifts, following the model of the GOODS project. Alternatively, the necessary color information for photometric redshifts can be obtained with HST, at the cost of reducing the total area coverage to about 10 square degrees for the same total observing time.

The UWFC program would improve the quality of the shear measurement by a factor ~ 5 with respect to the ACS+WFC3 program described above (see Figure 4.2-1), thus enabling more accurate testing of the models for the growth of structure and discriminating between cosmologies. However, the major qualitative improvements that the UWFC program can achieve are (1) probing the growth of structure on very large scales, up to hundreds of

comoving Mpc and (2) revealing the cosmic web.

On large scales, the distortion due to dark matter scales roughly as $1/\sqrt{\theta}$, while the measurement error scales as $1/\theta$. At some point, it becomes possible to “see” the large-scale structure itself, i.e., to produce a map of the structure where individual mass concentrations can be identified and measured. For the depth of the proposed survey, this cross-over occurs around $30'$, corresponding to a typical precollapse total mass $\sim 10^{15}M_{\odot}$. The proposed survey will produce a map of the dark matter distribution in which individual structures $30'$ and larger will be revealed. Smaller structures will be visible, but noisier; individual high-density regions will be seen clearly.

An even more ambitious survey could be carried out by taking only one filter and one orbit worth of data per pointing, and relying fully on ground-based images for the photometric redshift information. A survey along these lines, using one year of HST data, would cover about 150 square degrees, or half the area of the SNAP wide-field survey, at an effective depth of $I \approx 26$, and with only one filter. This survey would improve the measurement of dark matter clustering by an additional factor of 2, and enable detection of mass concentrations on scales up to 10° . However, this survey would rely more heavily on ground-based observations for photometric redshift information, and it would lack the built-in shape confirmation that comes from measuring galaxy shapes in two independent filters.

4.3. Young Galaxies

While the surveys proposed in §3.3 can provide basic statistics of the galaxy population at $6 \lesssim z \lesssim 8$, these programs certainly push the limits of the current and planned instruments on HST. A more complete understanding of galaxy formation and evolution, however, requires larger samples of typical galaxies extending to even higher redshifts (see Figure 4.3-1). This necessitates surveys covering significant comoving volumes at each redshift of interest, or several tenths of a square degree at a minimum. Surveys for Lyman-break galaxies at $z \approx 3$ have demonstrated how the physics of galaxy formation (beyond just the demographics of number counts) can be tackled with samples of thousands of galaxies. Their clustering tells us about masses and biasing. Their spatial distribution relative to quasar absorption-line clouds tells us about the dynamic interplay between star formation and the surrounding IGM. Morphologies of thousands of galaxies from the GOODS program will characterize the modes, scales, and structures of star formation at $z \lesssim 4$. However, neither GOODS, the UDF, nor even the extended-mission ACS+WFC3 survey described above, are yet capable of providing this wealth of information for younger galaxies in the redshift range $4 \lesssim z \lesssim 9$.

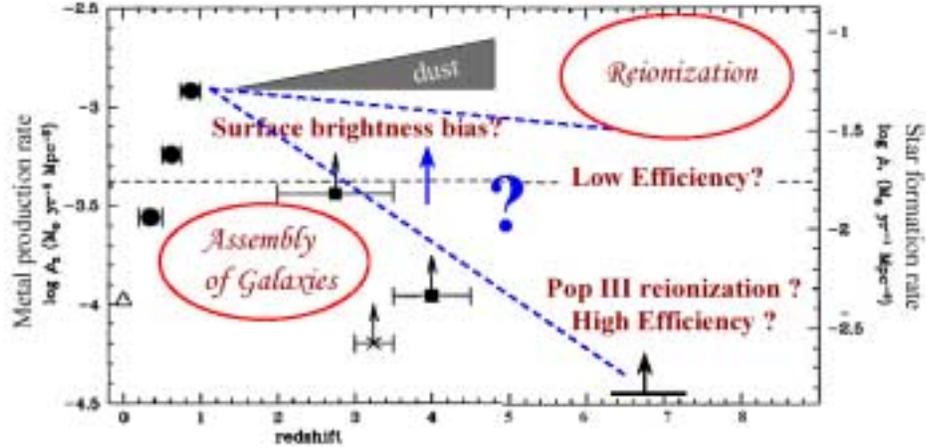


Fig. 4.3-1. Schematic diagram showing present uncertainties in the cosmic star-formation and metal-production rates. The data points are from a variety of ground-based and HST observations. In the galaxy assembly phase, at $z \lesssim 6$, these rates are uncertain because of the possible roles of dust and surface brightness bias. In the reionization phase, at $z \gtrsim 6$, the expected properties of objects could change considerably depending on whether the sources produce escaping ionizing radiation with high or low efficiency (compared to non-ionizing radiation). The low-efficiency sources would be easier to detect (for a given global ionization rate).

UWFC offers a powerful tool for achieving this goal, making it possible to cover ~ 1000 square arcmin to UDF depths or fainter, thus providing thousands of galaxies per unit redshift in the reionization era. To reach redshifts substantially beyond $z \approx 6$, UWFC must have excellent sensitivity at wavelengths redder than $1 \mu\text{m}$, i.e., outside the silicon CCD range; here, we assume the sensitivity cutoff is at $1.4 \mu\text{m}$, although $1.8 \mu\text{m}$, where the thermal emission from HST starts to dominate, would be even better. Scaling from previous calculations, this survey could cover 1000 square arcmin (10 UWFC fields) to UDF depths (150 orbits per filter) in four bands over 6000 orbits.

Some of the best probes of the reionization era from UWFC, however, would come from slitless spectroscopy. The galaxies at $z > 6$ detected in deep HST imaging surveys will mostly lie beyond the limits for spectroscopic confirmation by 10-meter ground-based telescopes. Although ACS+WFC3 slitless spectroscopy can measure redshifts from continuum breaks and emission lines, even the ambitious grism project described in §3.3 can only scratch the surface: dispersed spectroscopy always probes shallower continuum flux limits than does imaging, and the ACS+WFC3 imaging program only reaches the most luminous galaxies at $z > 6$. A slitless capability for UWFC would enable extremely long spectroscopic exposures (many hundreds of orbits) covering substantial areas (100 square arcmin per field), and would measure accurate redshifts and emission-line strengths for many hundreds of galaxies

at $6 \lesssim z \lesssim 8$ near the ACS+WFC3 imaging limit.

UWFC could further achieve a large increase in emission-line sensitivity by combining grisms with broad- and medium-band blocking filters to reduce sky noise and confusion. The observing time required to detect a narrow emission line in background-limited grism data is proportional to $\Delta\lambda$, the bandpass contributing to the sky background. The maximum viable integration time before spectral overlap becomes a problem increases with decreasing $\Delta\lambda$, and can thus be extended with filtered grism spectra. Placing the grisms and filters in independent selection mechanisms would provide very flexible spectroscopic modes. While the filters would reduce the redshift coverage Δz of any single observation, the sensitivity gain, coupled with the solid angle of UWFC, would result in a large net gain in the number of detected sources. The improvement in detection limit achieved through reduced sky background and longer usable integrations would enable Ly α searches at least to $z = 10$, and possibly up to $z = 12$.

The greater emission-line sensitivity and volume coverage of UWFC would enable another qualitatively new test of reionization. Because Ly α galaxies are visible only when the gas near them is predominantly ionized (cf. §3.3), detailed grism mapping of the Ly α galaxy distribution will effectively provide a three-dimensional map of the ionized regions at ~ 1 Mpc (physical) resolution. The geometrical effects of the interfaces between the ionized and neutral gas can thus be distinguished from the overall evolution in the Ly α source population, which is obtained by averaging among many lines of sight, while local variations in galaxy density can be controlled for using a sample of continuum-selected Lyman-break galaxies from the same surveys. At the highest redshifts, this mapping will allow identification of individual ionized bubbles in a generally neutral universe. As reionization progresses to the overlap phase, the Ly α source distribution will reflect the larger volume of the ionized regions. Moreover, topological statistics like the genus curve will reveal the transition from single ionized bubbles to complex, multiply connected regions of ionized gas. Because this transition occurs when the neutral fraction is ~ 10 –50%, this regime cannot be probed by either the Gunn-Peterson test (which is sensitive to much smaller neutral fractions, $10^{-3\pm 1}$) or by the CMB polarization (which is sensitive to the total column density of ionized gas and so is dominated by fully ionized regions).

Finally, high sensitivity over a large solid angle may allow us to identify very low metallicity, primordial sources from their HeII 1640 emission. Such “protogalaxies” are often regarded as the holy grail of galaxy formation studies. The intensity of the HeII emission from metal-poor objects is expected to be about the same as that of their H β emission.

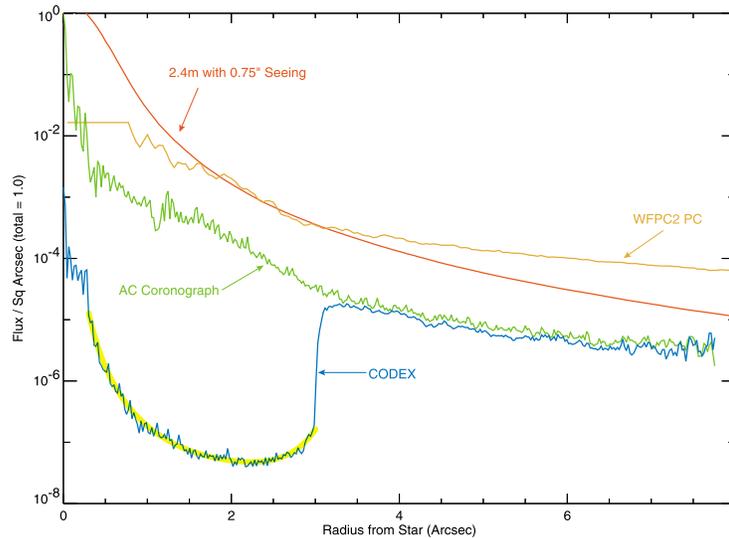


Fig. 4.4-1. Comparison of the CODEX off-axis PSF with those of WFPC2, ACS, and good ground-based seeing conditions, all at 5000 \AA . The CODEX detection zone is highlighted in yellow. The CODEX curve was modeled realistically for a lab-demonstrated deformable mirror with 140×140 actuators over the pupil, a practical alignment algorithm, a Gaussian focal-plane mask, and 10% spectral bandwidth.

4.4. Extrasolar Planets

A high-contrast coronagraph would dramatically advance HST capabilities in the realm of direct planet detection. CODEX serves as a specific example of such an instrument. In addition to an expected rich science return, the development and use of an advanced coronagraph on HST would provide invaluable technical experience toward achieving the more ambitious goals of TPF.

The detection zone of CODEX is an annular region around the star that extends in radius from $0''.3$ to $3''$ for observations at 5000 \AA (see Figure 4.4-1). Sufficiently bright planets appearing within this zone can be detected and tracked as they revolve around the star. From such observations, the physical properties and orbits of the planets can be determined in a range of mass and semi-major axis that depends on the distance and brightness of the target star.

Assuming a planet lies in the detection zone, it is deemed “sufficiently bright” for detection if two conditions are met: (1) the expected signal-to-noise ratio in the integration time is greater than about five (i.e., $S/N > 5$ considering both planet and background photons) and (2) the expected ratio of planet image intensity to background intensity is greater than about one tenth (i.e., contrast ratio $Q > 0.1$). The first of these conditions is

based on information theory and photon statistics; the second limits the risk of confounding planets with speckles and avoids the systematic effects that attend excessively long exposure times.

For each target star, the detection zone subtends a particular range of planet-star separations (or combinations of semi-major axis and phase angle). The planet flux depends on the phase angle, the planet-star distance, and the radius, geometric albedo, and phase function of the planet. [For Earth- and Jupiter-equivalents observed at greatest elongation (90° phase angle) the planet fluxes are 2×10^{-10} and 8×10^{-10} of the star flux, respectively.] The planet mass can be inferred within about a factor 2 from its brightness and orbit.

Figure 4.4-2 shows the discovery space for CODEX observing one of the two nearest stars, α Cen B, which is 1.3 pc distant. In less than 10 hours of integration time, it would detect a planet half the mass of Earth in a Mars orbit, or Earth in an Earth orbit. Figure 4.4-3 shows the case of ϵ Eri at 3.2 pc, for which CODEX could find Uranus- and Neptune-type planets on Mars/Jupiter-type orbits. Figure 4.4-4 shows the case of Gliese 150 at 9 pc; here, CODEX would discover Jupiter and Saturn.

Direct detection by a high-contrast coronagraph has several qualitative advantages, and few if any disadvantages, compared with the exhaustive indirect detection techniques, which are based on the reflex motion of the star, namely, radial velocity and astrometric searches. First, discovery is immediate, whereas the indirect techniques require that observations span most of a planet orbital period to obtain results. Second, the spectral content and secular variation of the planet light can be studied to learn about the surface or atmosphere of the planet and potentially to discover rotation periods, weather, or climactic change. Meanwhile, the orbital elements and planet mass can be determined with sufficient precision to answer questions about formation mechanisms and orbital evolution.

A high-contrast coronagraph on HST would find extrasolar planets in this decade that are below the sensitivity of radial velocity searches and that the Space Interferometry Mission (SIM) would discover in the next decade. The coronagraph would characterize those planets and compare them with solar system planets, based on their spectral and photometric properties.

With a year of observing time, a CODEX-like instrument on HST could provide an exhaustive survey for planets around all of the 25 nearest/best target stars. In addition, the next 100 more distant stars could be surveyed for favorably placed Jupiter-type planets. Application of advanced analysis techniques, e.g., the spectral deconvolution discussed in §3.4.1, may yield additional gains in observing efficiency and detection limits. In this way, we will begin to learn how to utilize advanced coronagraphs on space observatories.

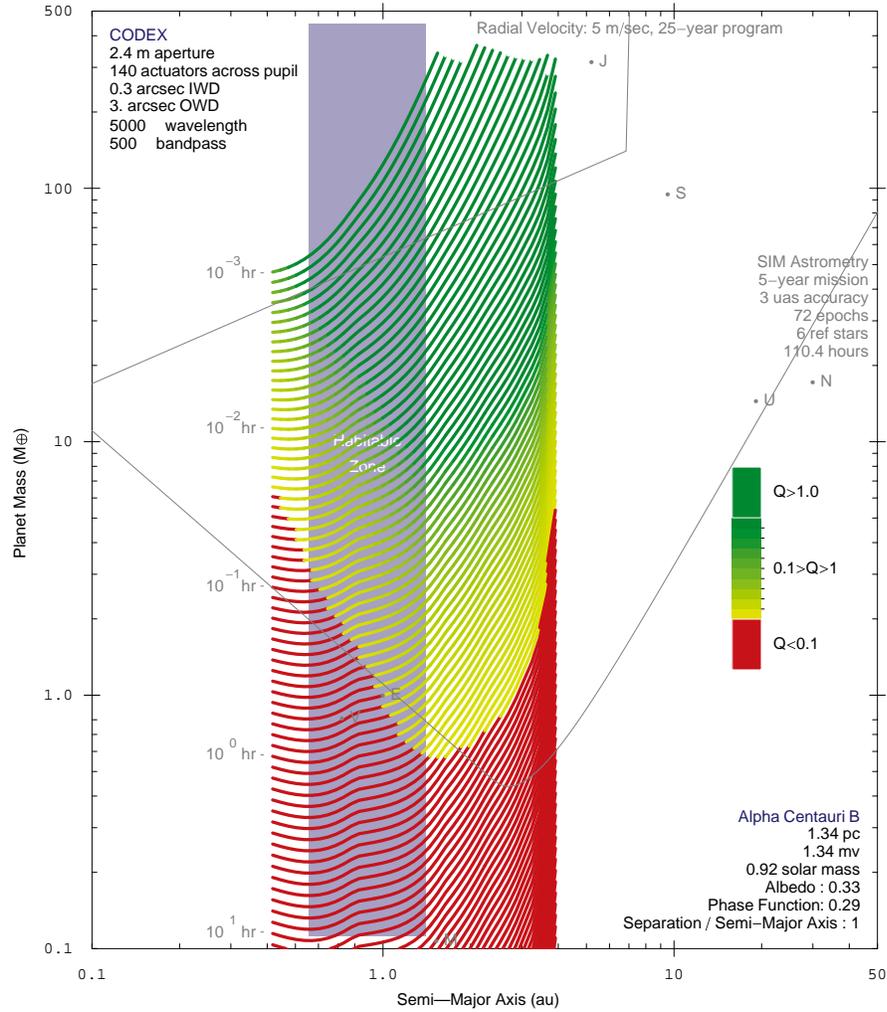


Fig. 4.4-2. CODEX detects Earth around the nearest stars (specifically, α Cen B). Each colored curve shows parametrically the mass and semi-major axis of the planet detected with $S/N = 5$ for the indicated integration time. The color indicates the contrast ratio between the planet image and background light ($Q > 0.1$ means “detectable”). Solar system planets are shown as points with letter labels. The “habitable zone” is the range of semi-major axis where water can exist in liquid form. The radial velocity curve shows the minimum detectable planet mass versus semi-major axis for a minimum detectable amplitude of 5 m s^{-1} and a 25-year program duration. The SIM curve shows the minimum detectable mass for 110.4 hours of observation spread over the five-year mission.

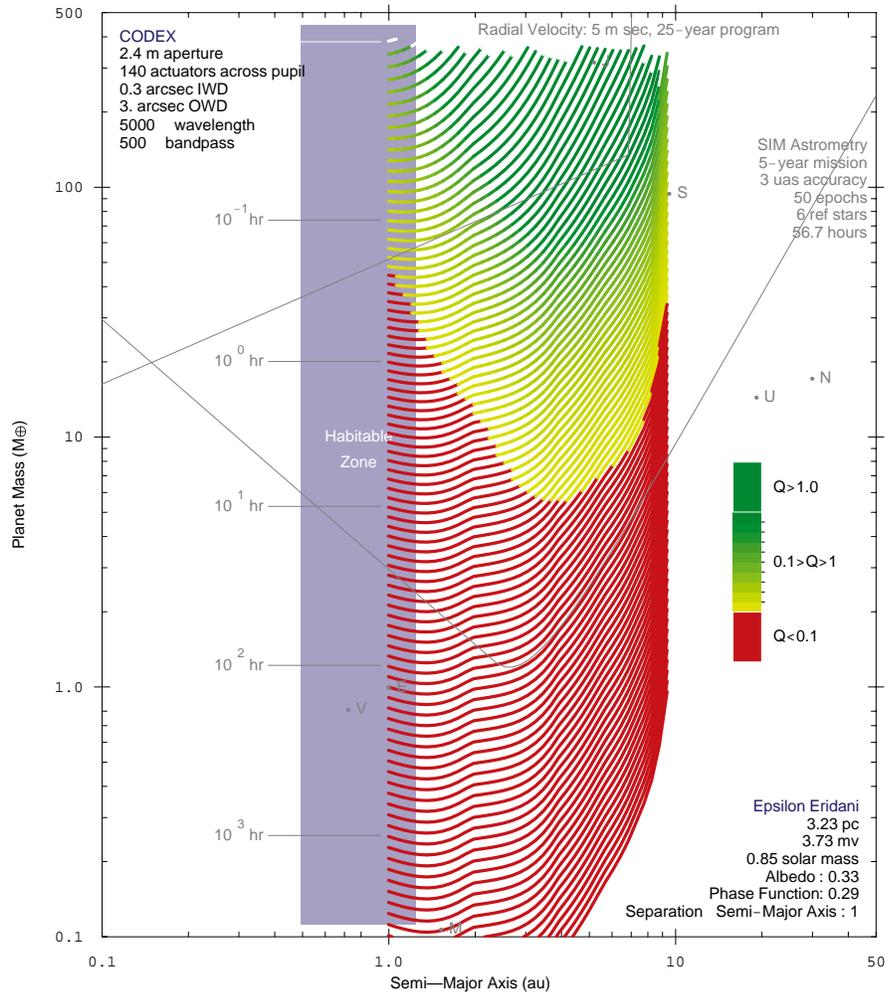


Fig. 4.4-3. CODEX discovers and studies Uranus/Neptune-type planets on Mars-Jupiter orbits around stars to about 5 pc.

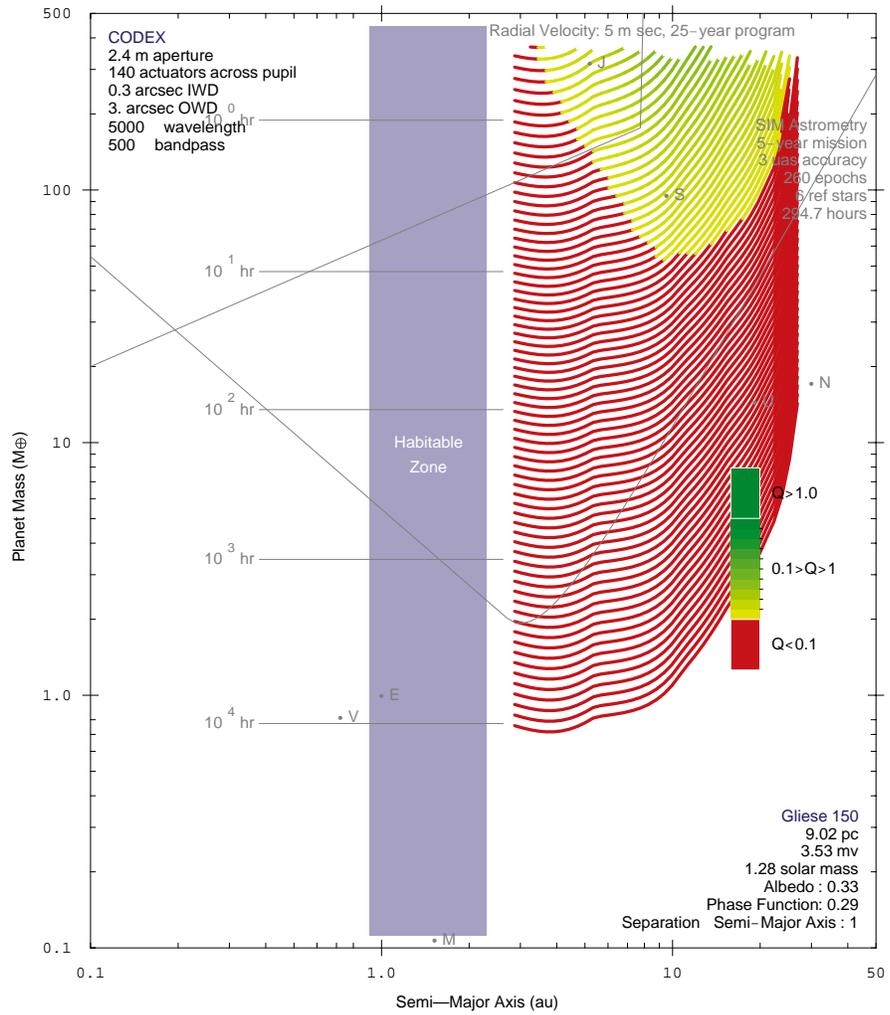


Fig. 4.4-4. CODEX discovers and studies Jupiter/Saturn-type planets around stars to about 10 pc.

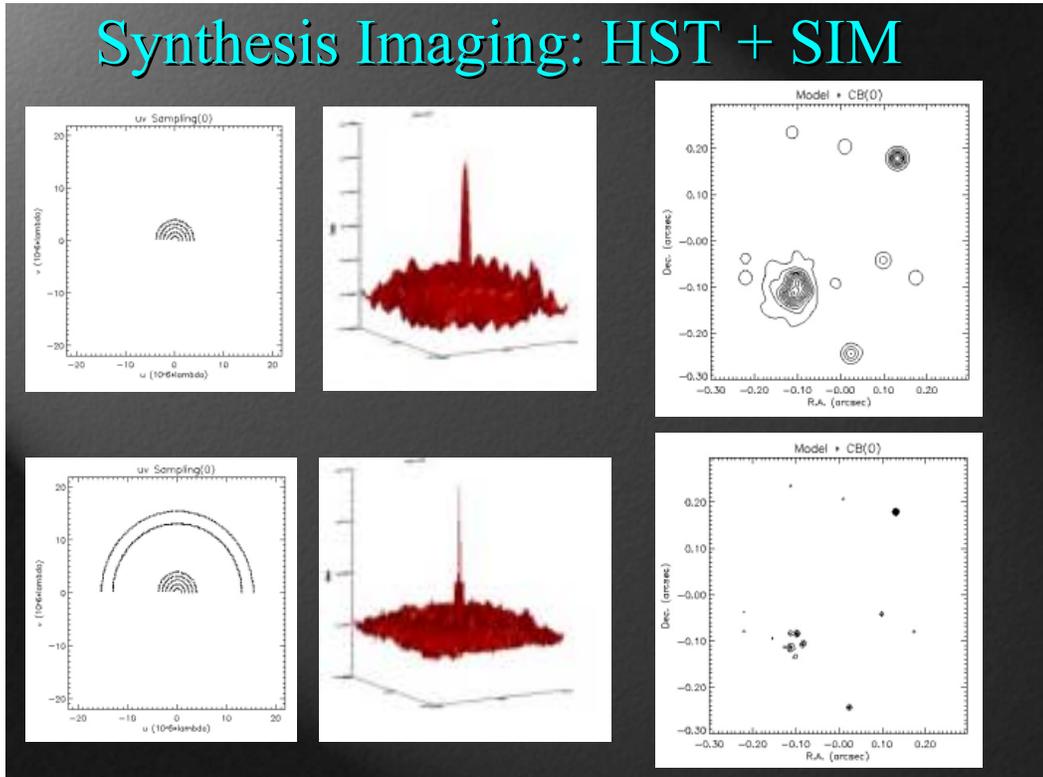


Fig. 5-1. In order to get a rough idea of what might be possible, we have carried out a crude calculation using the computer-based simulator for SIM developed at STScI. A relatively uncomplicated field of about a dozen stars (all of the same temperature) has been simulated and “observed” with an HST model alone (upper row of three panels) and combined with 8.5- and 10-meter baseline data from SIM (lower row). The improvement with SIM is quite significant, as a comparison of the upper and lower versions of the third panel shows. An object that appears as an extended source in the lower left part of the HST field is resolved into a compact cluster of ~ 7 stars when SIM data are added. Conversely, the HST data are essential to providing the first guess to the true distribution of brightness, as required for the image restoration step.

5. A Possible Synergy between HST and SIM

HST provides information about the distribution of brightness on the sky on spatial scales corresponding to interferometers with baselines ranging from 0 to 2.4 meters, with relative weight varying from 1 to 0 more-or-less linearly over this range. The HRC mode of ACS provides 25 mas “critical” sampling of images made in the *V* band (and longward) for

the ~ 50 mas PSF.

After launch around 2010, SIM will provide information on angular scales of order 10 mas over small ($\sim 1''$) fields of view using its 10-meter baseline science interferometer and its 8.5-meter guide interferometer. This information can in principle be added to small subfields of HST images to improve significantly the resolution of the final image; the potential exists to increase the resolution of such fields by a factor of 4 to 5. This will work well only in selected cases where the image structure is known to be not too complicated. Possible applications include the cores of dense stellar systems, nearby planetary systems, circumstellar matter, and star clusters and supernovae in nearby galaxies. Figure 5-1 gives a glimpse of what might be possible.

Although it is not required that the combined observations be carried out at the same time, various practical considerations suggest that it is not likely one will find suitable ACS/HRC data in the HST archive on every field potentially of interest for combined observations. Such data sets can only be guaranteed if HST remains in operation well beyond the launch of SIM.

A study needs to be conducted to determine the difficulties with actually carrying out this combination using realistic models for HST and SIM and with model targets of varying complexity.

6. Conclusion: Extended HST Mission

HST is making and will continue to make crucial contributions to some of the great scientific issues of our time. The issues we have focused on here—dark energy, dark matter, young galaxies, and extrasolar planets—are likely to remain central themes of physics and astronomy for at least another decade. Indeed, these themes have largely defined the NASA Origins Program in general and the JWST and TPF missions in particular. Until these missions fly—in 2010 and 2015, or later—HST will provide unique high-resolution imaging capability with ACS and WFC3, enhanced by SIM, and unique spectroscopic capability with STIS and the Cosmic Origins Spectrograph (COS). The installation of one or more new instruments would make HST even more powerful. A wide-field camera such as UWFC would lead to major advances in our exploration of dark energy, dark matter, and young galaxies, while a high-contrast coronagraph would enable new searches for extrasolar planets. These general-purpose imaging capabilities, and the corresponding opportunities for scientific discovery and public outreach, will not be equaled by any other existing or planned missions until JWST and TPF are launched.

HST is unique among space science missions in another important respect. It is the only observatory that has been upgraded over its lifetime with the installation of new and better instruments. Consequently, it has remained at, or more accurately, it has defined, the forefront of astronomical research since its launch in 1990. In terms of observing capabilities, HST is never older than its last servicing mission. All other space science missions inevitably fade into obsolescence, usually within a few years. This means that the rationale for the end of the HST mission must also be different. The science HST can do now is just as important and exciting as the science it could do ten years ago. We have argued here that the science HST could do ten years from now with an extended lifetime and new instruments would be equally important and exciting. Thus, we believe the decision to terminate HST should not be based on the somewhat arbitrary assumption made before its launch about what its useful lifetime might be. We believe instead that the decision should be based on careful consideration of the costs, risks, capabilities, and opportunities for scientific discovery and public outreach of HST within the present context of existing and planned missions.