A Visible Imager For NGST

CSA Instrument Study Final Report

September 28, 1999

Paul Hickson[1], Andrew Bell[2], Gary Buttner[2], Laurent Drissen[3], Timothy Hardy[4], John Hutchings[4], Christopher Morbey[4], Richard Murowinski[4], Neil Rowlands[2], Eric Steinbring[1,5], Peter Stetson[4]

Affiliations:

1. The University of British Columbia
2. EMS Technologies Canada Inc.
3. Université Laval
4. National Research Council of Canada
5. The University of Victoria

Document Number:

UBC-VI-003

Table of Contents

1 Science

1.1 Executive Summary

The Next Generation Space Telescope offers unprecedented opportunities for scientific discovery. A key goal is to observe and study the formation and evolution of galaxies from the time that the first light was emitted until their present form is attained. To do this requires observations over a vast range of distances and equivalently, redshifts. HST observations have shown that, at a given flux level, high-redshift galaxies are rare and that special techniques are needed to find them. Almost all that are known have been discovered by virtue of a strong discontinuity in their spectra at 121 nm, the Lyman break, which results from absorption of short-wavelength photons by intergalactic hydrogen. This method requires observations at wavelengths both shorter and longer than the break. This means that sensitive imaging at visible wavelengths are needed to identify these galaxies over the redshift range $z \sim 3 - 8$. This large range of redshifts encompasses the range of cosmic time where it is believed that the bulk of star formation occurred. It is also the period when galaxies were being assembled by the merging of smaller subunits. Clearly this redshift region is key to the understanding of galaxy formation. In fact, at present we do not know of *any* galaxies that have redshifts greater than this.

While the light from the older stellar populations in these galaxies will be redshifted to near-infrared wavelengths, these objects actually emit most of their radiation in the rest-frame ultraviolet due to newly-formed luminous hot stars. If absorption by interstellar dust is not severe in these primitive objects, this radiation will be seen at the visible region. Thus, visible observations will be essential to study the morphology of the star-forming regions in these galaxies.

There are also many important observations in the local universe that are best done at visible wavelengths. Observations of the faintest white dwarf stars can provide an independent and reliable estimate of the age of the oldest stellar populations in our galaxy. These stars are blue and very faint. Simulations indicate that they can only be reliably measured by an optimised Visible Imager such as the instrument proposed here. Cepheid variables are an important distance indicator. These stars must be observed at visible wavelengths because they do not vary significantly in the near-infrared. Dense star clusters are best observed in the visible where crowding will be less severe.

Observations at visible wavelengths also offer significant performance advantages over those at longer wavelengths for many scientific programs. The NGST image quality is expected to be quite good at visible wavelengths, with most of the light being concentrated in a near-diffraction-limited image core. At a wavelength of 0.6 um, star images are three times smaller than at 1.8 um and the central intensity of the image is about five times greater. This improvement in contrast results in greater sensitivity to features agains a bright background, such as star-froming regions in galaxies. Moreover, the sensitivity of all observations at low to moderate spectral resolution will be limited by background light. The relevant quantity is the number of background photons per spatial resolution element and this is lowest in the visible region. With its 8-m primary mirror, NGST will be over a hundred times more sensitive than HST at visible wavelengths.

It is important to emphasize that the gains in resolution and sensitivity offered by visible imaging can only be realized if the image is properly sampled. Large pixels, such as those envisioned for the baseline near-infrared camera, blur the image so severely that the sensitivity gain is entirely lost. While it may be possible to recover some of the resolution by dithering (repeated observations with small position offsets), the loss of sensitivity and signal-to-noise ratio is irretrievable.

The instrument that we propose takes maximum advantage of the performance gains available in the visible. For stellar photometry in crowded fields, it provides more than a magnitude gain in sensitivity over the near-infrared camera. The improvement in resolution and contrast of faint galaxies is equally dramatic. This Visible Imager provides a large 2-arcmin field of view with 10 mas sampling and a detector that is optimised for visible radiation. This combination will have no rival among any other space or ground-based telescopes that are currently foreseen.

The Visible Imager, as presently envisaged, employs a three-mirror relay system to illuminate a 12K × 12K pixel mosaic array of silicon p-i-n diodes. The optical system has an accessible pupil at which are located two filter wheels. These provide a total of 32 positions and allow dispersive elements (grisms) and band-selection filters to be used together. No shutter is required. Focus and alignment are achieved by piezoelectric actuators that control the position and orientation of the detector array. The entire instrument can be accommodated within one quadrant of the NIR camera, or, it can be a stand-alone instrument sampling an independent region of the NGST focal plane.

The Si p-i-n arrays have high quantum efficiency (typically over 90%) over the entire visible spectrum. These devices have lower noise than InSb, excellent uniformity of response, are radiation hard, and perform well at 30K. As their response cuts off around 1.1 um, the problem of filter red leakage with infrared detectors is avoided. Being hybrid detectors, they use similar readout schemes and cold electronics as infrared arrays. The mosaic is built from 2K × 2K arrays. An innovative packaging scheme provides a very high (87%) filling factor. The entire array can be read out in less than 2 minutes. Sub-areas can be read out more rapidly. This would allow the Visible Imager to provide fine guidance information to the NGST control system.

Even though the proposed detector contains a large number (150 million) of pixels, it need not impose undue requirements on the NGST ISIM and other resources. The power dissipation of the array, and the data processing/transmission load, depend on the frequency at which the array is read out. The exposures for visible imaging can and will generally be quite long, limited only by the accumulation of cosmic ray hits. With a well-sampled image, cosmic rays are more easily removed, so exposure times can be longer. Efficient data compression algorithms exist which can provide a reduction in data quantities by as much as two orders of magnitude with very little loss of information. The cold electronics are optimised to minimize parasitic heat losses.

In summary, we propose an instrument that fully exploits the high resolution and sensitivity that NGST can provide at visible wavelengths. This enables, or augments, a wide range of core science programs. In particular, it will allow us to see directly the galaxy formation process and follow it from the time of the earliest known galaxies to the current epoch. Equipped with such an instrument, the NGST will be unrivalled

by ground and space telescopes for wide-field high-resolution imaging. The detailed and panoramic images produced by this camera will go far in promoting space astronomy to the public.

1.2 Science Capability

Visible-wavelength imaging with the NGST provides new scientific opportunities and offers significant performance advantages in many areas. In recognition of this, the NGST spectral coverage requirement now extends to a minimum wavelength of 0.6 um. Despite the importance of the visible, the baseline NGST instrument complement contains no instruments that can fully exploit this spectral region. Imagers optimised for infrared wavelengths naturally provide less-than-optimal performance, both in resolution and sensitivity, when observing at wavelengths several octaves below their design wavelength. We begin by illustrating quantitatively the potential gains offered by visible imaging, and the technical requirements to achieve them. We then examine the scientific opportunities presented by an optimised Visible Imager and consider the relationship of visible imaging to the NGST Design Reference Mission (DRM).

1.2.1 Advantages of Visible Imaging

There are three primary motivations for visible imaging with NGST:

- *Spectral coverage* – some key scientific programs, such as the detection and study of galaxies over a wide range of redshifts, cannot reasonably be done without observations at visible wavelengths.

- *Resolution* – visible wavelengths offer resolution that is several times better than that which can be obtained in the near infrared. This is extremely important for detailed morphological studies and photometry in crowded field.

- Sensitivity – The sharper images result in lower backgrounds, per resolution element, at visible wavelengths than in any other spectral band accessible to NGST. This, combined with the high sensitivity and low noise of visible detectors and the fact that many objects are most luminous at visible wavelengths, lead to significant sensitivity gains.

Visible observations from space offer many advantages compared with ground-based observations. While much progress has been made at achieving high resolution by means of adaptive optics at NIR wavelengths, the results at visible wavelengths have been much less dramatic. The achieved Strehl ratios are low, of order 0.1 or less, even at 1 um wavelength, and the available field of view (FOV) is small. At 1 um, the isoplanatic radius (the distance from the reference star at which the Strehl ratio drops by a factor of two) is only 15 arcsec and decreases at shorter wavelengths. While this area may be increased by means of multiple conjugate deformable mirrors and laser guide stars, it is unlikely that the performance, and FOV will rival that of NGST.

1.2.1.1 NGST image quality

The NGST point-spread function (PSF) at visible wavelengths depends on parameters, such as mirror surface roughness, that are not yet fully known. However, we know from basic optical theory and experience that the PSF, at wavelength , will contain a sharp core having angular diameter of order /*D*, where *D* is the maximum diameter of the telescope primary mirror. For *D* = 8 m, this diameter ranges from 13 to 26 mas over the visible band (= 0.5 to 1.0 um). Small-scale mirror errors and surface roughness result in some of the light being scattered into a diffuse halo. The fraction of light in the core can be estimated from the well-known approximation for the Strehl ratio $S = \exp\left[-(2\pi s / \lambda)^2\right]$, where is the RMS wavefront (optical path difference) error. Tests of prototype mirrors indicate that this may be quite high, with typically 50% or more of the light being in the core at visible wavelengths. Figure 1.2-1 illustrates the expected variation of the Strehl ratio with wavelength, normalized to a value of 0.3 at 0.5 um. Also plotted is the central intensity of the image, or image contrast, which governs the detectability of faint point sources

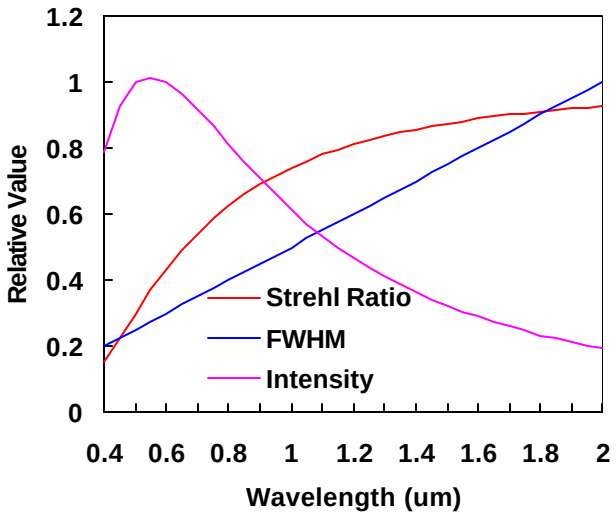


Figure 1.2-1. Expected variation of FWHM, Strehl ratio and PSF central intensity vs wavelength. The intensity of the image core peaks in the visible.

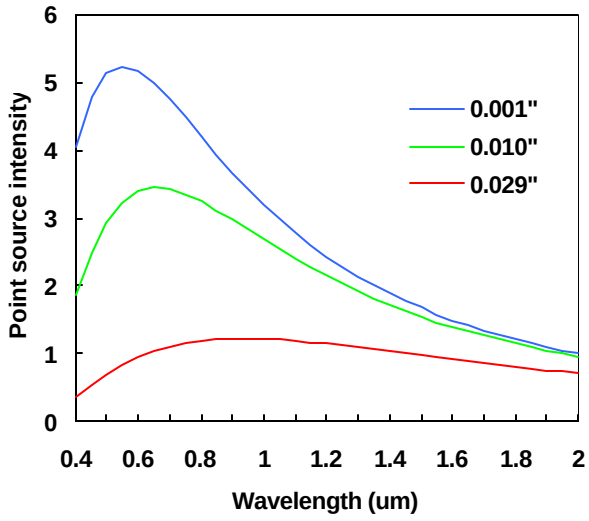


Figure 1.2-2. Effect of pixel size on image core intensity. The contrast gain in the visible is lost with large pixels.

in the background-limited regime. It can be seen from the figure that the contrast reaches a maximum near 0.6 um where it is about a factor of 4 greater than at near-infrared (NIR) wavelengths.

It is important to realize that this gain in contrast, and the corresponding increase in sensitivity for the detection of unresolved objects and features, requires good sampling of the PSF. With a finite pixel size, what is actually measured is the *effective PSF*, which is obtained by convolution of the telescope PSF with the response function of the pixels. This results in image blurring and loss of contrast. This effect is illustrated in Figure 1.2-2, which shows the image contrast for several pixel sizes. It can be seen that *with the large (29 mas) pixels of the baseline NIR camera, the contrast (and resolution) gain offered by visible wavelengths is entirely lost*. A pixel size of order 10 mas or less is needed to realize most of the gains.

It is also important to stress that *the loss in sensitivity caused by too-large pixels cannot be recovered by dithering or other multiple sampling technique*. This is because it is always the effective PSF that is sampled.

1.2.1.2 Visible Imager Performance

To illustrate the potential gains offered by visible imaging, realistic simulations were conducted. These simulations used a PSF that accurately models the expected diffraction and scattering from the NGST

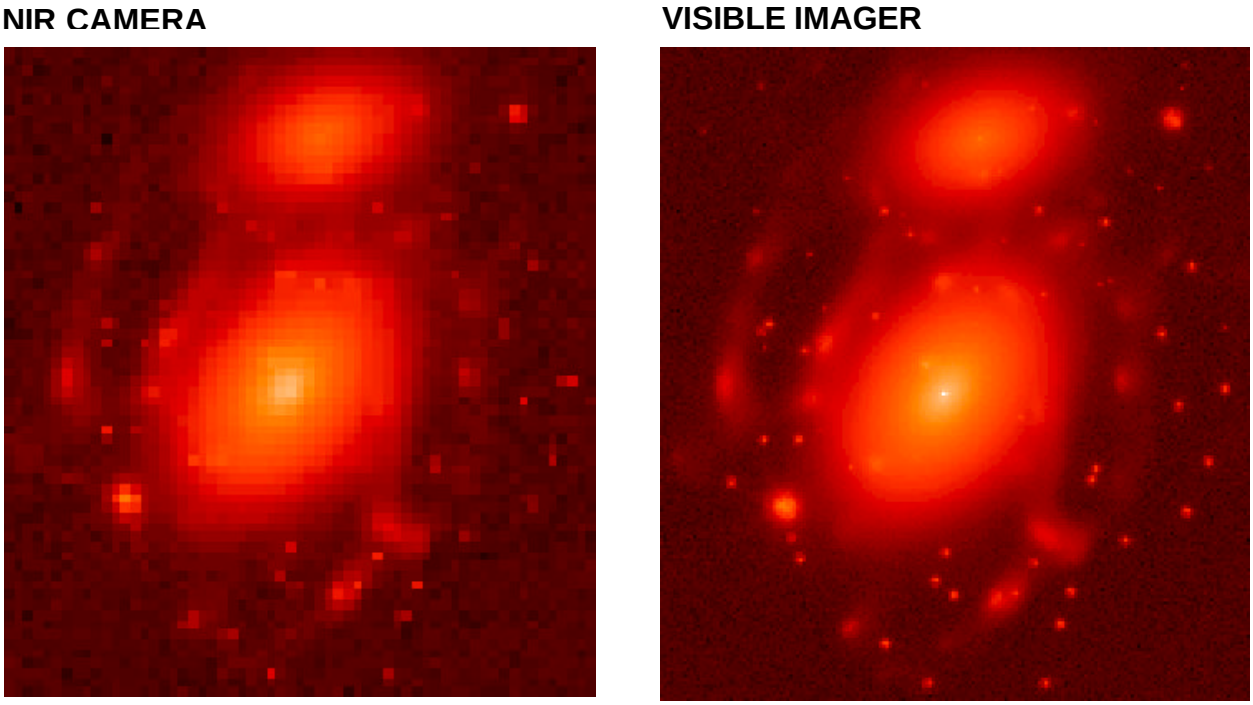


Figure 1.2-3. The left panel shows a simulated galaxy image as it would be seen in the visible by the baseline NIR camera. The right panel shows the same image as seen by the Visible Imager. Note the improvement in resolution of the nucleus and star clusters and the fainter detection limit of the latter.

optics, and the properties of distant galaxies. The simulation program is a powerful tool that will allow us to explore the performance of the Visible Imager for a wide variety of scientific programs and telescope parameters. More details of the simulation process can be found in the paper presented by Eric Steinbring at the September 1999 Hyannis meeting and from his web site (http://wwwastro.phys.uvic.ca/~steinb). Figure 1.2-3 compares a simulated galaxy observation with the proposed Visible Imager and the baseline NIR camera (operating at the same visible wavelength). The improvement in contrast and resolution provided by the Visible Imager is evident. Figure 1.2- 4 shows the result of photometry performed on a simulated globular cluster image. There is a substantial improvement in the stellar photometry with the Visible Imager, amounting to a gain of more than a magnitude. This allows the white-dwarf sequence to be clearly delineated, whereas it is only marginally seen with the baseline NIR camera.

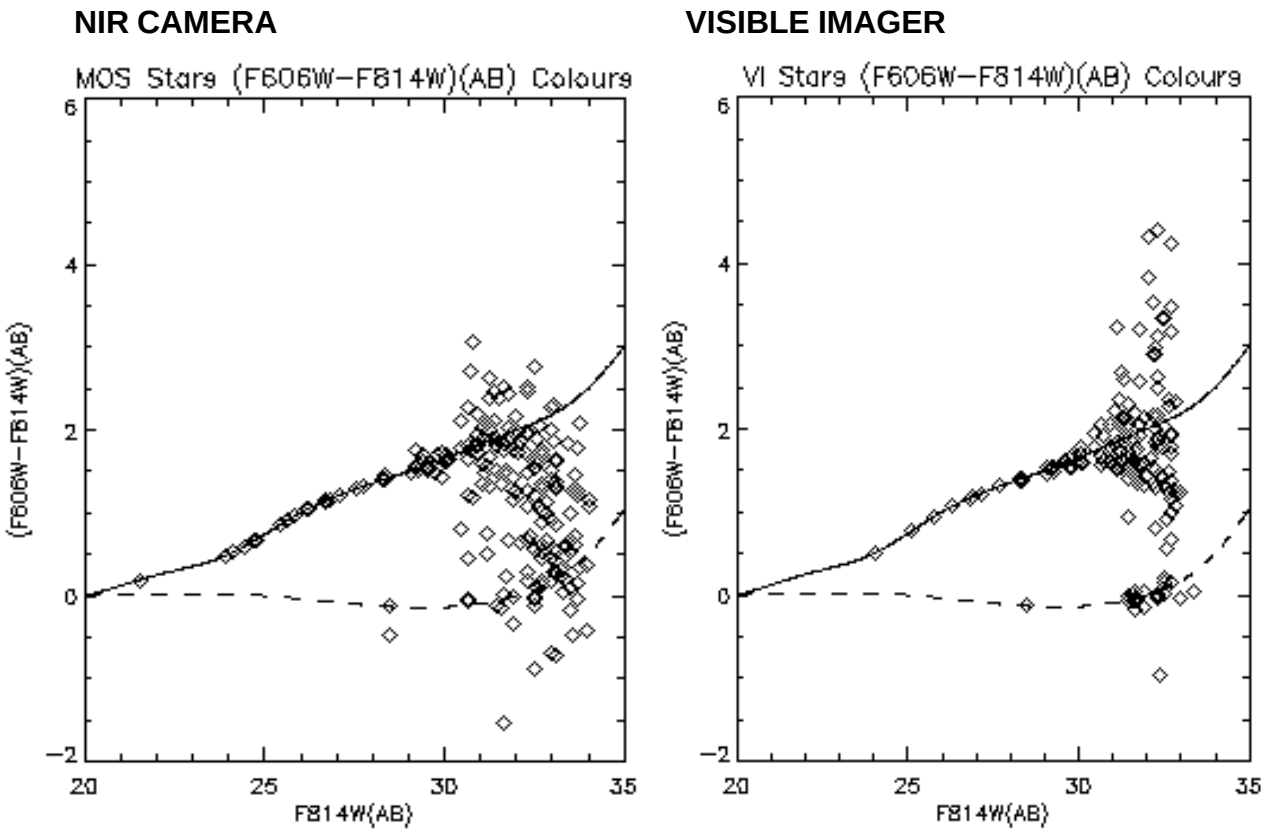


Figure 1.2-4. Comparison of stellar photometry. The left panel shows the results of photometry of stars in a simulated globular cluster image taken with the baseline NIR camera. The right panel shows the results for the visible imager. In each panel, 0.6 um – 0.8 um colours are plotted vs 0.8 um AB magnitude. With the Visible Imager, the white dwarf sequence (dashed line) can be clearly seen.

1.2.2 Science Programs

1.2.2.1 Galaxy Formation and Evolution

The study of the formation and evolution of galaxies is one of the key scientific programs of the NGST. While the emphasis is on infrared observations (because of the high redshift of the sources, and to take advantage of the much lower infrared background in space), there are good reasons why visible imaging is desirable, and even essential. In the sections that follow, we highlight the importance of visible imaging.

1.2.2.1.1 High-redshift Galaxies

Recent studies (Madau et al. 1996, Lilly et al. 1997) suggest that the star-formation rate in the Universe peaks at redshift $z \sim 2$, but many galaxies are now known which have redshifts greater than $z \sim 5$ (Hu et al. 1999, Fernández-Soto et al. 1999). With few exceptions, these high-redshift objects have been found

by virtue of the strong absorption of their ultraviolet spectrum at wavelengths shorter than rest-frame 121.7 nm, which corresponds to the Lyman- resonance line, by intergalactic hydrogen gas. A typical example of this is illustrated in Figure 1.2-5.

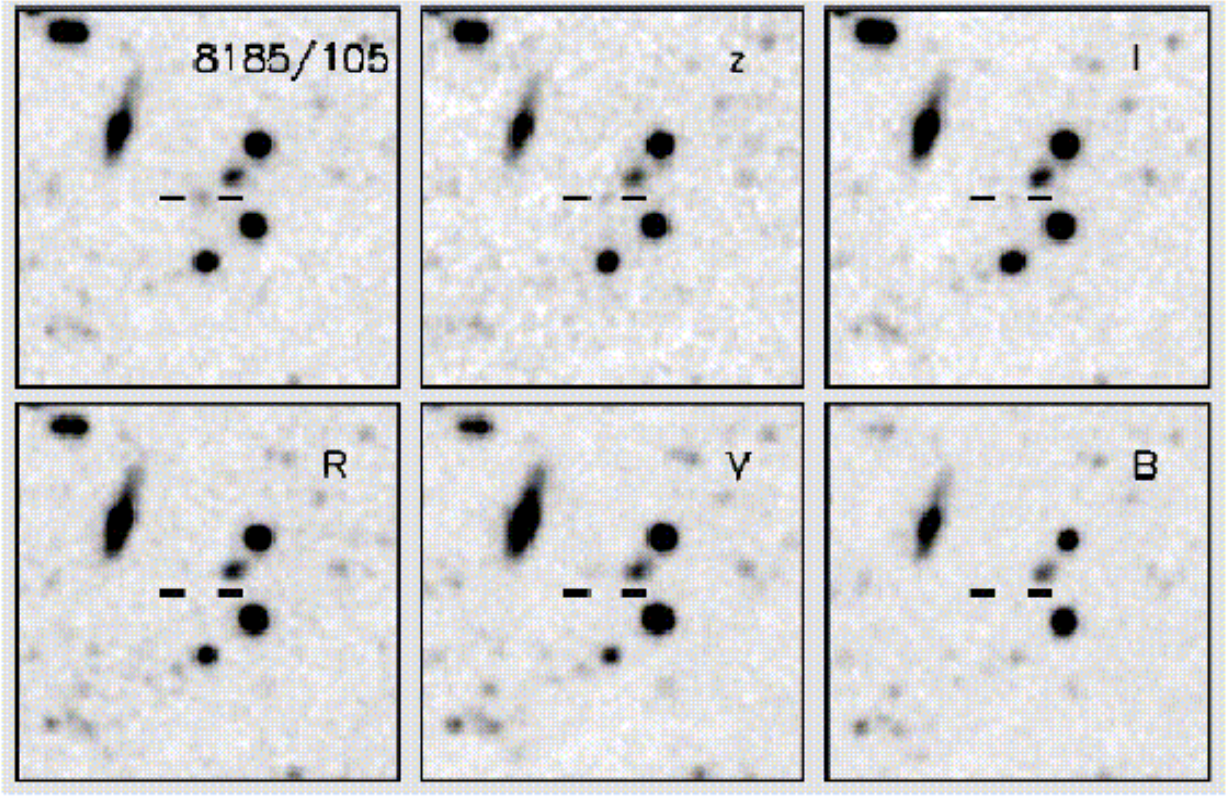


Figure 1.2-5. Visible and near-infrared imaging of a galaxy at redshift 5.7 showing ultraviolet blanketing (Hu et al. 1999).

In order to exploit this effect in searches for high-redshift galaxies, it is necessary to observe at wavelengths both shortward and longward of the redshifted Lyman- line. For redshifts in the range $z \sim 3 - 8$ this *requires* imaging at visible wavelengths. Access to the rest-frame ultraviolet spectral region is important not just for the detection of high-redshift galaxies, but also for their study. On the long-wavelength side of the Lyman- line, the continuum flux is produced primarily by hot stars whose energy distribution strongly peaks at 91.2 – 250 nm (see Figure 1.2-6). These stars play a major role in the early chemical evolution of galaxies, being the first contributors of processed elements (and eventually, dust) to the interstellar medium of young galaxies. The strength of the ultraviolet continuum can therefore be used to trace the star-formation rate over cosmic time. Moreover, only in the rest-frame UV spectral regime can we clearly observe the spectroscopic signatures of young stellar populations: analysis of the wind lines provides crucial information about the young massive star population in starbursts, such as the age of the burst and the slope and the upper mass limit of the initial mass function (Robert et al. 1993). These important diagnostic lines fall within the visible range for galaxies at redshifts between 2 and 5. While *integrated* spectra of star-forming galaxies at $z \sim 3$ have been obtained with the Keck telescope (Lowenthal et al. 1998), *spatially-*

resolved spectroscopy, with the Visible Imager equipped with gratings, will allow us to study the star formation *history* for distinct starburst knots within a given galaxy.

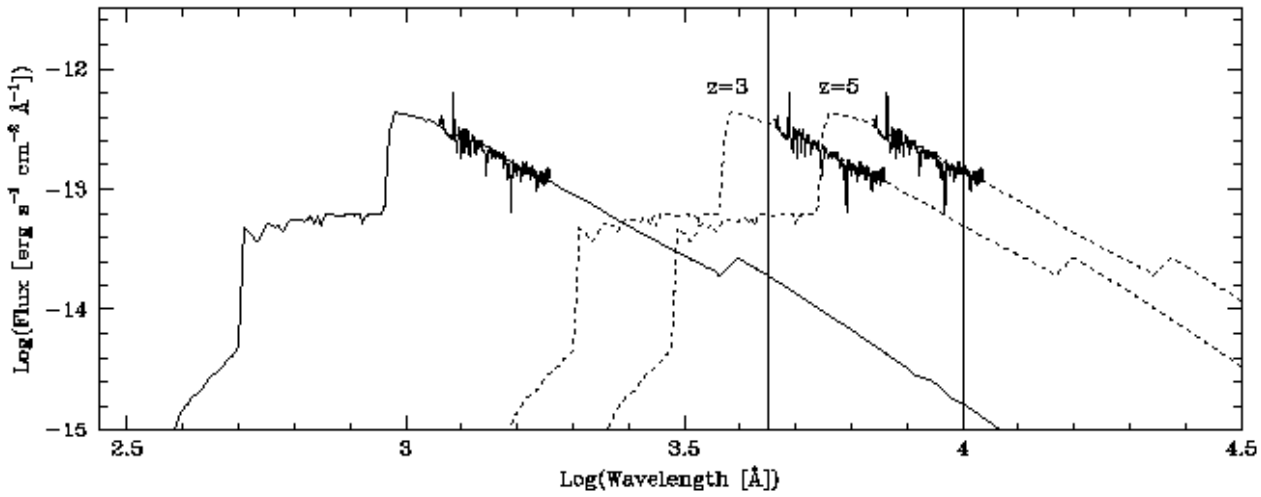


Figure 1.2-6. Energy distribution of a 6 Myr-old instantaneous starburst at different redshifts. The full SED was computed using stellar atmosphere models in an evolutionary synthesis code (see Leitherer et al. 1999). The UV spectrum observed for a starburst galaxy is superimposed on the model. The model and observations were shifted to 2 different redshifts (dotted lines). The range covered by the proposed optical imager is outlined.

1.2.2.1.2 Galaxy Morphology and Evolution

The optical morphology of galaxies can provide important clues to their formation process. HST imaging of high-redshift galaxies has shown a remarkable variety of unusual shapes and morphological features in galaxies at redshifts $z \sim 2$ - 3. Because of its fainter limits and higher resolution, NGST will be able to provide more detail on this phenomenon and extend the studies to higher redshift. Unreddened starburst knots in proto-galaxies at $z \sim 3$ - 8 will be a factor of 10 - 25 brighter in *V* than in the near-infrared. Given this high contrast, a direct comparison of deep *V*, *R* and *K*-band high-resolution images of distant galaxies will provide valuable clues about the influence of the starburst process on galaxy evolution.

If galaxies formed by the accretion of smaller "protogalaxies" as predicted by hierarchical clustering models, it may be possible to see this process in progress. Indeed, narrow-band surveys conducted at $z \sim 2$ - 2.5 in the Hubble Deep Field (HDF) have revealed compact, clustered, emission-line objects which may be galaxy "building blocks" (Pascarelle et al. 1996). Higher-resolution and better signal-to-noise ratio images are needed to elucidate the true nature of these objects. Their discovery highlights the importance of imaging with relatively narrowband filters in order to detect and study emission-line objects. Equipped with a continuously variable filter, a Visible Imager on NGST could detect such objects over a large redshift range in a single observation.

When star formation occurs at a very high rate in a dwarf galaxy, the kinetic energy generated by stellar winds and supernova explosions can deplete its host galaxy of a significant fraction of its gas content by expelling it into the intergalactic medium. The high resolution and large collecting area of NGST will be ideally suited to image this phenomenon, and the Visible Imager equipped with a series of tunable filters is necessary to study it in the "local" Universe z 1. Ideal targets are rich clusters of galaxies where the contribution of each galaxy to the enrichment of the intergalactic medium and the influence of the intergalactic medium on the ISM of individual galaxies could then be studied.

1.2.2.1.3 Active Galaxies

Nuclear activity in galaxies is of interest for several reasons: (1) It represents the most powerful source of continuous energy generation known, this offering a laboratory for physics of extreme environments; (2) It represents a major contributor to the ionization of the universe, particularly at earlier epochs; (3) Active galactic nuclei (AGN) are easily observed indicators of galaxy and star-formation at early epochs; (4) AGN are probes of absorbers along the line of sight by non-luminous matter.

The spectral energy distributions of AGN generally peak at rest UV or shorter wavelengths. NGST will be aimed at the study of objects in the redshift range 5 to 10, which places the observed peak radiation of AGN in the visible. In addition to being brighter in the rest UV, many key AGN emission lines occur in this region: O VI, Ly- , N V, Si IV, C IV, and He II. Thus, observations in the 0.5 to 1.0 um region are needed to study high-redshift AGN.

NGST observations of AGN will aim at understanding if galaxy formation via merging of star-forming protogalaxies is causally linked to the AGN activity or merely a contemporary activity at higher redshifts. With NGST, we will locate and make statistical studies of early epoch galaxies and protogalaxies. The fraction that have AGN and their properties will be a key part of this work, and should enable definitive answers to the origin of AGN. The initial formation and growth of galaxy nuclei is part of the early formation of galaxies, and study of high-redshift AGN will be a key program with NGST. The rest UV (observed visible) SED and line emission will be an essential part of this work.

1.2.2.2 Gravitational Lensing

In the past two decades, gravitational lensing has emerged as a major astronomical tool which has been applied to study the distribution of mass in the universe over a remarkable range, from planets to the largest clusters of galaxies.

1.2.2.2.1 The Mass Distribution in Galaxy Clusters

The largest individual mass concentrations in the universe are found in rich galaxy clusters. While clusters can be found and probed by means of X-ray emission from hot gas contained within them, only gravitational

lensing provides a direct measure of the amount and distribution of all gravitating mass within the cluster. In rich clusters, the gravitational field can cause strong distortion and multiple imaging of background sources and distorts the images of the many background galaxies seen in projection around the cluster (Figure 1.2-7). Analysis of this distortion provides information about the amount and distribution of dark matter in the clusters (Tyson et al. 1990.)

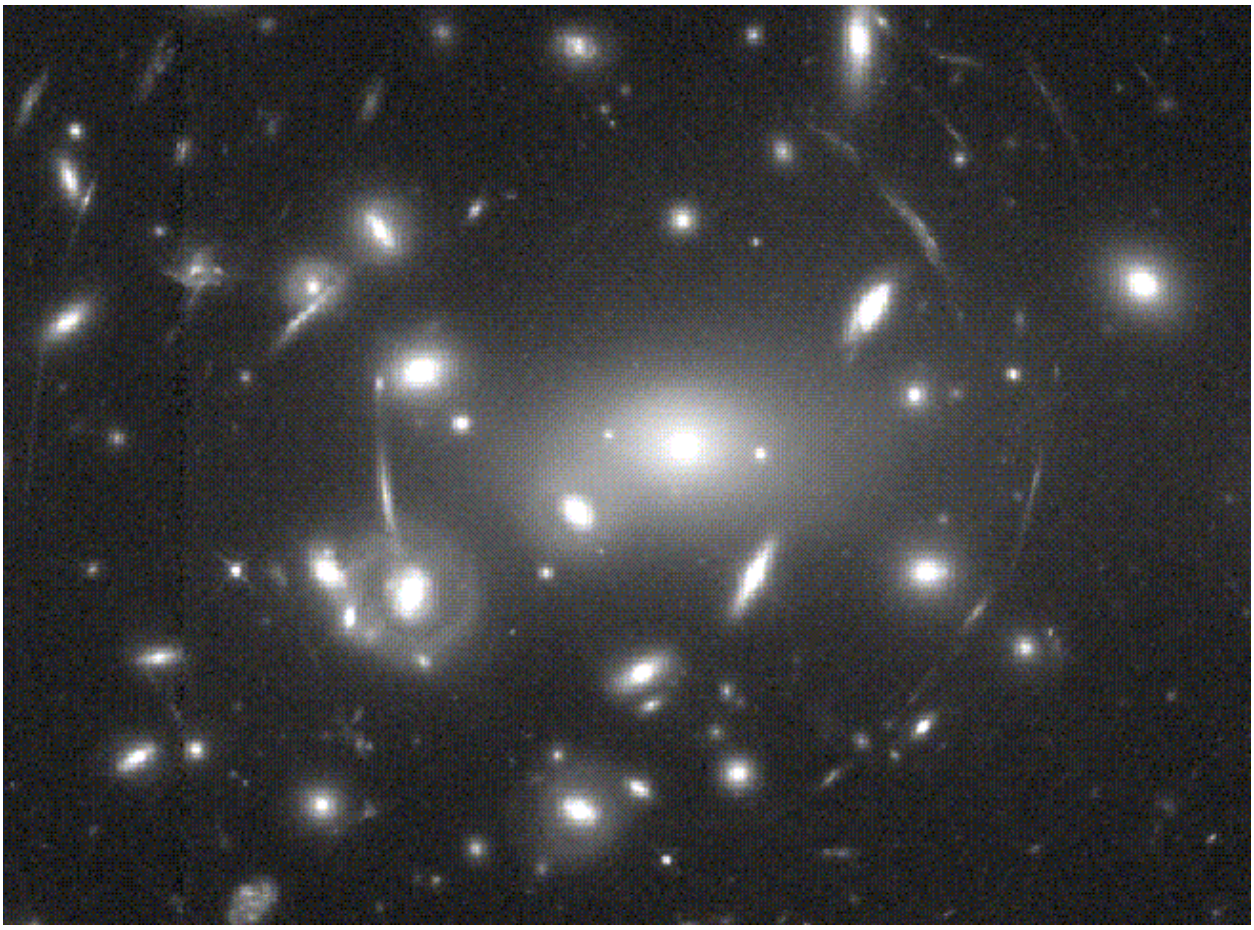


Figure 1.2-7. The rich cluster Abel 2218 showing effects of gravitational lensing (HST image courtesy of the Space Telescope Science Institute). Note the fine structure visible in the lensed images.

From the ground, these observations are difficult as high-resolution is critical in order to accurately measure the positions and morphology of the numerous small galaxy images. The consistent and uniformly-good image quality available over a large field of view from space makes the NGST an attractive instrument for such studies. Certainly, much work could be done using the NIR camera, but the best image quality will be available at visible wavelengths. Also, since most of the lensed galaxies are at moderate redshift ($z \sim 1 - 3$) they are brightest in the visible.

1.2.2.2.2 Microlensing

Gravitational microlensing has proved to be a very effective way to study populations of dark compact objects having masses ranging from those of planets to stars (Alcock et al. 1993). These studies have identified compact objects having masses of order a few tenths of a solar mass (M_o) as a major constituent of the dark matter in our galaxy. Gravitational microlensing can also distinguish binary lenses and even detect extra-solar planets (Pacz nski 1996).

A primary source of uncertainty in these studies is the fact that the distance to the lensing object is not known. One way around this is to observe galaxies in the Local Group and beyond (Crotts 1992). Such observations have not yet been conducted due to resolution limitations of ground-based telescopes. Higher resolution imaging provided by the Visible Imager on NGST would permit more than an order of magnitude increase in the number of stars that can be monitored.
The sources to be monitored are stars in nearby galaxies, which emit most of their light in the visible region of the spectrum.

1.2.2.3 Stellar Populations in Nearby Galaxies

1.2.2.3.1 White Dwarfs

White dwarf stars (WD) in the Milky Way Galaxy are of great interest as they are remnants of defunct stars that remain potentially detectable for all cosmic time. Since a WD undergoes predictable changes in brightness and color as it ages, an accurate census of these stars can trace the history of star formation in the Galaxy back to the earliest epoch. The Visible Imager on NGST will enable the study of the faint end of the white dwarfs luminosity function in the Galactic halo and globular clusters, which is very important for understanding both our Galaxy at early times and the galaxies currently seen at large redshifts.

The MACHO lensing surveys suggest that the halo contains unseen objects with halo kinematics and masses of order 0.5 M_o; White Dwarfs are obvious candidates. External constraints seem to require that the parent population of these objects was quite bizarre and unlike any population we see today. A deep survey with the Visible Imager will allow us to identify these objects by measuring their color, luminosities and proper motions. Using simple assumptions, a detection rate of about 20 halo WD per square arcmin is expected down to $V \sim 31$ (M_V 17.5 at 5000 pc). At a distance of 5000 pc or less, a typical halo velocity $v_T \sim$ 200 km s^{-1} would imply a proper motion 0.01 arcsec yr^{-1} (e.g. the size of a single pixel of the Visible Imager). Repeated observations of halo field over the lifetime of NGST will allow the precise determination of proper motions for the halo WD, as well as reliable (> 10) trigonometric parallaxes. Comparisons of the white-dwarf luminosity functions in the halo and among a well-chosen sample of globular clusters will provide an invaluable chronometer for the earliest stages of star formation and chemical enrichment of our Galaxy.

The $V - I$ color is optimal for distinguishing white dwarfs from M dwarfs, since it is only in this color that the real difference between these two populations shows up. Due to their comparatively blue colors, old white dwarfs will become progressively harder to detect at longer wavelengths.

1.2.2.3.2 Cepheid and RR Lyrae variables

The prospects for comparing the Population I (Cepheid) and Population II (RR Lyrae) distance ladders, as a function of galaxy morphology, galaxy luminosity, and metal abundance are enormous. Current studies of extragalactic Cepheids with HST are limited as much by crowding as by photon statistics. Because of its larger aperture and better resolution in the visible, NGST will allow an increase of a factor of 3 to 6 in the distances over which Cepheid variables can be studied, thereby increasing the accessible volume by a factor of 25 to 200.

Studies of extragalactic RR Lyrae variables would not be as limited by crowding as would the study of Cepheids, because these stars inhabit the low-density halos of galaxies. Exposure times for the study of RR Lyraes would probably have to be limited to ~1 hr, in order to avoid unacceptable phase smearing in these pulsating variables, whose periods range from 0.2 to 0.7 days. Even with that limitation, implying $V_{lim} \sim$ 30.0, RR Lyraes could be studied out to a distance modulus ~ 29.5 or a distance of 8 Mpc.

The amplitude of variation of Cepheids and - even more so - of RR Lyraes is strongly suppressed in the infrared compared to the optical; the variable star can be detected and its period can be determined *much* more effectively at visible wavelengths.

1.2.2.3.3 The Lower End of the IMF in Local Starbursts

The upper end of the IMF in a few nearby giant H~II regions has been studied in detail with HST/WFPC2 (Massey 1998), but these studies were generally limited to the most massive stars ($M > 6$ M_o) because of limited spatial resolution and sensitivity. The possibility of low-mass star formation in starburst regions has profound implications for many astrophysical problems: the formation of galaxies and their chemical evolution, the dynamical evolution of star clusters, the link between young super-star clusters and old globular clusters, not to mention the star-formation process itself. Very limited answers to this problem have been given by the HST/WFPC2 study of two nearby dense starburst cluster (NGC 3603 and R136). Definite answers will require a deep ($R = 30$) and detailed imaging survey (in R and I) of young starburst regions in different environmental conditions (metallicity, density, location within the galaxy and type of the host galaxy) in Local Group galaxies with a spatial resolution better than 0.1 pc (which corresponds to 0.03 arcsec at the distance of most Local Group galaxies). Although this study could be done in the near infrared, the high spatial resolution and large FOV provided by the Visible Imager are essential.

1.2.2.3.4 Imaging of Very-Low-Surface-Brightness Galaxy Disks

There is evidence from the existence of Low Surface Brightness Galaxies, from the star-formation rate in the very outer disk of spirals and from non-starbursting dwarf galaxies, of a continuous and quiescent mode of low star formation rate in galaxies. The best indication of this is the apparent existence of a pedestal at 1/40 solar in the abundance of oxygen in dwarf galaxies. This implies that there may exist a rich population of non-starbursting dwarf galaxies. These could only be found by photometrically-accurate deep imaging capability down to a surface brightness in *I* of 28 mag arcsec^{-2}. To strengthen this hypothesis, a program of very deep imaging should be conducted on star-bursting dwarfs to identify and measure a smooth underlying stellar distribution. NGST imaging in the R and I bands with very little background contamination is the prime instrument for conducting such a project.

1.2.2.3.5 The Immediate Environment of Nearby Supernovae

While Type Ia supernovae are now well understood, the origin of other types of supernovae (Ib and Ic in particular) are still unclear. Imaging the location of recent supernovae explosions with NGST in the optical will not only allow us to monitor their light curve for a longer time and with a much better precision than from ground-based telescopes, but it will also allow a detailed study of the immediate environment of the exploding star, providing important information about its age and mass.

1.2.2.4 The Outer Solar System

The outer Solar System has long been thought to be enveloped by a large halo of icy objects which are the source of comets . More recently, populations of objects orbiting the Sun beyond the orbit of Neptune, referred to as Kuiper-belt objects (KBOs) or Trans-Neptunian Objects (TNOs), have been postulated (Donnes 1997). The existence of TNOs has been demonstrated
(Jewitt & Luu 1993) and about 65 of these objects are now known. Their numbers, distribution and properties provide clues to the formation process of the Solar System. However, because of the difficulty of finding and studying TNOs even basic parameters such as their luminosity function remain poorly determined or controversial (Jewitt & Luu 1995).

NGST will be able to detect TNOs to a magnitude of $R \sim 31$ where the surface density will be such (Gladman et al. 1999) that we expect to find about 100 TNOs in a single exposure with the Visible Imager. Such observations will definitively establish the statistical properties of these objects and shed light on the formation and dynamical evolution of the Solar System. Because they shine primarily by reflected sunlight, efficient detection of TNOs requires imaging at visible wavelengths. The higher resolution and lower background provided by visible imaging will yield the faintest detection limits for these objects.

1.2.3 The Visible Imager and the Design Reference Mission

The Design Reference Mission (DRM) is a series of NGST proposals which serve as guidelines for the design of the observatory (http://www.ngst.stsci.edu/drm/programs.html). They also represent the core

of the science cases for the NGST. While most of the DRM proposals aim at the infrared regime, the following ones do require the Visible Imager.

Mapping the dark matter distribution at high redshift
V and *I* imaging are necessary for photometric redshift estimates and galaxy classification. High spatial resolution and large FOV are required.

Microlensing in the Virgo cluster and the role of baryonic dark matter in the Universe
High spatial resolution and especially very good sampling are necessary.

The formation and evolution of galaxies. I.
V and *I* imaging are necessary for photometric redshift estimates and galaxy classification. High spatial resolution and large FOV are required.

The formation and evolution of galaxies. III.
I imaging is necessary to observe the rest-frame UV (at 200 nm). High spatial resolution and good sampling are required.

The formation and evolution of galaxies. IV.
Rest-frame UV of targets is in the visible range. Good resolution and sampling, large FOV required.

The age of the oldest stars from the faint-end slope of the WD luminosity function in globular clusters
Targets are best studied in the visible. Good resolution and sampling, large FOV required.

The age and chemistry of the oldest stellar halo population
Targets best studied in the visible. Good resolution and sampling, large FOV required.

A survey of the trans-Neptunian region
Targets are best studied in the visible. Good resolution and sampling, large FOV required.

2 Engineering

2.0 Design Concept

From an understanding of the observational requirements of the visible science programmes it was possible to derive a set of instrument requirements which characterized the instrument needed to make the observations.

Table 2.1-1: Summary of Instrument Requirements

Field of View	2 x 2 arc min
Wavelength Range	0.5 - 1.0μm
Strehl Ratios	0.3 @ 0.5μm 0.7 @ 1.0μm
Angular Pixel Spacing	10 milli arc sec
RMS Read Noise per pixel	< 6 electrons
Fill Factor of Mosaic Array	> 90%

The visible camera data could also be used to support the mission directly as a Fine Error Sensor if a combination of diffusing element and broad band filter were used.

An initial survey of the available detector options was carried out as part of this study. After selection of a baseline detector format (18μm) the optical design which would meet the requirements was derived and the mechanical structure that would be required to provide physical support and controlled thermal environment was investigated (EMS Technologies 1999). It was found possible to accommodate the design within one quarter of the integrated structure of the GSFC NIR camera concept as indicated in figure 2.1-1. Alternative designs that could be accommodated outside of the NIR camera are possible and would require modification to details of the optical prescription developed for this configuration.

Figure 2.1-1 Visible Imager FPA Position Options

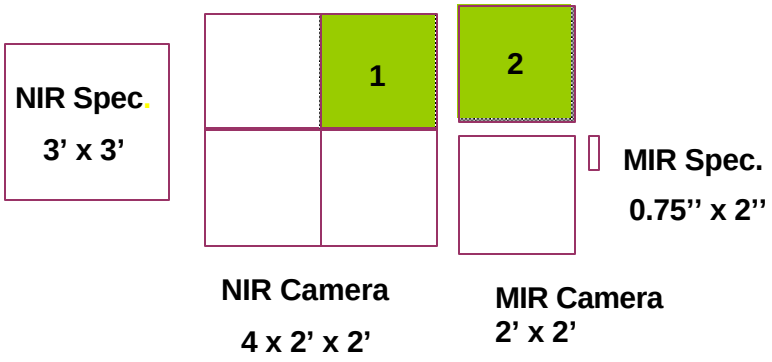
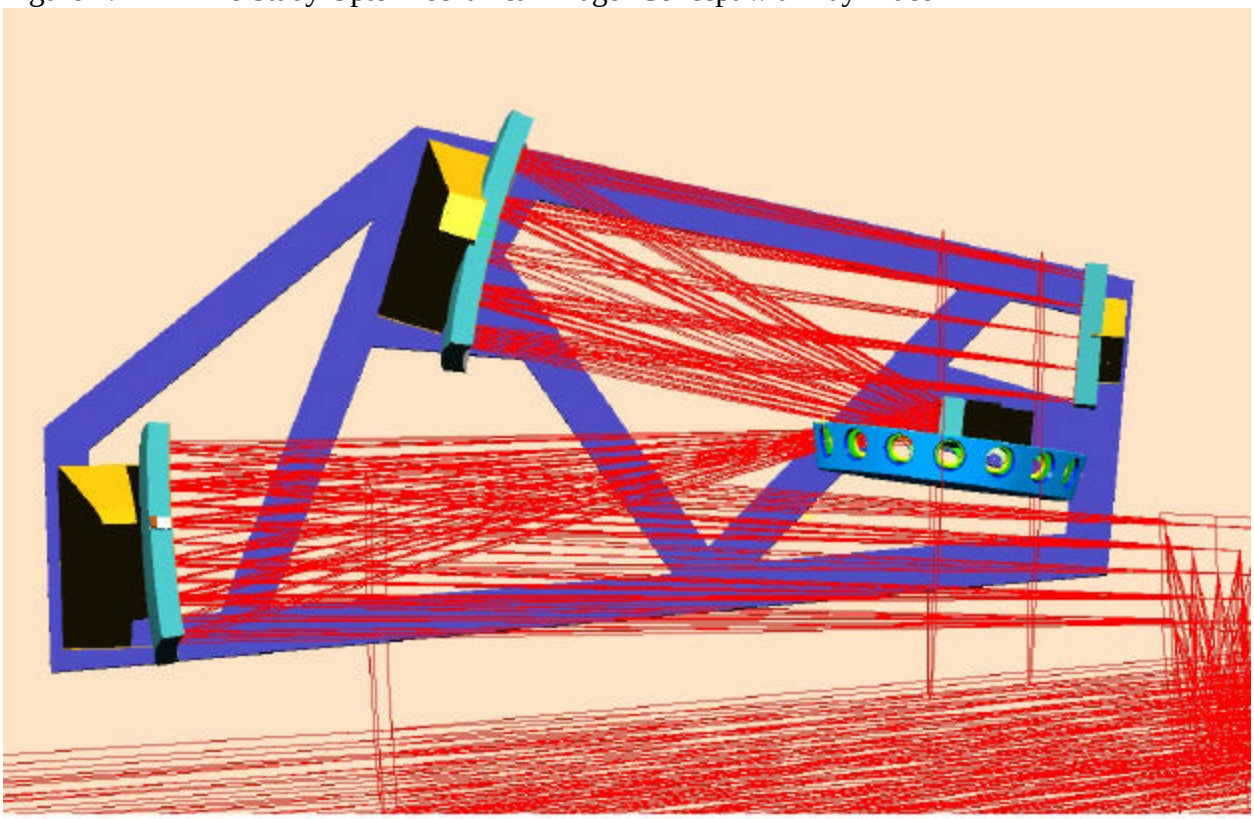


Figure 2.1-2 The Study Opto Mechanical Imager Concept with Ray Trace



2.1.1 Optical Design

The Visible Imager uses a similar layout to that of the GFSC NIR baseline, but uses a different optical design to meet the specific optical performance requirements.

- The visible imager uses a 237mm-square image sensor and the field-of-view has been mapped onto this size. This provides 10 milli arc second resolution over the 2x2 arc min field of view.
- The visible imager will receive a diffraction-limited wavefront at wavelengths as short as 600nm. The optical correction has been designed to be better than that of the NIR imagers such that the visible imager has diffraction-limited performance over the wavelength range 0.5 to 1.0 microns.

2.1.1.1 Paraxial Characteristics

The NGST and its associated imagers have very little symmetry. To describe this type of system it is best to restrict attention to the real geometry such as the fields of view, image size and shape, and pupil imagery.

The object-space fields of view used to design the visible imager were the corners of an inner and an outer square (1’ and 2’ in size respectivly); the final point was at the center. The combination of points was chosen both to define the field-of-view of the instrument and to assist in achieving optical correction over the field.

These field points are off-axis, with respect to the main telescope. The outer square maps onto the

perimeter of one of the facets of the pyramid structure within the GSFC ISIM.

The stop image was located, in a position between the imager mirrors, free of mechanical obstructions and away from the other light bundles.

- The stop provides a good location for filters and grisms. A filter at this point provides reasonably uniform filtering to the whole image.
- A single aperture at the stop surface can be simultaneously a good fit on all the image-forming ray bundles and is an important part of the strategy for controlling veiling glare.
- The visible imager is similar in size to the current imager modules and fits within the existing ISIM NIR camera package.
- The mirrors have a common optical axis about which the surfaces are axially symmetric. The mirror apertures, however, are decentered and non-circular.
- None of the mirrors has a common surface with any other mirror. Each mirror has a different surface form, with the pole of the surface at a different point along the common axis.
- Two of the mirror surfaces are conic-sections of rotation. The remaining mirror is an axially-symmetric asphere. No surfaces are spherical.

Mechanically, the design is largely planar. In this respect, it resembles the existing imager modules and lends itself to a similar mechanical structure. Alternatively, the system is compatible with a space-frame or a chassis-style mechanical design cast, or fabricated from plates. . An all aluminum construction is compatible with the chassis approach and a zerodur / composite construction would be compatible with the space frame approach.

2.1.1.2 Optical Correction

The required diffraction-limited performance has significant implications for the following issues:-

- The nominal, design optical correction of the NGST and visible imager combination
- The manufacturing, assembly and alignment tolerance budget for the NGST and the visible imager
- The environmental budget for items like the departure from isothermal behaviour or geometrical and mechanical changes during the life of the instrument.
- The performance budget for dynamic parts of the system like the wavefront sensor and corrector or any powered alignment and focus devices

At this stage only the optical correction for the nominal design was considered.

Figure 2.1-3 shows the wavefront aberrations of the system for a wavelength of 600nm. The wavefront aberrations are consistent with a diffraction-limited instrument . Indeed, some fields are very well corrected. Others, however, could leave a little more margin for the tolerancing, environmental and dynamic issues that will also degrade performance later.

The wavefront plot gives useful insight into the aberrations of the system. A better quantitative measure of the optical correction for well-corrected systems is the RMS wavefront error. For small aberrations, this value has a simple relation to the *Strehl* ratio.

A *Strehl* ratio of 0.8 is commonly taken as the threshold of diffraction-limited behaviour. This value corresponds to a RMS wavefront error of 0.074 waves

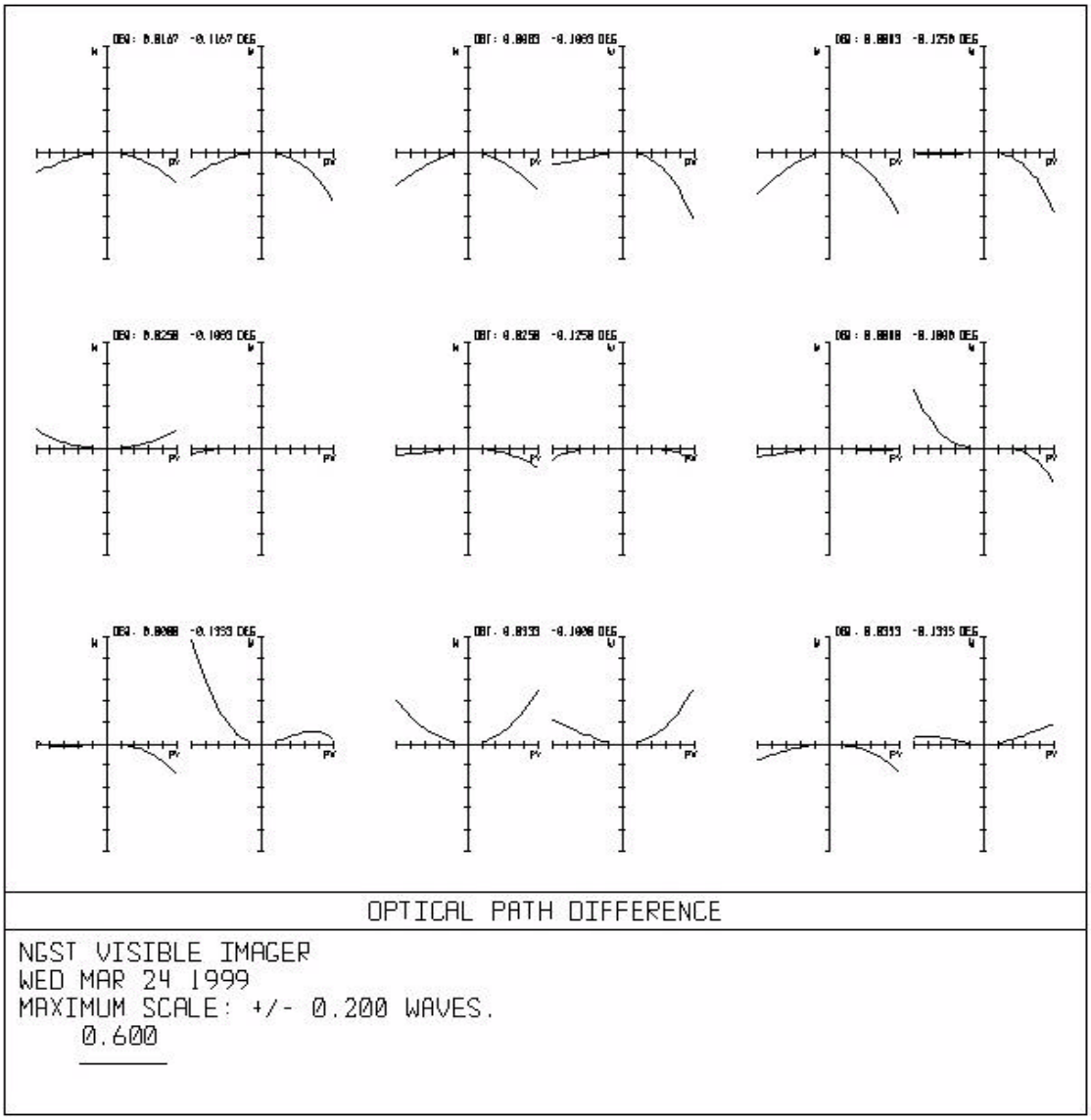


Figure 2. 1-3 Wavefront Aberrations of NGST with the Visible Imager

Fields 1, 2, 3, 4, 5, 8 and 9 leave most of the 0.074 wave budget available for the post-design degradation mechanisms that will occur. Fields 6 and7 consume half of the budget and may need some additional work as the other degradation issues become better quantified.

2.1.2 Operation at 30K

The visible imager thermal and structural performance at 30K has been examined. Assuming the preferred all aluminum fabrication it was found that the Imager would scale almost isotropically as it cools and such a change has no effect on the image quality. In practice, there will be some departure from isotropic scaling caused by these kinds of issue:-

- The image sensor, grism and filters will not share the thermal characteristics of the mirror and structural materials.
- The mirror and structural materials may be nominally the same, but their alloy composition, heat treatment, work-hardening history, etc will all be subject to some tolerance budget and vary from place to place in the instrument. Thus, the thermal characteristics will also vary across the system.
- The instrument will be subject to thermal gradients. There will be non-uniform heating and cooling of the instrument and parts of the system will dissipate power through the structure.

These problems can be reduced, to some extent, by active compensation during the mission. Tilt and displacement of the image sensor is a good candidate for adjustment.

The following images show the thermal gradients present in the structure and the associated distortion for a worst case FPA dissipation of 200mW. The distortion is well within that allowed by the available wavefront error budget.

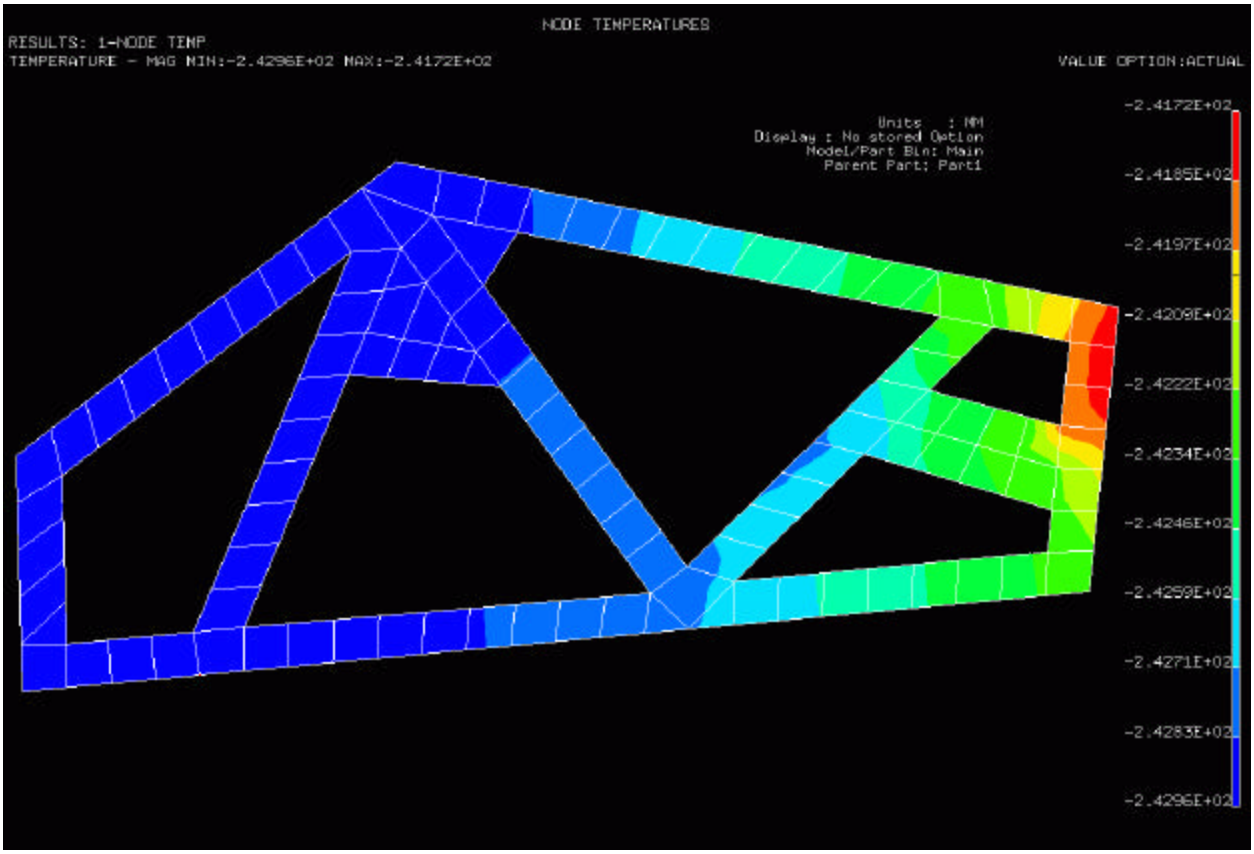
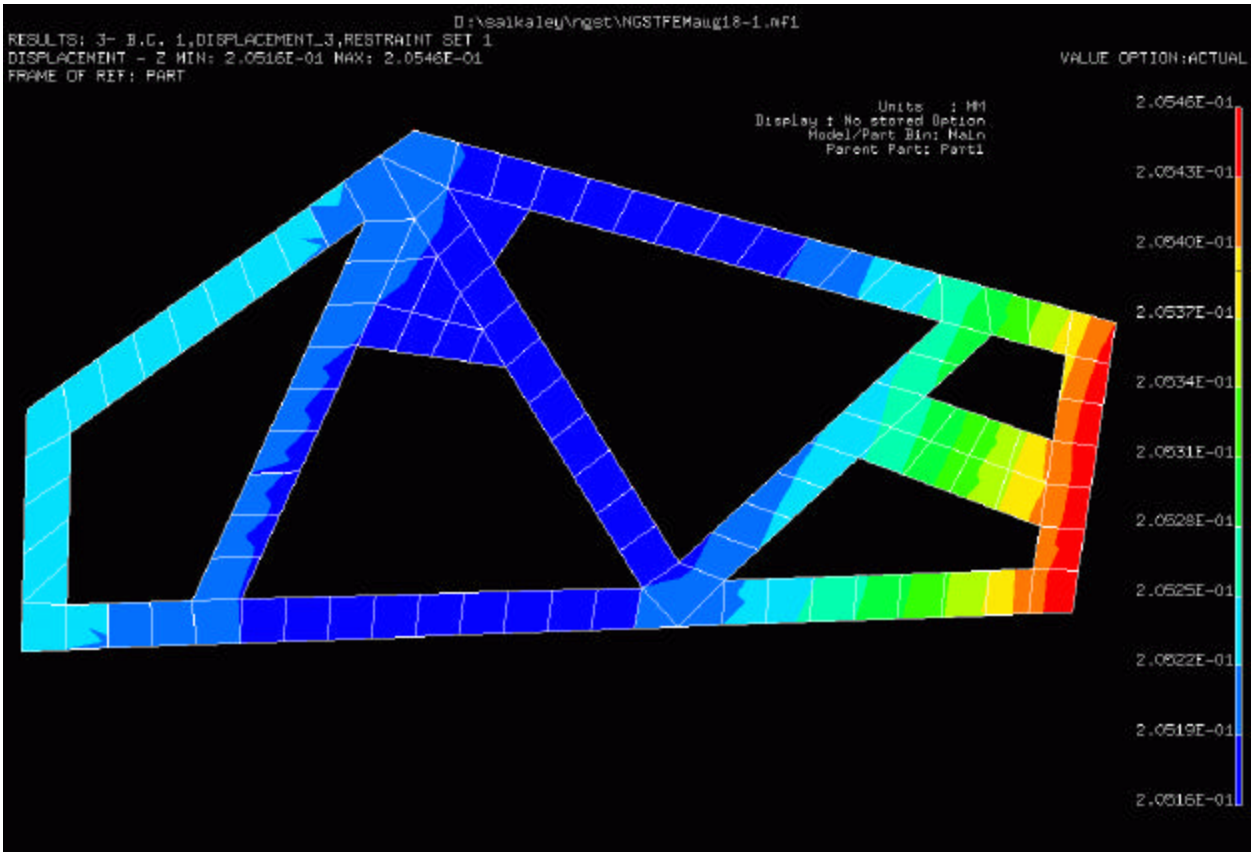


Figure 2.1-4 Temperature Gradients

Figure 2.1-5 Mechanical Distortion



2.1.3 Detectors

The following candidate technologies were investigated, in terms of their capabilities, and limitations, problems for development & fabrication, etc. This assessment is strongly based in what detector technology is currently available or will be available in the time frame during which design decisions for the NGST instruments must be made (1 - 2 years hence). It will obviously continue to be essential to track the ongoing development of all the candidate technologies as future instrument studies progress.

The format, pixel size and the other issues related to the choice of array detector for the NGST visible imager can be summarized in the following table.

Detector	Max Format	Pixel Size	QE (0.5- 1µm)	Noise	Issue
CCD	9k x 9k	7 – 25 µm	0.85 – 0.1*	< 6 e-	drivers, radiation
InSb	2k x 2k	27 µm	>0.70	>30 e-	pixel size, noise
HgCdTe	2k x 2k	18 µm	0.75	~20 e-	noise, dc uniformity
Si p-i-n	2k x 2k	18 µm	0.85 – 0.25	~20 e-	QE @ 1 µm
CMOS	1k x 0.8k	< 13 µm	0.85 – 0.1	~10e-	fill factor, format

Table 2.1-2 Detector Options

* Note: modeling of high-resistivity CCDs shows that quantum efficiencies above 0.8 should be feasible at 1 µm

2.1.3.1 Detector Baseline

For the purposes of developing a baseline instrument concept it was assumed that either a thinned HgCdTe detector or a Si p-i-n diode detector on a HAWAII-2 multiplexer would be available for the NGST Visible imager. The pixel size of 18 µm and the requirement of near diffraction limited imaging across the field of view drove the optical and mechanical designs to the maximum feasible physical sizes. It would not be possible to meet both FOV and resolution requirements simultaneously with a larger pitch pixel. Unless the visible imager were modularized as per the GSFC NIR Camera design.

The InSb detectors have two major drawbacks for visible work;

The pixel size is too large to allow simultaneous compliance with both field of view and resolution requirements.

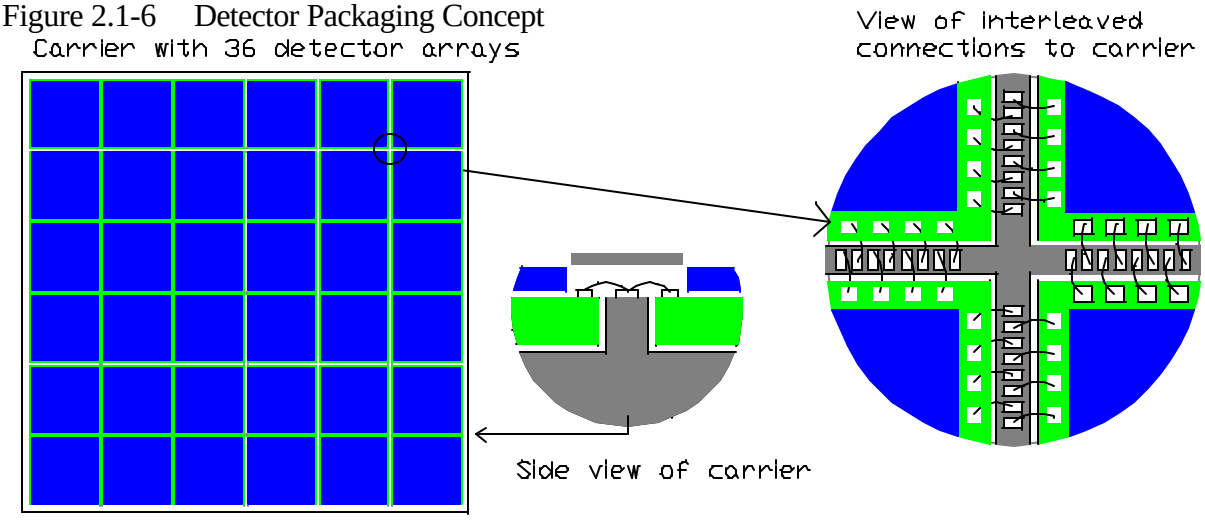
There is the presence of "black holes" in the visible data when the InSb detectors are cooled to low temperatures. The mechanism operating to cause this defect is unknown and not currently being investigated.

CCD's are regarded as not suitable for the NGST application for numerous reasons; reliability, charge transfer efficiency, power required to drive the readout, higher than 80K operating temp requirement etc.

2.1.3.2 Detector Packaging Concept

In order to provide the large field of view at the high resolution required, it will be necessary to provide a mosaic array of 36 detectors. Although there are yield issues with the hybridization of detectors it is believed that such an array is a reasonable proposition within the development timescale allowed. It is proposed to package the detectors in a custom ceramic (TBC) carrier. This carrier would allow a 1.5mm spacing but the multiplexer using interleaved interconnects (wirebonds). The detector spacing would be (1.5 +2.08mm), allowing for the nominal HAWAII-2 dimensions. No multiplexer redesign is proposed, since 87% fill factor (close to the 90% requirement) would be achieved.

Figure 2.1-6 Detector Packaging Concept



2.1.4 Electronics

The full electronic system required for the operation of the visible imager has been examined during the course of the study.

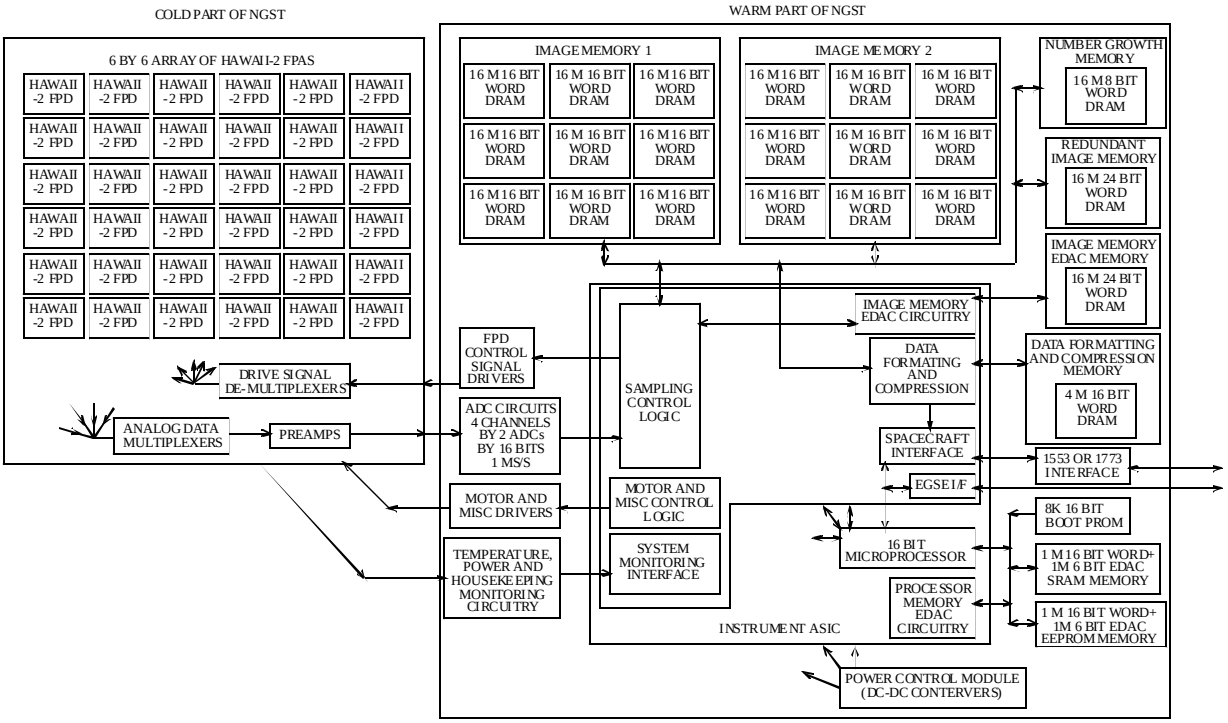


Figure 2.1-7 Electronics System Block Diagram.

2.1.4.1 Cold Side Electronics

The cold part of NGST will consist of the following:

1. A 6 by 6 array of HAWAII-2 Focal Plane Detectors (FPDs)
2. A drive signal de-multiplexer,
3. Analog data multiplexers and
4. Analog data preamplifiers.

36 FPDs will be mounted as an array of 6 by 6. The gaps between the 36 FPDs will be minimized by mounting the bare die on a carrier. The gaps between adjacent FPDs will be no more than 3mm. Each FPD supports a 2048 by 2048 pixel area with a 18 um pixel pitch. Thus a 6 by 6 array of HAWAII-2 FPDs will provide an FPA (Focal Plane Array) of 150 M Pixels

Each FPD consists of 4 quadrants which are totally independent of one another and hence can be scanned independently. However, the control signals of the 4 quadrants on each FPD are tied together to reduce the total number of control signals between the warm and cold parts of NGST. This results in each of the 4 quadrants being scanned together in an identical manner to provide 4 pixel outputs in parallel.

Figures 2.1-7 and 2.1-8 are drawn to show the simultaneous scanning of 4 FPD quadrants (within a single FPD). However, 4 quadrants from 2 FPDs could be scanned together to produce an 8 channel output, 4 quadrants from 4 FPDs for 16 channels, etc.

Short exposures will require fast readout rates in order to provide the highest exposure time resolution. Long exposures will tend to be of weaker signals and hence a slower readout would be used to obtain a lower self glow against the weaker image.

It is proposed that:

1. The de-multiplexer FET gate drivers are separated from the FET switches so that the FET gate drivers may be placed on the warm part of NGST and the FET switches on the cold part of NGST. This will ensure minimum dissipation on the cold part and
2. The FPD clock drivers be placed on the warm part of NGST and be passively switched between FPDs by the de-multiplexer so that the FPD clock driver dissipation is also placed on the warm part of NGST.
3. Differential signals will be required over the cold/warm interface to achieve low noise performance.

Details of the proposed I/F between cold and warm electronics is shown in figure 2.1-8

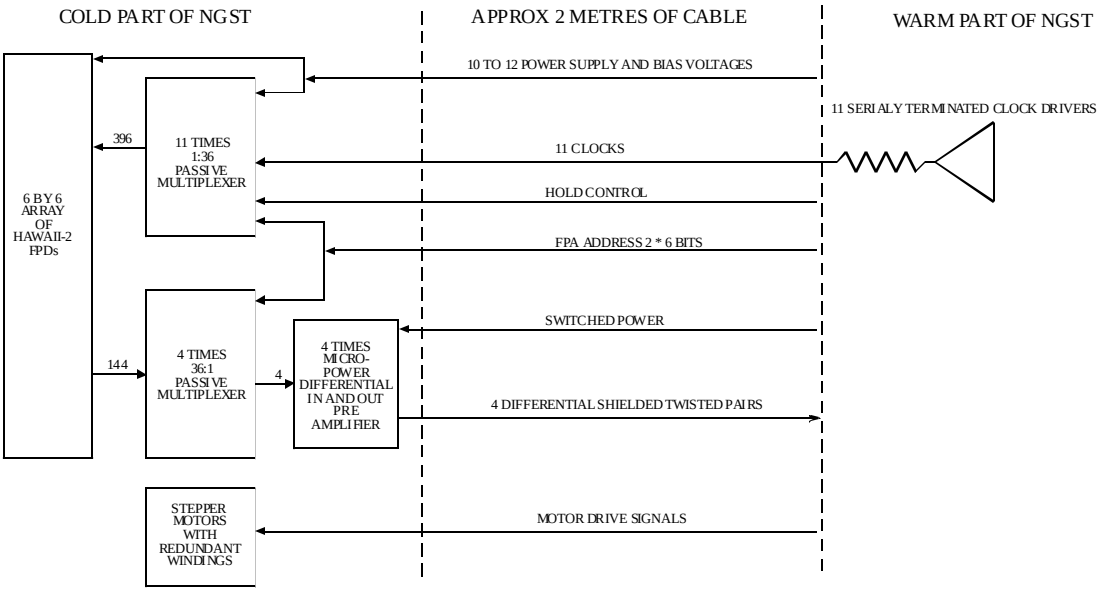


Figure 2.1-8

2.1.4.2 Warm Side Electronics

The Warm side electronics consists of the following basic blocks;

Instrument ASIC
- Sampling control logic
- 16 bit processor
- Motor control
- Memory control

Memory Blocks
- Image Memory #1, #2
- Number Growth Memory
- EDAC Memory
- Data Formatting & Compression Memory
- SRAM
- EEPROM
- Boot PROM

ISIM 1553 or 1773 Interface
Power Control Module
FPA control and signal processing
Engineering data circuitry

2.1.4.2.1 Image memories

There are 2 image memories. At any given time one is available for the capturing of data from the FPA for the current integration period while the other is available for data compression, formatting and down linking to earth of data from the previous integration cycle.

Each image memory consists of an array of 3 by 3 of memory blocks. Each of these blocks supports 4 HAWAII-2 FPDs. Thus, each memory block is associated with a 2 by 2 area of the 6 by 6 array of FPDs of the focal plane.

2.1.5 Data Sampling

Data sampling techniques have been studied and the recommended solution is to use Fowler sampling as shown pictorially in figure 2.1-9 as the baseline.

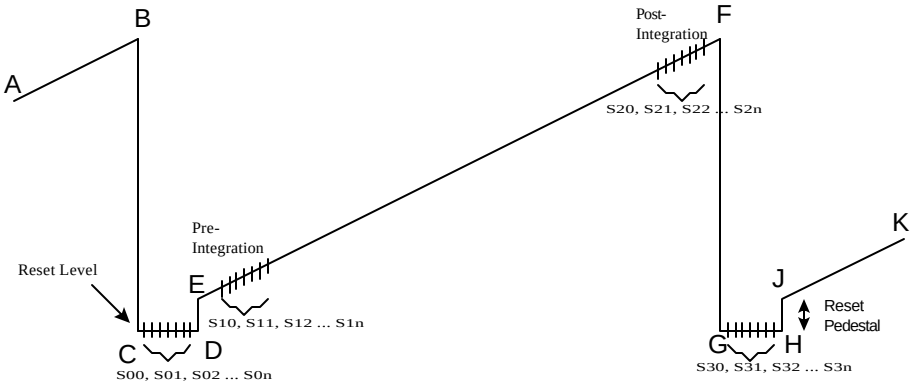


Figure 2.1-9 Pictorial Representation of Fowler Sampling

An enhanced technique, EMS term " Correlated Triple Fowler Sampling" may be required during long duration observations of faint sources to control 1/f noise sources. The technique would use multiple samples of the reset level before and after integration to provide a baseline.

2.1.6 Data Compression

The data compression should provide the maximum compression ratio whilst being lossless. The largest compression ratios are likely to be achieved by the following:

1. Using a lossy compression algorithm for the maximum compression ratio,
2. Calculating the resulting error,
3. Huffman coding the lossy compression parameters and error data and
4. Transmitting the resulting data to the ground.

For example assume one has an integer image which is N bits deep and one concludes that the final M bits are just noise. One may strip off the final M bits, and compress the (N-M) most significant bits in a lossless fashion. The final M bits are essentially incompressible. If the compressed version of the first (N-M) bits were transmit down, along with final M bits uncompressed. The original image can be reconstructed the in a loss-less way.

For a 12-bit image. Normal compression of the whole 12 bits, typically yeilds a compression factor of two. However, if the final two bits are stripped of, a compression the first 10 bits by a factor of 5 may be achieved. So the total compression factor for the entire image is

0.2 * (10/12) + 1 * (2/12) = 0.3333 this is effectively a completely loss-less compression of a factor of three. If the last three bits were removed, and the first nine bits compressed by a factor of 100. Then

0.01 * (9/12) + 1 * (3/12) = 0.2575 for a total compression of approximately 4.

Automatic algorithms exist for doing statistics on the contents of a given image and determining the optimum

number of bits to strip off and send uncompressed.

On the other hand, for a given scientific purpose it might be concluded that the last M bits (almost completely noise) are not required at all, and send down only the compressed version of the first (N-M) bits, for a slightly lossy compression factor of (in the first example above, M = 2) 0.2 * (10/12) = 0.166 implying a compression factor of 6, with the loss of the information contained in the least significant 2 bits.

2.1.7 Summary of Instrument Characteristics

	Optics Unit (cold)	**Electronics Unit (warm)**
Volume	1.5 x 0.6 x 0.4m	0.3 x 0.3 x 0.3m
Mass (inc 30% margin)	45 kg	8kg harness tbc
Power (inc 50% margin)	200mW (max) 40mW (average)	40W (active) 5W (quiescent)
Data rate (typical, no compression)	-	~ 19 Mbit per sec when observing (even when on long exposures, assumes 16bit measurement and Fowler sampling)

Table 2.1-3 Instrument Characteristics (all values tbc)

2.2 Technology Readiness

The Visible Imager is by intention a very simple instrument. There are no major technology developments required. Those developments that are necessary;

2.2.1.1 Detector Technology Readiness

Silicon p-i-n diode arrays are the preferred option for the Visible Imager because they represent straightforward technology. Rockwell Science Center is currently producing 640 x 480 Si p-i-n diode arrays on a commercial basis. RSC's concurrent development of 2048 x 2048 HAWAII-2 multiplexers also demonstrates technology readiness. Nevertheless, Si p-i-n diode arrays have yet to be mated to HAWAII-2 multiplexers and cooled to 30 K to measure performance. There may be issues with respect to the HAWAII-2 multiplexer performance at 30 K given the commercial processes used in their fabrication. There is little data on the performance of HAWAII-2 predecessor devices at temperatures below 40K. Cryogenic testing of a Si p-i-n diode array on a HAWAII-2 multiplexer is an obvious step that can be taken within the next year to further demonstrate technology readiness.

Should the on-going instrument configuration selection dictate the use of InSb detector arrays, either in their

current 1024 x 1024 format or larger, the current NASA sponsored detector developments should rapidly be directed to extensively evaluate the visible performance of these devices at 30K. The small amount of visible light InSb testing to date has revealed significant photo-response non-uniformities which are temperature sensitive and are temporally unstable. These non-uniformities, which take the form of localized reductions in quantum efficiency, some greater than 50%, are thought to be due to charge trapping in the surface layers of the InSb detector material. Visible photons are preferentially absorbed close to the surface, so that these effects are not present in the infrared (>1 μm) performance of the devices. Unless significant effort in the short term is expended in the resolution of this issue, the InSb devices will not be suitable for 0.6 to 1 μm imaging on NGST. It is instructive to consider an analogous issue in HgCdTe detector arrays, the persistence effect. This effect was identified over 10 years ago and it too is thought to be due effects in the surface trapping of charge. While the exact mechanism is likely not related to the InSb photo-response issues, it is significant that this persistence effect still has not been resolved.

2.2.1.2 Mirrors

The large off axis mirrors will probably require specialized equipment to perform the initial cutting for aluminum, polishing equipment already exists for such mirrors. For the zerodur option both cutting and polishing equipment ahs been identified Again this problem is common to the GSFC NIR camera and could change if the mission Prime contractors change the configuration for a more on axis system (as has already been proposed by one of the two teams)

2.2.1.3 Piezos

The piezos proposed for the focus mechanism are again the same inch worm devices being examined by NASA for the control of the main mirror and adaptive optic.

2.2.1.4 Data Compression

One area that does require further work prior to the start of the Phase A is the investigation of the use of slightly lossy data compression to limit the amount of data the instrument needs to transmit to ground. The science team have indicated that this type of approach is acceptable but a series of data simulations should be run and the results assessed.

2.3 Development Schedule and Integration and Test Plan

Proposed Schedule

01/2000 Technology developments

04/2001 Phase A Start

05/2002 PDR STM Delivery

06/2003 CDR EM Delivery

06/2005 PFM Delivery

2.3.1 Pre Phase A

During the Pre-Phase A period it is recommended that the NASA led developments outlined above be tracked. It has been stated at meetings with NASA that the detector development issues will be undertaken by the agency. It is of concern that the issues highlighted in the previous section are not yet being tackled.

It is recommended that work on lossy algorithms for data compression be further explored by exploiting the NGST simulation developed by DAO to test the scientific acceptability of various scenarios.

2.3.2 Phase A/B

Assuming that the technology development work is carried out by the NASA funded teams and this data is made available then the instrument team will be well positioned to move forward in Phase A to complete the design trade-offs and select a baseline design.

It is proposed that rather than follow the simple STM, EM, PFM model philosophy suggested by the NASA proposal that a more conservative approach be taken with respect to the "Optical Head" (all the equipment to be housed on the cold side of NGST). It is recommended that either the STM of the Optical Head be enhanced sufficiently that the Qualification of the structural design be completed prior to CDR or alternatively the EM of the optical head could be used for this mechanical qualification. This will allow the optical head of the Flight Model to be Acceptance tested only and thus subject to the minimum levels of stress possible. It is felt that this approach, whilst making the first two years of the project more difficult will retire a significant amount of risk on the design of the optical head. It is proposed that the EM electronics be limited to an electrically representative version only, ASICs being implemented by FPGAs etc. It is felt that because of the relatively benign environment on the warm side of the spacecraft that this is an acceptable risk.

Breadboards Optics (a full set of mirrors in realistic structure for optical testing)

Front end electronics (a small number of detectors to allow development of examples of electronic circuitry especially for operation at cryogenic temperatures.)

EM / STQM It is proposed that a full optical and structural model be made and mechanically qualified.

If detectors are available it would retire many risks to equip this model with a full FPA and readout set of electronics to allow cryogenic testing at the instrument level.

The electronics unit at this stage could be electrically representative only

2.3.3 Phase C/D

Given a successful Phase A/B a two year Phase C/D to provide a PFM model electronics unit and FM Optical Head seems a reasonable proposition. Ultimately success will depend on good communications within the instrument team and spacecraft teams and effective definition of interfaces and software. This will be especially true if the proposed level of instrument integration within the ISIM come to fruition.

FM / PFM The Optics unit would be FM standard and only need acceptance testing.

The electronics unit could be built as a PFM and qualified to protoflight levels as it is to be flown in a relatively benign environment.

3 Cost Estimate

Submitted to CSA only

4 References

Alcock C. et al. 1993, Nature, 365, 621
Crotts, A. P. S. 1992, ApJ, 399, L43
Dones, L., in *From Stardust to Planetesimals*, Y. J. Pendleton & A. G. G. M. Tielens Eds., ASP Conference Proceedings, 122, 347
Gladman, B., Kavelaars, JJ, Nicholson, P. D., Loredo, T. J. & Burns, J. A. 1999, AJ, in press (astro-ph/9806344)
Hu, E. M., McMahon, R. G., & Cowie, L. L. 1999, Apj, 522, L9
Jewitt, D. & Luu J. 1993, Nature, 362, 730
Jewitt, D. & Luu J. 1995, AJ, 109, 1867
Fernández-Soto, A, Lanzetta, K. M., & Yahil, A. 1999, ApJ, 513, 34
Leitherer, C., et al. 1999, ApJS, 123, 3.
Lilly, S. J., LeFevre, O., Hammer, F. & Crampton D. 1996, ApJ, 460, L1
Lowenthal, J., Simard, L., & Koo, D. C. 1998, in The Young Universe, ASP Conf. Ser. 146, p. 110
Madau P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996, MNRAS, 247, 1P

Massey, P. 1998, in The Stellar Initial Mass Function, ASP Conf. Ser., vol. 142, p. 17
Paczýnski, B. 1996, ARAA, 34, 419
Pascarelle, S., Windhorst, R. A., Keel, W. C. & Odewahn, S. C. 1996, Nature, 383, 45
Robert, C., Leitherer, C., & Heckman, T. M. 1993, ApJ, 418, 749
Tyson, J. A., Wenk, R. A., & Valdes, F. 1990, ApJ, 349, L1
EMS Technologies Canada, NGST Visible Imager Study Report, EMS-RP-1054-XXX, Oct 1999

Appendices

A Visible Wavelength Detectors for NGST

A.1 Introduction

We have considered several detector technologies for use in the NGST Visible Imager, including charge coupled devices (CCDs) and a variety of hybrid detectors based on photodiodes and CMOS multiplexers. These were evaluated according to the requirements for the NGST Visible Imager, which are as follows:

- angular field of view at least 2.0 arcmin square
- fill factor of at least 90%
- pixel spacing of order 10 milli arcseconds (resulting in a total array size of 12000×12000 pixels)
- quantum efficiency (QE) greater than 80% over the entire visible range (0.5 – 1 um)
- rms read noise per pixel less than 6 electrons
- operable at, or close to, the ambient instrument support structure temperature (~ 30 K)

Each of the detector types and its suitability for use on NGST is described in detail in the following sections. We consider CCDs first and then the various hybrid detectors. Although there are many possible materials for the photosensitive material in a hybrid detector, we look at three. Indium antimonide (InSb) and Mercury Cadmium Telluride (HgCdTe) are both candidates for near-infrared detectors on NGST and appear to have reasonable performance in the visible wavelengths. Silicon p-i-n material has roughly the same wavelength responsivity as a CCD, due to the silicon bandgap.

A.2 CCDs

Charge coupled devices (CCDs) are currently the standard astronomical detector for imaging and spectroscopy in the visible wavelength range. They rapidly took the place of photographic film at all major astronomical observatories because of their superior sensitivity, low noise, high linearity, large dynamic range, good geometric accuracy and broad spectral response. CCDs are a mature technology with extensive experience in astronomical imaging, including space missions such as the Galileo probe to Jupiter and the Hubble Space Telescope, and they are readily available in large formats. The most common one now being produced for astronomical work is 2048×4096 pixels, buttable on 3-sides for assembling "mosaics" of multiple devices with minimal dead space between devices. Small pixels (typically ~15 um, though the range goes down to 7 um) lead to an overall reduction in the size of the optical imager and its optics. Small pixels also reduce the already very low dark current per pixel, typically 30 e^-/pixel/.hour at 200 K for 15 um pixels and essentially immeasurable at 30 K. On the other hand, reducing the pixel size also reduces the full well capacity. Current scientific grade CCDs exhibit less than 4 electrons of read noise at slow scan rates (<500 kHz).

Silicon, with a bandgap of 1.12 eV at 300 K, is sensitive to wavelengths from around 300 to 1100 nm, but reflections and absorption in non-active regions reduce the quantum efficiency of actual devices. CCDs are usually optimised with anti-reflection coatings and other process parameters for either the red or blue end of this band and can reach a peak QE of over 90%. However, the optical transparency of silicon in the near infrared is quite high, and the QE usually drops rapidly beyond 800 nm, even for red-enhanced devices, because the absorption depth of the long-wavelength photons exceeds the thickness of the active region of the detector (typically around 20 um for a thinned CCD). For 1000 nm photons, the absorption depth is about 100 um and the QE of a typical CCD is well below 20%. Unfortunately, the transparency in the near-infrared also increases significantly as the temperature is lowered so that at 30 K the absorption depth is about 1100 um (using models from [1]). The only way to detect these long-wavelength photons is to increase the active region of the detector. This approach is taken in recent devices produced from high-resistivity silicon, which have active regions about 300 um thick and have shown much improved quantum efficiency in the near infrared (up to 50% for 1000 nm light at 153 K) [2], [3], [4]. One draw back to this larger active volume is increased dark current (50 e^-/pix/hour at 150 K [3]); however, the dark current will still be insignificant at 30 K.

Despite their excellent performance in ground-based instruments, there are several difficulties with using CCDs as a visible wavelength detector on NGST. These include the susceptibility of CCDs to damage from space radiation, complex support electronics, and poor performance at the low ambient temperature of the NGST baseline (30 K).

A.2.1 Radiation susceptibility

The space environment includes a large amount of high-energy particle radiation, which can cause permanent damage to CCDs (as well as other types of detectors). High-energy particles, especially protons, can create dislocations in the silicon crystal lattice, some of which create mid-gap energy states (traps) in the device. We performed numerous tests on CCDs that had been subjected to proton radiation, similar to what would be expected in space [5]. We found that the radiation-induced traps have two main harmful effects: they can increase the amount of dark current, particularly in high-intensity localised events referred to as “hot pixels”, and they can reduce the charge transfer efficiency (CTE) of a CCD [5]. Both of these effects have been observed in space-based CCD instruments such as the Wide Field / Planetary Cameras (WFPC1 and WFPC2) and the Space Telescope Imaging Spectrograph (STIS), all of which were or are on board the Hubble Space Telescope.

Whereas increased dark current and hot pixels may be experienced with any photovoltaic detector, degradation of CTE is a problem unique to CCDs. The photo-generated charge in a CCD must be transferred through the device to the output, and this process is not perfectly efficient. Trapping states can remove charge from the signal as it is being transferred. The loss of signal can be severe – in the worst case examples from the WFPC2 after 5 years in orbit it is 40% [7] - and is particularly problematic in large format devices in which the signal must be transferred greater distances. The effects of poor CTE are difficult to calibrate out because the losses depend on numerous factors, including temperature, signal strength, background level, image details and even the contents of previous images [6].

There have been successful attempts to reduce the effect of radiation damage on CTE, such as the introduction of a highly doped trench in the charge channels of the CCD, which constricts small signals and

reduces their exposure to traps. Another possibility is to produce the CCD with p-type doping in the channels. The radiation-induced traps in p-type silicon are not as efficient at capturing the signal charge and CCDs of this type have been shown to have little degradation in CTE even after irradiation to 40 times the level expected for NGST [EMS report]. The high-resistivity silicon devices produced by Lawrence Berkeley National Laboratory are coincidentally made with p-type channels and therefore would be expected to be radiation hard. However, because of the thick active layer of these devices, they show an increased sensitivity to cosmic rays [3].

The production of traps by radiation damage may also be suppressed at very low temperatures. Saks and Modolo report an almost complete suppression of damage from ionizing radiation at temperatures below 80K. However, the suppressed damage will appear if the device is allowed to warm up afterward (reverse annealing) [8]. Low temperatures may also reduce the effect of the damage that does occur. The trapping mechanism is highly temperature sensitive because of the varying thermal energy of the carriers. At a low enough temperature, the carriers don't have enough energy to escape the traps and the traps remain full, rendering them inert. In our tests, the dominant trap at a level of around 0.4 eV was almost completely ineffective below 180 K [1]. We were not able to cool the CCDs below 140 K, but we did test some transistors identical to those at the output of a CCD at temperatures down to 40 K. These transistors had also been subjected to radiation and we did see significant radiation-induced generation of a trap that would dominate in the range 50-100K, but this is still above the NGST baseline temperature [5].

A.2.2 Support electronics

The drive circuitry for a CCD is more complex than the simple TTL-level drivers needed for a CMOS multiplexer. High voltages (typically 10-30 V) are needed and care must be taken in the design of the clock drivers to ensure good CTE. High-resistivity CCDs require even higher bias voltages to ensure full depletion. However, at slow readout rates, the power requirements are not high. A rough estimate, for a typical 2048×4096 pixel CCD with 2 output channels read out in approximately 11 seconds (around 370 kpixels/s from each output), is that the clock drivers and bias supplies would dissipate an average of about 140 mW during a readout. For a 12k×12k pixel focal plane array (18 CCDs), that becomes about 2.5 W. However, the average power is considerably less because the duty cycle is low. For a 30 minute exposure, average power consumption is 15 mW. The current power budget on NGST is about 3W per instrument. The power consumption can be reduced by using a hybrid device like those developed at the Jet Propulsion Lab (JPL). These devices consist of a CCD bonded to an integrated CMOS circuit containing both drive and digitization electronics. A prototype 256×512 array was able to readout at 50 kpixels/s with less than 5e- noise and consuming less than 100 mW of power [8].

A CCD would also require a shutter for proper operation, which is not the case with multiplexer-based hybrid detectors. The latter are electronically shuttered as a result of the readout process.

A.2.3 Low temperature operation

The basic difficulty with operating CCDs at very low temperatures appears to be carrier freeze out. That is, the donor or acceptor levels themselves, though extremely shallow, become effective traps because of the low thermal energy of the carriers. We know from our experience with radiation induced trapping levels at higher temperatures that traps reduce the charge transfer efficiency (CTE) through signal charge capture

and deferral [6]. Freeze-out trapping at low temperatures is a particularly serious problem because of the high density of dopant impurities in the channel.

In a typical buried channel CCD, the active n-type region is doped with phosphorus, which has an energy level of 0.044 eV below the conduction band. At room temperature, and even down to liquid nitrogen temperatures, the photo-generated electrons have plenty of thermal energy to escape this donor level quickly. But below 50 K, the emission time starts to become significant, and charge deferral effects start to be seen.

The main method explored so far to alleviate the freeze out problem in CCDs is simply to reduce the concentration of donor levels in the channel [10], [11]. A lower concentration of donors means that fewer signal electrons are trapped, and therefore an improvement in the transfer efficiency is seen. However, a reduced donor concentration in the buried channel also reduces the full-well capacity. A compromise solution evaluated by Lattes and Tsaur [11] is to use a two-step buried channel implantation. The first implant is a low dose implant to reduce the trapping effects. The second implant is a shallow, high dose implant to provide extra charge capacity. Smaller signal packets, which are more severely affected by trapping, are confined to the low dose region of the channel, while larger signal packets can spill over into the higher dose "overflow" region.

The effect of trapping on the transfer efficiency is dependent not only on temperature but also on clocking rates [6], [10]. Kimata *et al*. tried adjusting various aspects of the clocking patterns to examine the effects. In their experiments, simply slowing down the clocks, which can be quite effective with deeper traps at higher temperatures, had only a limited ability to improve the CTE at carrier freeze out. However, they found that increasing the pulse fall time yielded a much greater improvement in CTE. Their explanation was that low temperatures, in addition to reducing the ability of signal charges to escape shallow traps, significantly reduce the rate of thermal diffusion, so that charges emitted from the impurity levels during transfer need the extra push of a drift electric field in order to efficiently rejoin the charge packet in transit.

Figure A-1 shows modelled performance of a CCD array similar to the one on which we did our radiation testing [6]. The solid curves show the charge transfer *inefficiency* (CTI =1-CTE) at the nominal readout speed (100 us clock pulse widths) for two different channel doping concentrations. The curve on the right is the normal doping (peak of 2×10^{16} cm^{-3}) and the left curve is for a reduced doping concentration (2×10^{15} cm^{-3}). The dotted curves show the results for the same two doping concentrations, but at a much slower readout speed (1 ms clock pulse widths). It appears from the figure that the high concentration of donor impurities causes a sharp increase in the CTI in the temperature range 30 – 60 K, although the accuracy of the simple model in this range has not been verified. In particular, the model does not account for the reduction in the rate of thermal diffusion mentioned above. However, the model is in agreement, at least qualitatively, with the experimental results in [10].

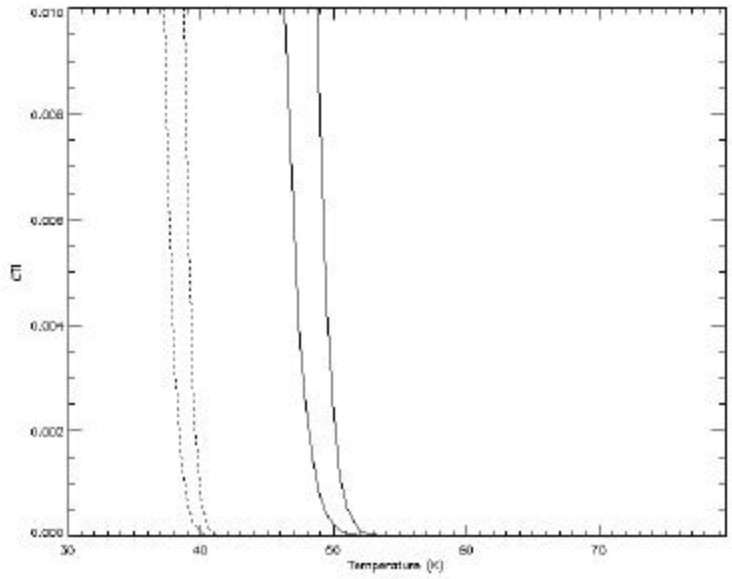


Figure A-1: CTI vs. temperature for different clocking scenarios and doping levels

Banghart *et al*. describe the effect of the electric field on trapping at low temperatures [12]. The electric field distorts the potential well of a trap in what is referred to as the Poole-Frenkel effect. The distortion lowers the barrier potential of the well in one direction, making the trap effectively shallower. However, their experiments show only a limited improvement in the CTE when the electric field in the channel was increased by increasing the voltage swing of the CCD clock lines.

There are two other possibly deleterious effects of freeze out. The freeze-out effect may increase the resistance of the doped polysilicon clock lines, thereby increasing the demands on the clock driver circuitry in order to maintain good CTE. Also, in our tests of radiation damaged transistors (as are found at the output of a CCD) we found an increase in the noise due to radiation induced trapping [13]. There may be a similar increase in the noise at very low temperatures due to carrier freeze out.

A.3 Hybrid detectors

Hybrid detectors consist of a slab of photoresponsive material containing an array of photodiodes bonded via indium "bumps" to a multiplexer array. The incoming photons create electrical carriers in the photodiodes that are collected and stored on the capacitance of the diodes and the input nodes of the multiplexer. Transistor switches connect each of the multiplexer input nodes to one of the outputs enabling the control electronics to read out the array one pixel at a time.

Hybrid detectors have been in development for many years (partly due to the great interest in infrared imagers for military applications), but have not yet reached the same performance level as CCDs. The technology is about ten years behind CCDs, and it should be noted that detectors in space are about 10 years behind ground based ones. The largest array sizes currently being manufactured are 2048 × 2048 pixels, and the smallest pixel size is 18 um. They exhibit much higher noise than CCDs, typically around 10-30 electrons rms, but this can be reduced by multiple sampling to less than 10 electrons. A common sampling technique, referred to as Fowler sampling, involves taking an equal number of samples of the reset

level and signal level. One Fowler sampling is equivalent to correlated double sampling (CDS) which is also used for destructive readout devices like CCDs. Other problems found with hybrid detectors are non-linearity, cross talk between outputs, and glow of the shift registers or output amplifier at wavelengths detectable by the photosensor.

Hybrid detectors are attractive, however, for several reasons. They could share a common readout architecture with the infrared science detectors, reducing development costs. They are designed to operate at low temperatures to reduce the background in infrared imaging, so they will have no difficulty with the ambient NGST temperature. Because there is no charge transfer involved in the readout of these devices, they are more resistant to radiation damage than CCDs, and the random access allows faster readout of small regions of interest. Hybrid detectors do not exhibit "blooming" (charge bleeding into neighbouring pixels when the capacity of a single pixel is exceeded). They are also low-power devices. And finally, hybrid detectors offer flexibility in wavelength response by selecting an appropriate photosensitive material. Three types of photosensor material are described in the next sections.

A.3.1 Indium Antimonide (InSb)

Indium Antimonide is a type III-V semiconductor with a bandgap of 0.2 eV and has a fundamental wavelength cutoff at around 5 um. Its performance at wavelengths less than 1 um has not been well characterized, and there is concern that the higher energy photons will produce multiple electron-hole pairs and that this process will increase the noise of the device at short wavelengths. Work is in progress on this, and preliminary results indicate good quantum efficiency down to 450 nm and seemingly no problems with excess noise [14].

One difficulty with InSb detectors is the crosstalk between pixels due to carrier diffusion in the material. Current detectors have 27 um pixels. At this size, modelling indicates that the crosstalk is about 5% into the neighbouring pixels and it increases to about 10% at 18 μm [15]. However, this can be reduced by thinning the InSb layer. Measurements of the devices used on SIRTF that have 30 um pixel spacing and have been thinned so that the InSb slab is only 7 μm thick indicate a pixel-to-pixel crosstalk of 0.3% at an operating temperature of 15 K [16].

Some poorly understood problems with InSb detectors include persistence (bright sources producing ghosts in subsequent images) and "black holes", which are areas of seemingly depressed responsivity which appear randomly in flat field images.

The most advanced InSb arrays currently available are the ALADDIN II arrays produced by the Raytheon Infrared Center of Excellence (IRCoE). These 1024 × 1024 pixel devices exhibit 35 e^--read noise for dual correlated sampling (DCS), but this can be reduced through multiple sampling to less than 10 e^- (32 Fowler sampling). Dark current is about 0.05 e^-/pix/s at 35 K. Quantum efficiency is about 90% in the 1 – 4.5 um region and drops to 65% at 5 μm [16]. The recent test results indicate that with an appropriate anti-reflection coating, a QE of over 90% down to 450 nm is achievable [14].

The Raytheon IRCoE is currently in the second year of a three-year program to develop improved readout circuits in the same format as the ALADDIN arrays. The aim is to reduce the noise, amplifier and shift register glow, and power consumption [17]. A parallel development program is aimed at producing larger format (2048 × 2048) detectors.

A.3.2 Mercury Cadmium Telluride (HgCdTe)

Mercury Cadmium Telluride is a semiconductor alloy whose bandgap can be varied by varying the composition of the alloy. The most common alloy in use in astronomical detectors, such as the NICMOS detectors on HST, has a wavelength cutoff of 2.5 um. The advantage of the shorter wavelength cutoff is a lower dark current which allows operation of the device at a higher temperature. However, even for material with a similar wavelength cutoff, HgCdTe has a much lower dark current than InSb [18].

The best HgCdTe arrays currently available are the HAWAII arrays made by the Rockwell Science Center. These are 1024 × 1024 pixel arrays with 18.5μm pixels. Mated with PACE 2.5 um detector material, these devices exhibit less than 9 e^- read noise with CDS and down to 3 e- with Fowler sampling [18]. Dark current has been measured to be around 3 e^-/hour at 43 K [20]. Quantum efficiency increases from 50% at 0.8 μm to 60% at 2.4 μm and drops off sharply at 2.5 um.

These detectors suffer from many of the same problems as the InSb arrays, such as crosstalk due to diffusion, and persistence. ESO tests on a Rockwell HAWAII array indicate up to 15% crosstalk between adjacent pixels, although it is unclear how much of this is due to the finite point spread function of the stimulating light [19]. As with InSb, diffusion crosstalk should be reducible by thinning. The ESO tests also revealed multiplexer problems of shift register glow, non-linearity and electrical crosstalk between rows.

Rockwell is currently developing a second generation array (HAWAII-2) which has 2048 × 2048, 18 um pixels. As of April, 1999, the first multiplexers had been operated and mating with detector material (initially PACE 2.5 μm) was expected within a month [21]. Performance is expected to be similar to the HAWAII arrays, with a reduction in both shift register glow and electrical crosstalk [22].

The HAWAII-2 arrays will later be mated with detector arrays produced with Rockwell's advanced double layer planar heterostructure (DLPH) process to produce high-performance 5 μm cutoff hybrids. The DLPH process uses molecular beam epitaxy (MBE) to grow a double layer of wide and narrow bandgap HgCdTe on CdZnTe substrates [23]. The result is superior surface passivation and reduced defect density which means less persistence and lower dark current. This technology approaches the theoretical limit for dark current at a given cutoff wavelength [18], [23]. Unfortunately, the CdZnTe substrate, unlike the sapphire substrate used for PACE, is opaque at visible wavelengths and must be removed to operate below 0.85 μm. Rockwell claims a quantum efficiency of 60% at 400 nm for detectors with the substrate removed [21]. This substrate material is, however, difficult to produce. The largest array size available so far in the DLPH technology is 640 × 480 pixels. Other substrate options such as CdTe/Si have been explored, but currently these have a large dislocation content [18].

A.3.3 Silicon p-i-n

A third option for the detector material of a hybrid is regular silicon. As noted above for CCDs, the fundamental wavelength coverage of a silicon detector is 0.3 to 1.1 um. In fact, a bare multiplexer is sensitive to light and can be used for imaging, a concept which is exploited by active pixel sensors (APS) which have recently been challenging the CCD's favoured status as an optical imager. APSs are attractive because they can take advantage of the highly developed standard CMOS manufacturing processes, they use less power, and they can be integrated on a single die with drive circuitry and even an analog to digital converter. The main advantage of using a hybrid with a separate slab of detector material is increased sensitivity. Because of the circuitry in each pixel, the fill factor of a bare multiplexer or APS is about 50%,

whereas a hybrid has a fill factor of close to 100%. The detector material can also be optimised to get better QE, especially at the longer wavelengths.

The p-i-n photodiode configuration consists of a typically 150-300 um thick substrate of intrinsic (i) silicon with a p-type implant on one surface and an n-type implant on the other. This creates an active region that extends across the whole thickness of the device, increasing the chances of absorbing long wavelength photons in a similar way to the high-resistivity CCDs. Also similar to these CCDs, the p-i-n photodiode will have a relatively high dark current (though insignificant at 30 K), and will be strongly affected by cosmic rays. Because the depletion region, and therefore the electric field, also covers the whole thickness of the device, diffusion crosstalk should be virtually eliminated.

Currently, Rockwell is advertising the "HyViSi", a silicon p-i-n hybrid detector based on their TCM6600 multiplexer, which has 640×480 27 μm pixels [24]. The dark current is ~14000 e^-/pixel/s at 273 K. With an antireflection coating for 680 nm and operated at 257 K, the QE is 40% at 400 nm, increasing to 90% for the range 600 to 900 nm and then dropping sharply to 50% at 1 um. An even broader QE response for a single silicon photodiode is reported in [25]: 70% at 400nm, 90% at 480 nm and decreasing to 70% at 1 um.

There seems to be no inherent obstacles to creating larger silicon arrays and mating them to the HAWAII-2 multiplexers to create 2048×2048 pixel hybrids, and Rockwell intends to pursue this option soon.

A.4 Summary

Table A-1 give a general summary of some key characteristics of the visible detectors discussed. This is a composite of currently or imminently available technology.

Table A-1: Key characteristic comparison

	Requirement	CCD	InSb	HgCdTe (2.5 um)	Si PIN
Array size (pixels)	12000×12000	2048×4096	1024×1024	2048×2048	2048×2048
Pixel size (um)	< 20	15	27	18	18
Read noise with Fowler sampling (e-)	< 6	2	<10	3	3
Dark current (e-/px/hour @ 30K)		~0	<180	<3	~0
Wavelength range (m)	0.5 – 1.0	.3 – 1.1	0.4 - 5	0.3 – 2.5	.3 – 1.1
Quantum efficiency (500nm)	80%	85%	90%	60%	70%
Quantum efficiency (750nm)	80%	75%	90%	50%	90%
Quantum efficiency (1000nm)	80%	50%	90%	50%	50%

Figure A-2 shows quantum efficiency curves for some representative devices. The short-wavelength InSb data, from tests conducted at the University of Rochester on 256 × 256 devices from the SIRTF program, show the effects of an infrared AR coating. If the coating is taken into account, the response is much flatter.

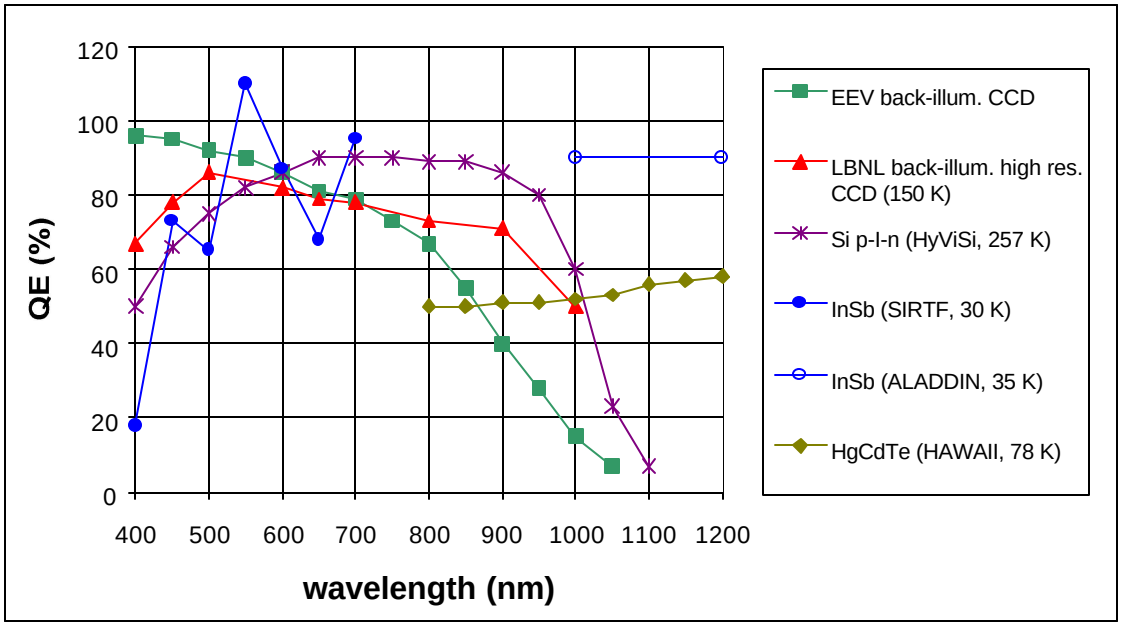


Figure A-2: Quantum efficiency curves

The main advantages and disadvantages of the various technologies are:

- CCDs pros: mature technology
  low noise
  large array sizes available
  small pixels
  low dark current

  cons: higher power consumption
  low temperature operation not demonstrated
  radiation susceptibility (CTE)
  shutter/special readout electronics required

- InSb pros: broad spectral response

  cons: "black holes"
  small array size
  large pixels
  high dark current
  high noise
  persistence
  multiplexer glow

- HgCdTe pros: variety of spectral ranges available

cons: persistence
multiplexer glow
lower quantum efficiency

- Si p-i-n pros: low dark current

cons: cosmic ray effects

References

[1] D. E. Groom, S. E. Holland, M. E. Levi, N. P. Palaio, S. Perlmutter, R. J. Stover, and M. Wei, "Quantum Efficiency of a Back-illuminated CCD Imager: An Optical Approach," SPIE Vol.3649, pp. 80-90 (1999).

[2] R.J. Stover *et al.*, "A high performance CCD on high resistivity silicon," SPIE Vol. 3505, pp. 13-18 (1998).

[3] *ibid*, "Characterization of a fully depleted CCD on high resistivity silicon," *Solid State Sensors Arrays: Development and Applications*, SPIE Vol. 3019, pp. 183-188 (1997).

[4] P.S. Heyes, P.J. Pool, and R. Holtom, "A new generation of scientific CCD sensors," *Solid State Sensors Arrays: Development and Applications*, SPIE Vol. 3019, pp. 201-209 (1997)

[5] T. Hardy, M.J. Deen and R. Murowinski, "Effects of radiation damage on scientific charge coupled devices," in *Advances in Imaging and Electron Physics*, Vol. 106, Academic Press (1999).

[6] T. Hardy, R. Murowinski and M.J. Deen, "Charge transfer inefficiency in proton damaged CCDs," *IEEE Transactions on Nuclear Science*, Vol. 45, No. 2, pp. 154-163 (1998).

[7] B. Whitmore, I. Heyer, S. Casertano, "Charge transfer efficiency of the Wide Field and Planetary Camera 2," Space Telescope Science Institute web site (http://www.stsci.edu/instruments/wfpc2/wfpc2_advisory.html), (April 1999).

[8] N.S. Saks and J.M. Modolo, "Radiation effects in n-channel MNOS CCDs at low-temperature," *IEEE Transactions on Nuclear Science*, Vol. 30, No. 6, pp. 4197-4204 (1983).

[9] G. Domingo, "Hybrid Imaging Technology," NGST Detector Workshop, Space Telescope Science Institute, Baltimore, MD, April 20-21, 1999.

[10] M. Kimata, M. Denda, N. Yutani, N. Tsubouchi and S. Uematsu, "Low-temperature characteristics of buried-channel charge-coupled devices," *Japanese Journal of Applied Physics*, Vol. 22, No. 6, pp. 975-980 (1983).

[11] A.L. Lattes and B.-Y. Tsaur, "Buried –channel CCDs with high charge transfer efficiency and large charge capacity for low-temperature readout of long-wavelength infrared detectors," *Infrared Readout Electronics*, SPIE Vol. 1684, pp. 212-219 (1992).

[12] E.K. Banghart, J.P. Lavine, E.A. Trabka, E.T. Nelson, B.C. Burkey, "A model for charge transfer in buried-channel charge-coupled devices at low temperature," *IEEE Transactions on Electron Devices*, Vol. 38, No. 5, pp. 1162-1174 (1991).

[13] T. Hardy, R. Murowinski and M.J. Deen, "Low-frequency noise in proton damaged LDD MOSFETs," *IEEE Transactions on Electron Devices*, Vol. 46, pp. 1339-1346 (1999).

[14] W. Forrest, private communication.

[15] K.M. Merrill, "NOAO experience with the ALADDIN (InSb) detectors," NGST Detector Workshop, Space Telescope Science Institute, Baltimore, MD, April 20-21, 1999.

[16] A.W. Hoffman *et al*., "Near IR arrays for ground-based and space-based astronomy," *Infrared Astronomical Instrumentation*, SPIE Vol. 3354, pp. 24-29 (1998).

[17] J.D. Garnett, "ORIGINS – Pathway to NGST," NGST Detector Workshop, Space Telescope Science Institute, Baltimore, MD, April 20-21, 1999.

[18] L.J. Kozlowski *et al*., "HgCdTe 2048^2 FPA for infrared astronomy: development status," *Infrared Astronomical Instrumentation*, SPIE Vol. 3354, pp. 66-76 (1998).

[19] G. Finger *et al*., "Infrared detector development programs for the VLT instruments at the European Southern Observatory," *Infrared Astronomical Instrumentation*, SPIE Vol. 3354, pp. 87-98 (1998)

[20] G. Finger, "Experience with near IR detector Arrays used for low background ground-based applications," NGST Detector Workshop, Space Telescope Science Institute, Baltimore, MD, April 20-21, 1999.

[21] K. Vural, "HgCdTe detector development at Rockwell," NGST Detector Workshop, Space Telescope Science Institute, Baltimore, MD, April 20-21, 1999.

[22] Preliminary data sheet for HAWAII-2 arrays, Rockwell Science Center (1999).

[23] R. Bailey *et al*., "Prospects for large format IR astronomy FPAs using MBE-grown HgCdTe detectors with cutoff wavelength > 4 μm," *Infrared Astronomical Instrumentation*, SPIE Vol. 3354, pp. 77-86 (1998)

[24] HyViSI data sheet, Rockwell Science Center (1999).

[25] S.E Holland, N.W. Wang and W.W. Moses, "Development of low noise, back-side illuminated silicon photodiode arrays," *IEEE Transactions on Nuclear Science*, Vol. 44, No. 3, pp. 443-447 (1997).

B NGST PSF Degradation with Mirror Surface Roughness

It is likely that efforts will be made to fabricate the NGST mirrors (primary and secondary) such that any dust on them will not dominate the intrinsic mirror microroughnesses. In order to estimate what effects the scatter will have on the PSF, models were computed using the program, ASAP, from Breault Research Corporation. At this stage we consider only the effects of specular scatter from shiny surfaces. Diffuse

scatter, although simpler to calculate, is much more difficult to estimate at the detector because the geometry and surface properties of structures in the vicinity of the incident and internal light beam need to be defined beforehand. In this study we will, therefore, only examine the specular scattering so that an empirical relation can yield scattering estimates based on cleanliness levels and surface microroughness.

Before discussing the relative scatter from both sources for a wavelength of about 1.6 microns and the visible it is important to understand how the estimates of scatter are made and what definitions are used. Much of the material derives from empirical studies and because of the large number of parameters and difficult-to-achieve conditions, the errors are large; that is, the precision of measurement of scattering estimates from various measuring instruments at various laboratories is at least as much as an order of magnitude.

B.1 Estimates of scatter

The Strehl ratio for a shiny mirror where the surface irregularities are small wrt the wavelength is given by

$$S = \exp\left[-(2\pi s / \lambda)^2\right]$$

The phase aberration $2\pi s / \lambda$ contains σ, the wave aberration. Note that σ implies an rms wavefront difference, $s\sqrt{2}$, or an rms surface difference, $s/\sqrt{2}$. The rms wavefront deviation is 2X the surface rms deviation.

According to Mahajan (1983, Applied Optics 22, p 3035) the aberrated PSF, within the Airy disk, is practically identical to the unaberrated PSF, except that it is scaled by the Strehl ratio (*S)* as given above. Bennett and Porteus (1961, J. Opt. Soc. Am. 51 p123) show that the specular reflectivity of a "rough" surface mirror is also given by this same definition of Strehl ratio.

Of course, the removal of scattered specular power (watts) from the PSF implies that this is distributed into the PSF wings. However, calculations by Mahajan where the rms wave deviations are introduced into the pupil, show that the form of the PSF inside the Airy disk is maintained; only the level is decreased within this radius and the estimate is considered appropriate for Strehl ratios greater than 0.1. For the specular scattering from the NGST mirrors, the Strehl ratios caused by scattering alone are expected to be greater than 0.99. The scattered light is then a very small fraction of the total and this is distributed over a very much larger area in the wings.

The TIS or total integrated scatter, is related to the rms roughness (δ) of a surface by

$$TIS = 1 - \exp\left[-(4\pi d / \lambda)^2\right] \quad \text{or} \quad TIS \approx (4\pi d / \lambda)^2$$

It is the ratio of light scattered into a hemisphere to the total incident specular light.

The most detailed specification of scattering is given by the bidirectional scattering distribution function

(BSDF) and is the scattered surface radiance/incident surface irradiance where radiance is expressed in watts/sq. m./steradian and irradiance is expressed in watts/sq. m. – so BSDF has units sr^{-1}. More specifically, the bidirectional reflection scattering function is referred to as BRDF.

The BSDF depends on two directions of incidence and scatter, and its dynamic range is very large. BSDFs are approximated by two numbers, an intercept and a slope. The intercept is the BSDF at 0.57 degree from the specular direction and the slope defines the change in scattered power as a function of the angle from the specular direction. Both parameters define a best-fit straight line (Harvey-Shack model) in a log-log plot of BSDF as a function of scatter angle.

Incident surface irradiance is the light flux (watts) on the surface per unit of illuminated surface area (not beam cross-sectional area). The scattered surface radiance is the light flux scattered per unit surface area per unit projected solid angle (solid angle x cos(θ)).

$$BSDF = (dP_{out}/dw)/(A\cos(q))\,//\,P_{in} \quad \text{so} \quad BSDF = 1/P_{in} * dP_{out}/d\Omega \quad \text{where } \Omega = dw\cos(q)$$

$TIS = \int_{\Omega} BSDF$ which yields $BSDF(b) = b(|\,b - b_0\,|/b_p)^m$. Here, b is the intercept and m is the slope of the BSDF designation. b is the BSDF at b_p, b is the sin of the scatter angle, and b_p = *sin(0.01)*. b_0 is *sin(specular angle – typically 0)*.

Using the formalism above, the rms scatter (*d)* can be related to the BSDF parameters b and m by the following equations…

$d(b,m,l) = l/4p * \sqrt{2pbb_p^{-m}k(m)}$ where $k(m) = \left[(b_2)^{2+m} - (b_1)^{2+m}\right]/(2+m)$ integrating between b_1 and b_2 over the whole solid angle.

For b=0.05 and m=-1.5 which represents the scattering parameters for a good mirror in the visible (550 nm), like the Hubble mirror after it was aluminised, d = 1.04 nm rms. For b=0.5, m=-1.2, which were the measurements for the Hubble mirror after several days in a class 100 clean room, δ = 5.4 nm rms. For comparison, a car windshield has a microroughness of about 20 nm rms.

For some surfaces and conditions the BSDF parameter, b, can be scaled by

$$b(l_2) = b(l_1)(l_1/l_2)^{4+m},$$

which for the visible to the infrared (550 to 1600 nm, say), is a factor of about 0.05 for m=-1.2.

If the rms microroughness is 2.5 nm as estimated for the NGST mirrors by Bely, Petro, and Mehalick (1999), the Strehl reduction of the PSF is about 0.9967.

The surface microroughness in terms of an rms deviation yields an even scattering over all angles; there is no specification of how the angle dependence changes the amount of the scattering. Although this is a useful description of matte surfaces such as dull painted or etched surfaces it is not the best description for shiny surfaces such as polished glass or metal. In those cases there is substantial scattering difference depending on what the incidence and reflection angles are. For this reason we choose the BRDF-for-shiny-surfaces model where the specular reflection is many orders of magnitude greater than the off-specular scattering caused by the microroughness.

If the scattering level is considered to be the same over the mirror surfaces it is easy to compute the fraction scattered by means of the BSDF definition. For example, the scattered light incident the secondary mirror from the primary mirror is just the power on the primary times the BRDF of the primary times the solid angle of the primary subtended by the secondary. The calculation can proceed for each surface until the detector is reached. Overall, uniform scatter at the detector would include the scattered light originating at each surface, each component calculated in the same way and added.

In what follows we will take a look at what might be expected from scattering caused by particulates and then compare it to the distribution of scattered light from the intrinsic microroughness of the mirror surfaces.

B.2 Particulate scatter

A method for determining estimates of scattering from dust on smooth mirrors is outlined in Spyak and Wolfe (1992) and they have provided a number of graphs of BRDF levels for various surface cleanliness levels at 4 wavelengths. They assume that the particles of dust scatter radiation according to the Mie theory and do so independently. Also, that the radiation scattered in the forward direction reflects from the mirror and does not interact again with the particle.

Interpolation of their graphs yield (to within about ¼ order of magnitude) the following BRDFs as represented by the Harvey-Shack model (parameters b and m)…

Table B-1.

cleanliness level 500	b= 5.0	m = -2.84
cleanliness level 200	b= 0.025	m = -2.67
cleanliness level 100	b= 0.0014	m = -2.31
cleanliness level 50	b= 0.00011	m = -2.18

The distribution and number of particles on a mirror determine the amount of the resultant scatter. Spyak and Wolfe have used a number of sample distributions for their contamination specifications. Specifically, for two distributions, the scattering tends to similar values at angles greater than about 10 degrees from specular for the wavelength region between 1.15 and 3.39 microns. It is not difficult to fit the empirical relationship for the BRDF (see above BRDFs for particular cleanliness levels) through this region. This, then, gives a reasonable estimate of the microroughness that would be a sensible specification for the NGST

mirrors. The assumption being that the particulate scatter does not have the same degree of wavelength dependence as does that for microroughness.

B.2.1 Intrinsic mirror scatter

In addition, there is the microroughness of the mirrors themselves. Following the discussion given by Spyak and Wolfe (1992), an accurate empirical relationship for the BRDF of a clean mirror is the following:

$$BRDF = 2k^4 \sigma^2 l^2 / \left[\pi (1 + (kl(\beta - \beta_0))^2)\right]$$

Here $k = 2\pi / \lambda$, σ is the rms roughness, l is the correlation length(from the autocorrelation function = 1/e the correlation at zero shift), β is the sin of the scattered light from specular, and β_0 is the sin of the incidence angle. According to Elson, Rahn, and Bennett (1983, Applied Optics 22, No. 20, p 3207.), short correlation lengths (*l=0.2m)* refer to surfaces such as conventional polished glass and long correlation lengths (*l=10m)* refer to chemically polished or diamond turned surfaces.

Although the above empirical relationship differs from the Harvey-Shack model used in the ASAP software, it is easy enough to extract a slope (m) and a 0.57-degree-from-specular scatter estimate (b) from the plotted curves given by Spyak and Wolfe (1992). Knowing these quantities allows ASAP to compute the resultant overall scattering contribution in addition to the PSF calculation. ASAP uses the method of gaussian beams for the computations and, therefore, accurate estimates of interference effects are also included for non-scattered rays. The ASAP computer program does suffer from a normalising effect in that more rays (beams) generate more scatter but the effect appears to be easily within an order of magnitude.

B.3 Overall scattering contribution

Typical BRDF parameters for a qualitative range of mirror cleanlinesses when the particulates are relatively small and form more of a haze are given by the following table. In the Spyak Wolfe papers the cleanlinesses refer to various distributions of particulates on perfect mirrors. Their BRDF slopes (m) are closer to the slopes for very clean mirrors but their BRDF intercepts (b) are higher. This would demonstrate that the presence of larger particulates causes greater scattering. In fact they show that very few relatively large particles are required to dominate the scatter profile of a smooth mirror in the far IR. Also, they show that particulate scatter does not have as strong a wavelength dependence as microroughness scatter. That means that particulates contribute more to the overall scatter at longer wavelengths.

Table B-2.

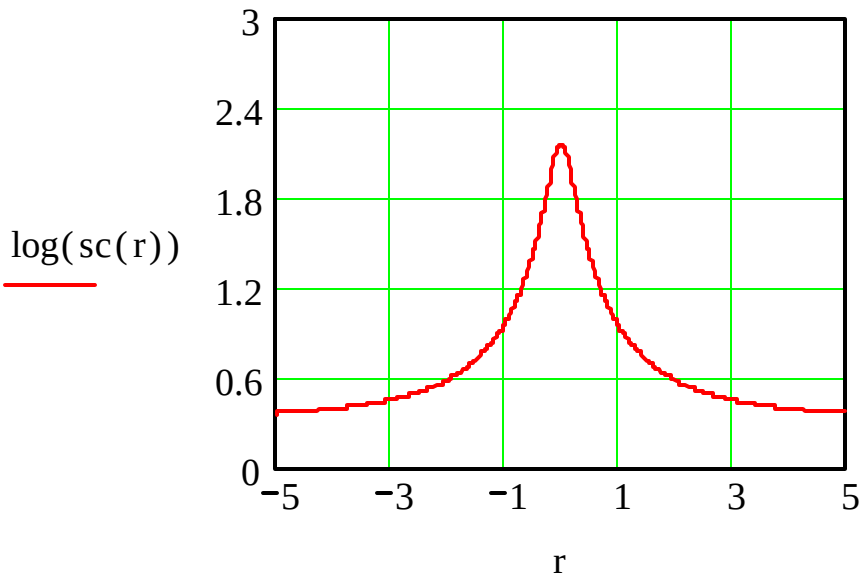
NGST Mirror Description	Approximate BRDF (b, m)
Visible average clean	0.05 -1.5
Visible extra clean	0.02 -2.5
Visible dirty	0.15 -1.0

IR average clean	0.005 -1.5
IR extra clean	0.002 -2.5
IR dirty	0.015 -1.0

Asap computes the resultant scatter at an image by applying the Harvey-Shack law to the various optical surfaces. For the present study only the primary and secondary mirrors were assumed to be contributors to overall scatter. To this end models of the PSF were computed using an 8 metre circular pupil. An 8-petal configuration requires an unwieldy number of beams for proper definition and a corresponding long computation time. Little if anything is gained since using the circular pattern over the petal pattern merely adds a small fraction of scattering.

Models using a large range of b (.0001 –> 10) and m (-1.1 -> -3.0) values showed that the resultant scatter can be reduced to a two parameter empirical relationship for the 2 mirror study. The fitting of the actual curves is certainly not exact but in no cases were the deviations more than an order of magnitude different between the ASAP model and the empirical relationship.

One has only to use Table B-1. of BSDFs and the estimates of intrinsic microroughness to obtain particular values for b and for m. Given these, the resultant scatter can be found as a function of radial distance from the centre of the PSF. Assuming that logarithm of the central brightness of the PSF is 10.5 (the number of rays used in the ASAP calculations resulted in this figure) the empirical relationship is the following:



$$sc(r) = a^3 / \left[a^2 + (a^{1.3} r)^2\right] - m \quad \text{where} \quad a = b(10)^{-2.72m-2.31} / 0.95$$

For b=0.1 and m=-2.0 the LOG of the scattering is shown below as a function of radial distance from the peak of the PSF.

For this situation, then, the central scattering brightness is down from the unscattered PSF brightness by 8.5 orders of magnitude.

References

Bély, Pierre Y., Petro, Larry, Mehalick, Kimberly I., Perrygo, & Charles M., 1999, Straylight analysis of the Next Generation Space Telescope, Internet preprint.
Bennett, H. E. & Porteus, J. O. 1961, J. Opt. Soc. Am. 51, 123.
Bennett, Jean M. & Mattson, Lars 1989, Introduction to Surface Roughness and Scattering, Optical Society of America, Washington D.C.
Elson, J. M., Rahn, J. P., & Bennett, J. M. 1983, Appl. Opt. 22, 3207.
Elson, J. M. & Bennett, J. M. 1979, J. Opt. Soc. Am. 69, 31.
Glenn, P., Reid, P., Slomba, A., & Van Speybroeck, Leon P. 1988, Appl. Opt. 27, 1539.
Hill, J. M. 1994, Steward Observatory Technical Memo UA-94-01.
Lyon, R. G., Dorband, J. E., & Hollis, J. M. 1997, Appl. Opt. 36, 1752.
Mahajan, Virendra N. 1982, J. Opt. Soc. Am. 72, 1258
Mahajan, Virendra N. 1983, J. Opt. Soc. Am. 73, 860.
Mahajan, Virendra N. 1983, J. Opt. Soc. Am. 22, 3035.
Rasigni, G., Varnier, F., Rasigni, M., & Palmari, J. P. 1983, J. Opt. Soc. Am. 73, 222.
Schroeder, Daniel J. 1987, Astronomical Optics, Academic Press, San Diego.
Spyak, Paul R., and Wolfe, William L. 1992, Opt. Eng. 31, 1746-1784.
Stover, John C. 1990, Optical Scattering: Measurement and Analysis, Optical and Electro-Optical Engineering Series, McGraw-Hill.

C A Data Simulation Study of the NGST Visible Imager

C.1 Summary

The NGST will be optimized for the near-infrared (NIR). However, it will still be a powerful tool for wavelengths as short as 0.5 microns. How best to exploit the NGST imaging capabilities here - whether to use the core NIR instrumentation outside its optimal wavelength range or to include a dedicated visible imager in the instrument suite - is an important question. Numerical simulations are discussed which simulate observations with the NGST utilizing a visible imager (VI) and a NIR camera. The goal is to provide a means of direct comparison of the two camera options including different filters, detectors and choices of pixel-scales by means of carrying out virtual observing programs. The user can select one of several telescope designs for which a PSF is generated. The simulation then uses archival data from the HST NICMOS Deep Field South and HST WFPC2 globular cluster fields to generate artificial NGST fields. The galaxy SEDs are generated using template spectra which are redshifted according to a photometric redshift catalog. The stellar colours are based on extrapolations of published HST colour-magnitude diagrams. A “cloning” procedure simulates the increased depth achievable with NGST. The simulation includes a model for the background illumination consisting of a scattered and thermal component of the

zodiacal light. The mirror coatings and filters are simulated with specified efficiency curves. The detectors are also simulated and include pixel scale, quantum efficiency, well-depth, gain, dark current, and readout. The VI and NIR camera permit imaging of the artificial fields through standard HST filters. Results suggest that the VI always achieves better results than the NIR camera in the wavelength range 0.5 - 1.0 um. This is due to a combination of better pixel sampling (0.01 arcsec/pixel at 0.6 microns compared to 0.03 arcsec/pixel for the NIR camera) and detectors optimized for visible light (Si-PIN or HgCdTe versus InSb). The amount of scattered light at short wavelengths due to the micro-roughness of the primary mirror is not as important as having a small pixel scale for the VI. The larger pixels of the NIR camera result in lost information. In simulated observations of galaxy fields, for example, this corresponds to easily-detected star-formation regions which are not detected in NIR camera observations. Simulated observations of extra-galactic globular clusters show that point-source detection and galactic background subtraction are much improved for the VI. The VI would detect the white-dwarf sequence in M15 if it were placed a distance modulus of 8 further away. The NIR camera would not be capable of this.

C.2 Introduction

The large primary mirror alone (8 m diameter) suggests powerful imaging capabilities for NGST not only for near and mid-infrared wavelengths but also for wavelengths shorter than 1.0um. The NGST will have a pupil area roughly 10 times that of HST, and thus, a diffraction limit roughly 10 times smaller than HST. For visible wavelengths it corresponds to a diffraction limit of approximately 0.01 arcsec compared to 0.1 arcsec for HST. It is easy to see that there will be a considerable advantage for NGST in collection area and spatial resolution. Also, this advantage will be compounded by the improved performance of more advanced detectors than were available for HST. What effect these improvements have on real images and how to best take those observations - for example, what type of detector or pixel sampling to use – is not so clear. At the suggestion of Paul Hickson (UBC) a simulation was developed to study the relative merit of a dedicated VI for NGST over using the NIR camera outside its optimal wavelength range. This was a natural extension of earlier numerical simulation work carried out by the author for the Canadian NGST MOS study. There the goal was to determine the advantages and disadvantages of various slit-mask designs for the MOS but included in this was a simulation of broadband imaging with a simulated NIR camera. It appeared that the best means of comparing the VI and NIR camera (which in this simulation would be the MOS imaging mode) would be a direct comparison of imaging of the same field with the two "virtual" instruments. A means of generating artificial NGST deep galaxy fields had already been developed for the MOS simulations but improvements would be needed due to the better pixel sampling of the VI. Also, the NGST PSF would play a crucial role in the VI studies. Discussions with John Hutchings and Peter Stetson (DAO/HIA/NRC) suggested that optimizing the pixel scale of the VI at a wavelength near 0.6 um would make one of it's advantages the detection of point sources. A simulation that quantified this advantage would be needed. It was decided that a simulation of a globular cluster field - one that could be superimposed on a galaxy background to simulate observations of extragalactic globular clusters - was needed.

C.3 A Virtual NGST

The instrumental PSF used in the earlier MOS simulations was a Gaussian with a FWHM given by the diffraction limit of the telescope. This was improved for the VI study by forming a high-resolution mask of the NGST pupil (for the three proposed observatory designs) and taking the Fourier transform. This represents the diffraction pattern at the focal plane. Work done by Chris Morbey (1999) (HIA/NRC) suggests that the Strehl reduction for wavelengths shortward of 1.0 micron for NGST are reasonably well described by the scattering law 1-exp[-(4 /)^2] where is the wavelength and is the RMS micro-roughness of the primary mirror. This law is applied to the diffraction pattern for the NGST to achieve the telescope PSF at the focal plane. The telescope throughput – including the number of gold surfaces - is also calculated.

C.3.1 Generating Artificial NGST Deep Fields

The methods used to simulate artificial NGST fields are now discussed. The galaxy simulation is very similar to that of Steinbring (1999A).

C.3.1.1 Galaxy Fields

To realistically simulate deep imaging for NGST one needs a realistic NGST deep field. One might begin from `first principles' and generate an entirely artificial galaxy field given an expected luminosity function, size-magnitude relation, etc. Perhaps it could even include galaxy evolution. But, this is dissatisfying because there will be no way to know what galaxy evolution NGST will see until NGST can observe it. An alternative is to take the deepest known observations - those of the HST deep fields and extrapolate them. This idea is similar to the method of Hubble Deep Field "cloning" by Bouwens et al. (1997). However, the goal of their method is mainly to alter the redshift distribution of the original HST WFPC2 deep fields without increasing the depth of the exposures. The goal discussed here is to add many new fainter galaxies to the deep fields - in effect, greatly increasing the depth of exposures. The simulation uses images from the HST NICMOS Deep Field South (HDFS) and photometric redshift survey data on the galaxies in this field to generate artificial NGST fields. It utilizes a model for the background illumination consisting of a scattered and thermal component of the zodiacal light. It has template spectra of four different galaxy types, E/S0, Sbc, Scd, and Irr which are redshifted according to a photometric redshift catalog. The galaxies are generated by spreading the *H*-band HST NICMOS image over the range of 1 to 5 microns according to the resulting SEDs.

The simultion generates the artificial NGST field by taking the faint galaxies from the HST field (say, $H_{AB} > 25$), shifting them fainter by a few magnitudes and to a factor higher in redshift, making them smaller, and then putting them back into the field in randomized locations and rotations.

Since the galaxies in the original HST field are undersampled and noisy they are replaced in the simulation by artificial galaxies. They retain the spectral classification, brightness, redshift, and size of the original galaxies but these properties are used to generate highly-oversampled artificial galaxy images. These morphological templates correspond to the four galaxy spectral types and are based on a two component

model of an $r^{1/4}$-bulge and exponential-disk. The disk is augmented with spiral structures, star-formation regions, etc., based on scaling an image of the brightest low-redshift spiral galaxy available in the HDFS. The resulting template galaxies are formed with user-supplied values for bulge and disk radii, and bulge-to-total ratios.

Note that the resulting spatial distribution of galaxies in the field from this method is not entirely random. The galaxies are preferentially put down (according to a normal distribution with a specified "correlation length") next to randomly chosen original galaxies. The correlation length here is the FWHM (in arcseconds) of the normal distribution. This simulates the projected spatial "clumpiness" of a real distribution. The simulation actually puts several `clones' back in the field for each faint galaxy to mimic the steep increase in source counts at these magnitudes. The background is calculated using a model that includes a scattered and thermal component of the zodiacal light (Simon Morris 1999). In this simulation, however, the scattered component is multiplied by the observed NIR solar spectrum to simulate the effects of solar absorption features.

The simulation has some difficulty with overlapping galaxies. The masks used to separate galaxies are circular and the galaxies can have very small separations. In the case of extreme overlap, the average magnitude is used in the updated catalog. The spectral types may be mixed under these circumstances. That is, a portion of a galaxy might have a SED that is intermediate to the different overlapping types.

C.3.1.2 Star Fields

The artificial star fields are generated in a similar manner to the galaxy fields. Photometry was carried out in a deep HST WFPC2 field in M15 (Obtained from the CADC HST data archive). Published HST WFPC2 photometry of NGC 6397 (King et al. 1998) was used to make template colour-magnitude diagrams for the main sequence and white-dwarf sequence and provide the colour information for the different filters in the simulation. Each star in the original HST field is replaced with a NGST PSF with the same brightness. As with the galaxy simulation, for each star fainter than a given magnitude several clones are placed in the NGST field within a correlation length of other stars. The net effect is a field with the correct number counts and the same spatial distribution function at the original HST field. The stars in this entire field can be shifted fainter and made more-dense to simulate their appearance at an extragalactic distance. Furthermore, they can be superimposed on a galactic background (a scaled HST WFPC2 image of NGC 5457) of a given surface brightness.

C.3.2 NGST VI and NIR Camera

Both a virtual VI and NIR camera were developed. The filters are simulated with a specified efficiency curves and correspond to the HST WFPC2 and NICMOS filters. The detectors are also simulated and include size, pixel-scale, well-depth, gain, quantum-efficiency, dark current, readout and shot-noise. The user can select a particular 4.0 × 4.0 arcsec galaxy or star field for study from the artificial NGST deep fields and perform broadband imaging on it. The exposure duration and number of co-adds can be

specified. Given the user-inputs for detector size, pixel-scale, etc., the simulation outputs what the imagers would actually record on the detector. The simulation performs automated data reduction and photometry. Diagnostic numbers are output, such as the magnitude, radius, and bulge-to-total ratio of the galaxy. Also, in the case of star fields, the colour-magnitude diagrams for a given choice of filters.

This method of performing virtual imaging surveys with NGST was written as an IDL tool. It is included in a larger simulation of NGST instrumentation (along with a virtual Multi-Object Spectrograph (MOS)) called 'NGST VI/MOS' (Steinbring 1999B). The goal of the MOS investigation was to determine the optimal size of slits and compare the various slit-mask designs. This was investigated in a separate study by the author and will not be discussed further here. The results are available in Steinbring (1999A). A copy of the NGST VI/MOS software is available from the author. See http://astrowww.phys.uvic.ca/~steinb/ for a web-based version of NGST VI/MOS.

C.4 Results and Discussion

The goal of this work was to determine the advantage of observing with a dedicated VI imager over the NIR camera for wavelengths shorter than 1.0 micron. This is a straightforward experiment for NGST VI/MOS and was investigated using "virtual observing runs". Two main tests were carried out. The first was to image high-redshift galaxies and attempt to detect star formation regions and globular clusters. This is similar to the program suggested by Lilly et al. (1998) in their Design Reference Mission (DRM) proposal. The second was to attempt to detect the white dwarf sequence in an extra-galactic globular cluster. This is similar to the DRM proposal of Rich and Margon (1998).

A realistic artificial galaxy field was generated. It was purposefully made to be "clumpy". That is, the correlation-length parameter was set to 3.0 arcsec. This value was chosen because it is the smallest value for which severe overlapping of galaxies is not a problem. Setting this value to, say 0.5 arcsec, will cause almost all cloned galaxies to overlap another galaxy. Setting the value to 25.0 arcsec produces essentially a random distribution of galaxies on the sky. Several magnitude 20 - 30 galaxies were selected and imaged with both the VI and NIR camera through *R* and *I* filters. In all cases the the VI will have a distinct spatial resolution advantage over the NIR camera for wavelengths shortward of 1 um. Here, also, the efficiency advantage of Si-PIN detectors (and HgCdTe arrays) over InSb arrays is evident. This is very advantageous for programs of studying star-formation regions in galaxies with $z > 1$. For example, see Figure C-1 and C-2. The smallest structures clearly visible for VI imaging of this $z = 1$ galaxy from the NGST Deep Field vanish with imaging by the NIR camera. Note that the measurement of bulge-to-total (*B/T*) is noisier in the NIR camera image. This is typical of simulated observations made with this instrument. The poorer pixel sampling of the NIR camera makes fitting of a bulge function to the image more difficult - and the result is a noisier *B/T*.

A realistic globular cluster field was generated along the lines of what Rich and Margon (1998) propose to observe. It too was observed with both the virtual VI and NIR camera in *R* and *I* filters. One can see from Figure 1.2-4 that for photometry of point sources the VI will perform better than the NIR imager for short wavelengths. Here, M15 has been shifted away by a distance modulus of 8. With faint stars so close together the better sampling and increased throughput of the VI makes detecting faint point

sources easier. Imaging with a NIR camera would not detect the faint white dwarfs (I_{AB} >30) so clearly evident in the VI photometry.

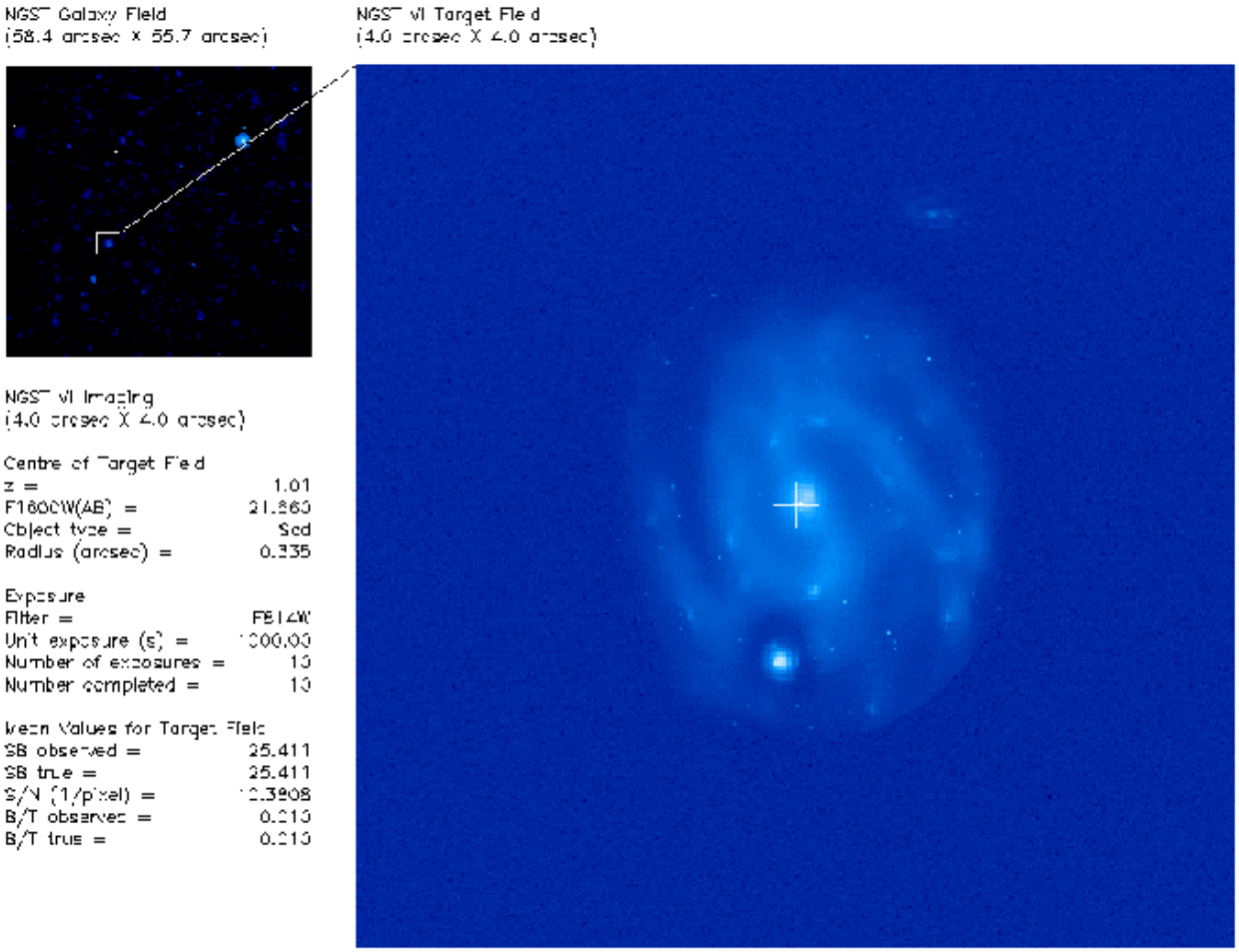


Figure C-1. An image of a bright, $z = 1$ galaxy with NGST VI. Note the resolve core and numerous bright globular clusters and star-formation regions in the disk.

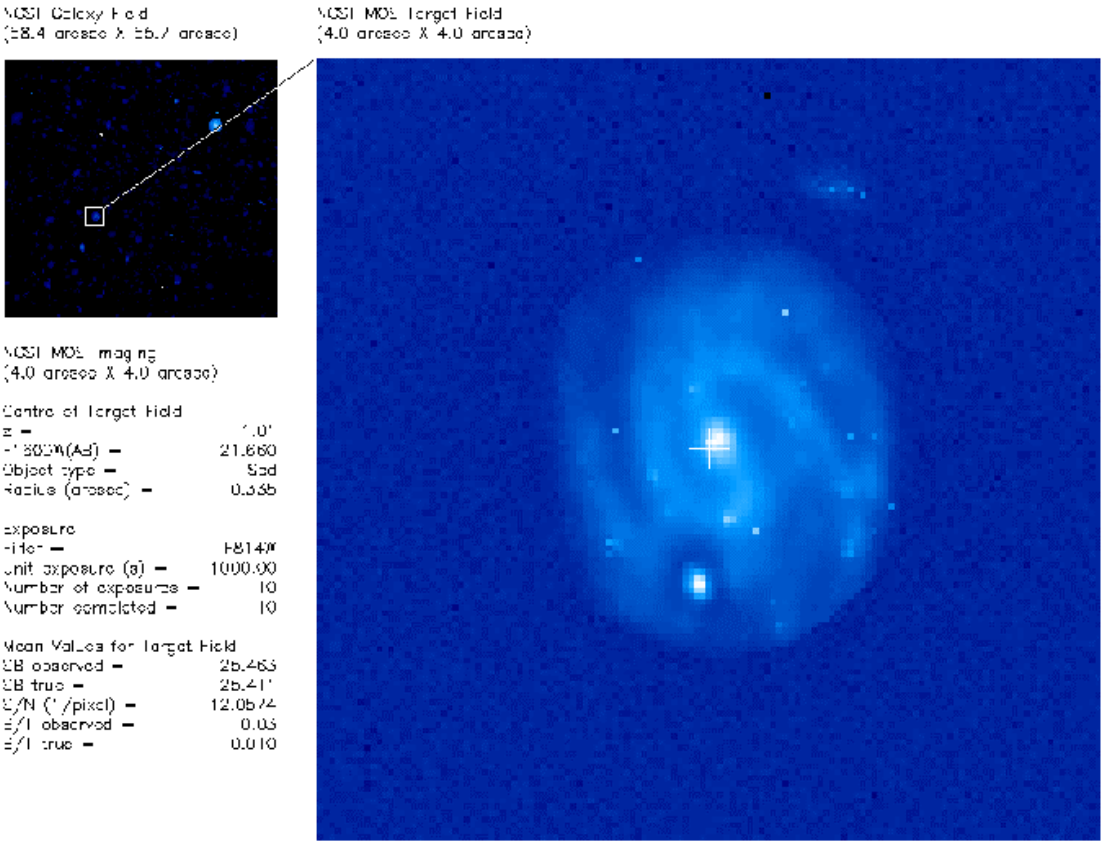


Figure C-2. The same field as in Figure C-1 imaged with the NIR camera. Note how the core is now blurred - most of the globulars and star-formation regions are now unresolved. The measurement of bulge-to-total is noisier than for the VI as well.

References

Bouwens, R., Broadhurst, T., Silk, J., 1997, astro-ph/9710291.

Burgarella, D., Bouwens, R., Broadhurst, T., Buat, V., Chapelon, S., Silk, J., 1998, astro-ph/9807169v2.

Burgarella, D., Buat, V., Dohlen, K., Zamkotsian, F., Mottier, P., 1998, astro-ph/9807168.

Gwyn, S., 1999, private communication.

King, I., R., Anderson, J., Cool, A., Piotto, G., 1998, ApJ, 492, L37.

Lilly, S., Fall, M., Stiavelli, M., Madau, P., Loeb, A., Gardner, J., Rieke, M., 1998, {\it The formation and evolution of galaxies I: The deep imaging survey(s)}, Ad-Hoc Science Working Group Design Reference Mission Proposal.

Lilly, S., Fall, M., Rieke, M., Stiavelli, M., Madau, P., Loeb, A., Gardner, J., 1998, {\it The formation and evolution of galaxies II: The deep spectroscopic survey(s)}, Ad-Hoc Science Working Group Design Reference Mission Proposal.

Morbey, C., 1999, *NGST PSF degradation with mirror surface roughness*, internal report, NRC.

Morris, S., 1999, private communication.

Rich, R., M., Margon, B., 1998, *The Ages and Chemistry of the Oldest Stellar Halo Populations*, Ad-Hoc Science Working Group Design Reference Mission Proposal.

Steinbring, E., 1999A, *A Data Simulation Study of the Next-Generation Space Telescope Multi-Object Spectrograph*, Canadian NGST MOS study, CSA.

Steinbring, E., 1999B, *NGST VI/MOS: A Virtual Next-Generation Space Telescope Visible Imager and Multi-Object Spectrograph*, NGST Science and Technology Exposition, Hyannis, MA, ASP Conf. Series, in press.

Wheelock, S., 1994, STScI internal report.

HST NICMOS website http//www.stsci.edu/ftp/instrument_news/NICMOS/nicmos_doc.html

HST STIS website: http://hires.gsfc.nasa.gov/

Kitt Peak Solar Observatory Data Archive website: http://www.nso.noao.edu/nsokp/dataarch.html

NGST website: http://ngst.gsfc.nasa.gov/

NGST VI/MOS website: http://astrowww.phys.uvic.ca/~steinb/